Smart Innovation, Systems and Technologies 368

Roumen Kountchev
Srikanta Patnaik
Kazumi Nakamatsu
Roumiana Kountcheva *Editors*

Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)

Artificial Intelligence and Wireless Communications, Volume 1

KES International

Springer

Smart Innovation, Systems and Technologies

Volume 368

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago, DBLP.

All books published in the series are submitted for consideration in Web of Science.

Roumen Kountchev · Srikanta Patnaik ·
Kazumi Nakamatsu · Roumiana Kountcheva
Editors

Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)

Artificial Intelligence and Wireless Communications, Volume 1

Springer

Editors
Roumen Kountchev
Technical University of Sofia
Sofia, Bulgaria

Kazumi Nakamatsu
University of Hyogo
Kobe, Japan

Srikanta Patnaik
Interscience Institute of Management
and Technology
Bhubaneswar, Odisha, India

Roumiana Kountcheva
TK Engineering
Sofia, Bulgaria

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-981-99-6640-0 ISBN 978-981-99-6641-7 (eBook)
https://doi.org/10.1007/978-981-99-6641-7

Preface

This is Volume one of the Proceedings of the International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023): *Artificial Intelligence and Wireless Communications.*

The conference took part on June 9–11, in Shenzhen, China.

The aim of ICAICT2023 was to provide a forum to researchers, academicians and practitioners to exchange ideas and discuss the latest achieved scientific results on artificial intelligence, communication technology and other related fields. The conference was focused at the establishment of an effective platform for institutions and industries to introduce the work of scientists, engineers, educators and students from all over the world.

After reviewing, 65 papers were accepted for presentation and publication, from which 33 are in this volume.

In memory of Prof. Dr. Roumen Kountchev, Dr. Sc., *Best Paper Award* was announced by the Organizing Committee. The *winner paper* is published as Chap. 1 in Volume one of this proceedings. The title of the paper is "Multi-Objective Dynamic Optimization Scheme for Unmanned Aerial Vehicles", with authors: Bao-Qiang Zhuang and Ya-Jun Wang.

The chapters, collected in this volume, present contemporary research works, based on artificial intelligence, deep learning and many other related algorithms like the improved PSO, security applications, lunar and space investigations, new methods for university and college training, intelligent image processing and remote sensing, etc. The chapters are arranged in four groups, which cover different parts of the related scientific areas.

The structure of Volume one is

Theme 1: Remote Sensing and Machine Learning (10 chapters)
Theme 2: Intelligent Image and Data Processing (13 chapters)
Theme 3: Health Systems and Security (5 chapters)
Theme 4: Intelligent Teaching Applications (5 chapters).

The book editors express their special thanks to IRNet International Academic Communication Center who organized this conference in correspondence with

their dedication to build platforms for scholars and researchers for better knowledge sharing, together with providing full-scale conference services that meet the standards from renowned publishing organizations.

We also thank Prof. Lakhmi Jain (Honorary Chair), Prof. Dr. Srikanta Patnaik (General Chair), Prof. Hang Li and Dr. Shoulin Yin (Organizing Chairs), Dr. Muhammad Ibrar, Dr. Asif Ali Laghari and Dr. Anas Bilal (Program Chairs), and Dr. S. R. Roumiana Kountcheva (International Advisory Chair).

The editors express their warmest thanks to the excellent Springer team for making this book possible.

Sofia, Bulgaria — Prof. Dr. Roumen Kountchev
Bhubaneswar, India — Prof. Dr. Srikanta Patnaik
Kobe, Japan — Prof. Dr. Kazumi Nakamatsu
Sofia, Bulgaria — Dr. Roumiana Kountcheva
July 2023

Contents

About the Editors

Roumen Kountchev is a professor at the Technical University of Sofia. He got his Ph.D. degree at the Institute of Telecommunications at Sankt-Petersburg, Russia and defended his dissertation for Doctor of Sciences at TU-Sofia. His scientific areas of interest are: digital signal and image processing, image compression, multimedia watermarking, video communications, pattern recognition and neural networks. Professor. R. Kountchev has 342 papers published in magazines and conference proceedings, and also: 30 books (20 published by Springer-Verlag, Germany); 50 book chapters and 20 patents (3 US). He had been principle investigator of 38 research projects. At present he is a member of Euro Mediterranean Academy of Arts and Sciences (EMAAS) and President of Bulgarian Association for Pattern Recognition (member of IAPR). Editorial board member of *International Journal of Reasoning-based Intelligent Systems*; *International Journal Broad Research in Artificial Intelligence and Neuroscience*; *KES Focus Group on Intelligent Decision Technologies*; *Egyptian Computer Science Journal*; *International Journal of Bio-Medical Informatics and e-Health*, and *International Journal Intelligent Decision Technologies*. He has been a plenary speaker in many international scientific conferences and workshops.

Prof. Srikanta Patnaik is presently working as director of Interscience Institute of Management and Technology, Bhubaneswar, India. He has received his Ph. D. (Engineering) on Computational Intelligence from Jadavpur University, India in 1999. He has supervised 32 Ph.D. Theses and more than 60 Master theses in the area of Computational Intelligence, Soft Computing Applications and Re-Engineering. Dr. Patnaik has published more than 100 research papers in international journals and conference proceedings. He is author of 3 text books and more than 100 edited volumes and few invited book chapters, published by leading international publisher like IEEE, ACM, Elsevier, Springer-Verlag, Kluwer Academic, SPIE, IOS Press and others. Dr. Patnaik is the Editor-in-Chief of *International Journal of Information and Communication Technology* and International *Journal of Computational Vision and Robotics* published by Inderscience Publishing House, England and also Editor-in-chief of Book Series on *Modeling and Optimization in Science and Technology* published by Springer, Germany. He is the Editor of *Journal of*

Information and Communication Convergence Engineering, published by Korean Institute of Information and Communication Convergence Engineering. Professor Patnaik is a Guest Professor to Hunan University of Finance and Economics, Changsha and Kunming University of Science and Technology, Kunming, China and visiting professors to some of the B-Schools of Europe and South East Asia. He is a member of Institute of Electrical and Electronics Engineering (IEEE) and Association for Computing Machinery (ACM). He is also Fellow of IETE, Life Member of ISTE, and CSI. Dr. Patnaik has visited various countries such as Japan, China, Hong Kong, Singapore, Indonesia, Iran, Malaysia, Philippines, South Korea, United Arab Emirates, Morocco, Algeria, Thailand and Vietnam for delivering Key note addresses at various conferences and symposiums.

Kazumi Nakamatsu received the Master of Science in Engineering and Doctor of Science from Shizuoka University and Kyushu University, Japan, respectively. His research interests encompass various kinds of logic and their applications to Computer Science, especially paraconsistent annotated logic programs and their applications. He has developed some paraconsistent annotated logic programs called ALPSN (Annotated Logic Program with Strong Negation), VALPSN(Vector ALPSN), EVALPSN(Extended VALPSN) and bf-EVALPSN (before-after EVALPSN) recently, and applied them to various intelligent systems such as a safety verification based railway interlocking control system and process order control. He is an author of over 180 papers and 30 book chapters, and 20 edited books published by prominent publishers. Kazumi Nakamatsu has chaired various international conferences, workshops and invited sessions, and he has been a member of numerous international program committees of workshops and conferences in the area of Computer Science. He has served as the Editor-in-Chief of the *International Journal of Reasoning-based Intelligent Systems (IJRIS)*, he is now the founding editor of IJRIS, and an editorial board member of many international journals. He has contributed numerous invited lectures at international workshops, conferences, and academic organizations. He also is a recipient of numerous research paper awards.

Roumiana Kountcheva got her Ph.D. at TU-Sofia, and became senior researcher in 1993. She is the Vice president of TK Engineering, Sofia and participates in R&D scientific projects of the Technical University of Sofia, Bulgaria. Her scientific interests are in image processing, image compression, digital watermarking, image tensor representation, CNCs, programmable controllers. She has more than 190 scientific publications and 5 patents and participated in 49 scientific research projects. R. Kountcheva is a member of Bulgarian Association for Pattern Recognition (IAPR); International Research Institute for Economics and Management (IRIEM); Institute of Data Science and Artificial Intelligence (IDSAI); Honorary Member of the Honorable Editorial Board of the nonprofit peer reviewed open access IJBST Journal Group, etc. Reviewer of WSEAS, NAUN, etc. She presented her research achievements in more 21 plenary speeches at international conferences and other scientific events and edited several books published in Springer SIST series.

Part I
Remote Sensing and Machine Learning

Chapter 1
Multi-objective Dynamic Optimization Scheme for Unmanned Aerial Vehicles

Bao-Qiang Zhuang and Ya-Jun Wang

Abstract In this paper, we analyzed the effective location of unmanned aerial vehicle (UAV) transmitting signal and the position adjustment of UAV in different formation, and established the mathematical model of multi-objective dynamic programming under the shortest path. The corresponding model is solved by using optimization algorithm and simulated annealing algorithm. The precise azimuth of a passive UAV is calculated from known polar coordinates and angles. Based on the three-target bearings-only triangulation method, the target triangle is obtained and the UAV position is determined by heart. The UAV with better position on the circumference is selected as the launch source.

1.1 Introduction

When the UAV is flying in formation, it is necessary to keep electromagnetic silence to prevent external interference, to reduce the electromagnetic wave signal sent out, and to avoid causing sparse sensor nodes and uneven distribution of anchor nodes [1]. Aiming at the problem of low localization rate and large localization error caused by the random distribution of anchor nodes, Huang [2] proposed a wireless transmission and localization algorithm based on the connectivity of anchor nodes. MDS-MAP algorithm [3] is used to reduce the positioning error. So it is common to adjust the UAV's position, maintain formation based on bearings-only passive positioning. The formation of UAV has a fixed number, and other UAVs relative position does not change. Chen et al. [4] divided the network into many hexagons and proposed a path planning algorithm for auxiliary beacon nodes, which reduced the location time and the average location error. Wang et al. [5] proposed a static anchor node

B.-Q. Zhuang (✉)
Department of Air Defense and Missile Defense, Air Force Engineering University, Xi'an 710051, China
e-mail: 2405538973@qq.com

Y.-J. Wang
Department of Basic Science, Air Force Engineering University, Xi'an 710051, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_1

layout optimization method based on geometric precision and degree factor, but the positioning precision is constrained by the size of anchor nodes and the geometric area of points. Huang et al. [6] proposed a breadth-first search (BFS) algorithm to select virtual anchor nodes, reducing the number of virtual anchor nodes and the length of the moving path of the nodes. Cheng et al. [7] proposed to use the grid method to model the known land, map, introduce vectors to guide the path direction in the algorithm, through multiple rewards and punishment measures to carry out the path planning. The angle between the receiving UAV and any two transmitting UAV is direction information.

A common model is exhibited in Fig. 1.1: FY01, FY02, and FY03 send signals, FY04 receives the signal angle θ_1, θ_2, and θ_3. In Fig. 1.1, FY00 is a signal transmitter under the same height circle formation, and it transmits signal through the two UAVs in the formation and the UAV in the center of the circle. For the model in Fig. 1.1, we established the location model to receiving UAV, and studied the minimum number of UAV transmitting signals to ensure the precise UAV location in Sect. 1.2. Meanwhile, we designed UAV position adjustment program by the initial data in Fig. 1.2. In Sect. 1.2, we also discussed corresponding UAV position adjustment scheme in a cone bearing-only formation (Fig. 1.3).

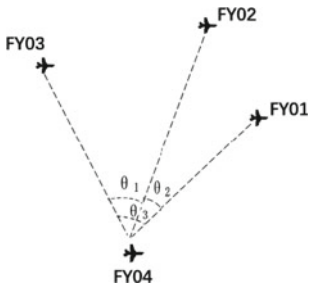


Fig. 1.1 Direction information of UAVs

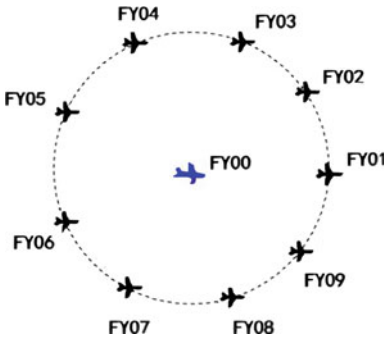


Fig. 1.2 Sketch of the formation of the circular UAVs

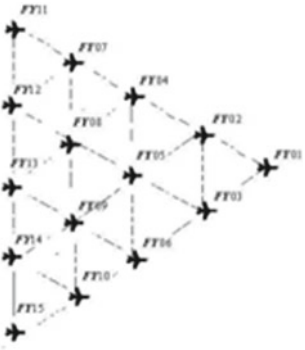


Fig. 1.3 Cone-shaped formation of UVAs

1.2 Analysis of UAV Position

1.2.1 Triangulation Method on UAV Positioning

The triangulation [8] method can detect the target position in different directions based on two or more UAVs, and can locate the target UAV position using the principle of plane triangle geometry. When the UAV transmits or receives a signal passively, its azimuth is the core parameter in target locating. The location model of passive received signal is established in this part. The azimuth angle of UAV can locate the target UAV, so that the target UAV can be located effectively. The position of the UAV is presented in Fig. 1.4.

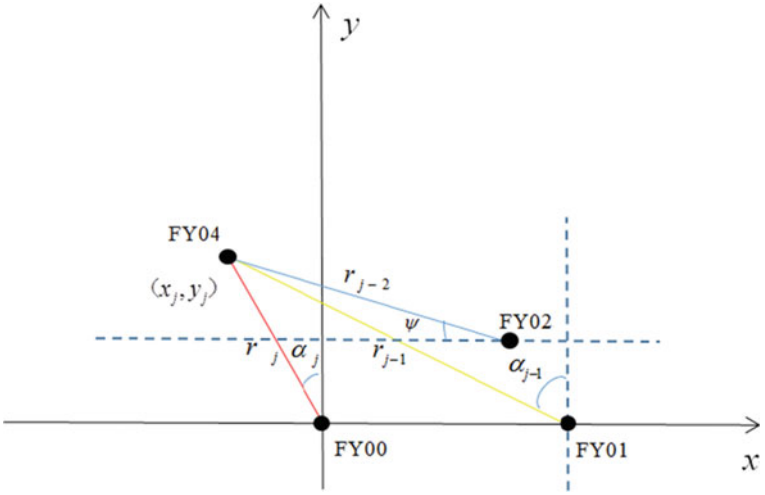


Fig. 1.4 Position of UAVs

Firstly, for special cases, we select special points FY01 and FY02 to send signals to FY04 to get the distance and angle, then establish the equation through the distance and angle, which can locate the position of FY04 effectively. From the position relation of the four UAVs in Fig. 1.4, we obtained:

$$\tan \alpha_j = \frac{x_j}{y_j} \tag{1.1}$$

$$\tan \alpha_{j-1} = \frac{r_j(\sin \alpha_j + 1)}{r_{j-1} \cos \alpha_{j-1}} \tag{1.2}$$

$$\begin{cases} x_{\mathrm{FY02}} = r_{j-2} \cos \psi - r_j \sin \alpha_j \\ y_{\mathrm{FY02}} = r_j \cos \alpha_j - r_{j-2} \sin \psi \end{cases} \tag{1.3}$$

Combining (1.1) to (1.3), three position relation equations of UAV can be obtained:

$$\begin{cases} x_1 = r_1 \sin \alpha_1 \\ y_1 = \frac{r_0^2 \sin 2\alpha_0}{2r_1(\sin \alpha_1 + 1)} \end{cases} \tag{1.4}$$

$$\begin{cases} x_2 = r_2 \sin \alpha_2 \\ y_2 = \frac{r_1^2 \sin 2\alpha_1}{2r_2(\sin \alpha_2 + 1)} \end{cases} \tag{1.5}$$

$$\begin{cases} x_3 = r_3 \sin \alpha_3 \\ y_3 = \frac{r_2^2 \sin 2\alpha_2}{2r_3(\sin \alpha_3 + 1)} \end{cases} \tag{1.6}$$

$\alpha_j (j = 1, 2, 3)$ is the azimuth angle of the target position (x_j, y_j), where the azimuth angle rotates clockwise from the positive half axis of the y axis, and the range is $[0, 2\pi]$.

1.2.2 Three-Target Bearings-Only Triangulation on UAV Positioning

Three-target bearings-only triangulation are a method of transmitting signals from UAV at different positions to the same target. If the UAV position is unbiased, three azimuth lines intersect at one point, the position coordinates of the passive receiving UAV can be obtained. However, the UAV is affected by the noise in the atmosphere caused by an electromagnetic wave signal it sends. So there are some position deviations in the actual circumstances, causing that three lines cannot meet at a point, but form a target triangle.

In Fig. 1.5, the selecting UAVs randomly FY0M, FY0N, and FY00 send signals to FY0K at the same time. After the target triangle is determined, we determine the

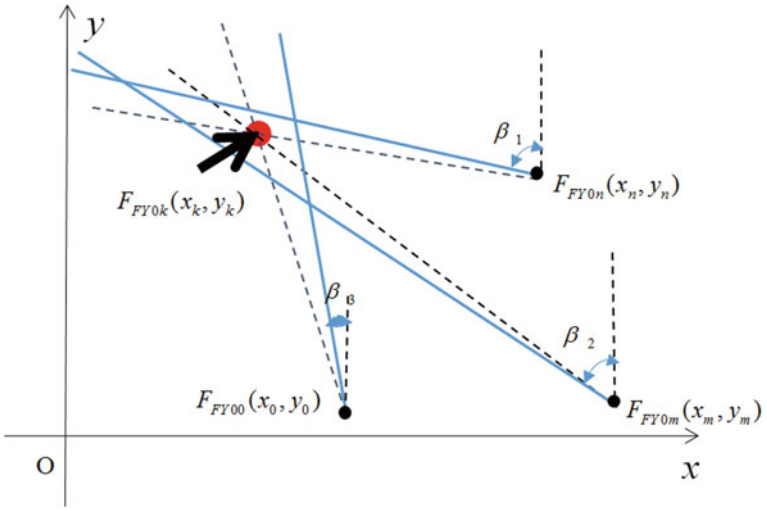


Fig. 1.5 Three-target cross-localization of geometric sketch

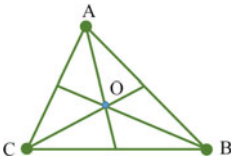


Fig. 1.6 Triangle heart with angle bisector

position of the UAV (Fig. 1.6) using the center method, where we take the intersection of the triangle's three bisectors as the target point.

As shown in Cartesian coordinate system in Fig. 1.6, we assume that the three transmitting targets are $F_{nm}(x_{nm}, y_{nm})$, $F_{n0}(x_{n0}, y_{n0})$, and $F_{m0}(x_{m0}, y_{m0})$, and the passive receiving target is FY0k(x_0, y_0). In this situation, the angles between receiving target and three transmitting targets are β_1,β_2, and β_3, respectively. According to the method of double-target triangulation, the intersection coordinates (x_{nm}, y_{nm}) of FY0n and FY0m are:

$$\begin{cases} x_{nm} = \frac{x_m \sin\beta_n \cos\beta_m - x_n \cos\beta_n \sin\beta_m + (y_n - y_m)\sin\beta_n \cos\beta_m}{\sin(\beta_n - \beta_m)} \\ y_{nm} = \frac{y_n \sin\beta_n \cos\beta_m - y_m \cos\beta_n \sin\beta_m + (x_m - x_n)\sin\beta_n \cos\beta_m}{\sin(\beta_n - \beta_m)} \end{cases} \tag{1.7}$$

The intersection coordinates (x_{n0}, y_{n0}) of FY0n and FY00 are:

$$\begin{cases} x_{n0} = \frac{x_0 \sin\beta_n \cos\beta_0 - x_n \cos\beta_n \sin\beta_0 + (y_n - y_0)\sin\beta_n \cos\beta_0}{\sin(\beta_n - \beta_m)} \\ y_{n0} = \frac{y_n \sin\beta_n \cos\beta_0 - y_0 \cos\beta_n \sin\beta_0 + (x_0 - x_n)\sin\beta_n \cos\beta_0}{\sin(\beta_n - \beta_0)} \end{cases} \tag{1.8}$$

The intersection coordinates (x_{m0}, y_{m0}) of FY0m and FY00 are:

$$\begin{cases} x_{m0} = \frac{x_m \sin\beta_m \cos\beta_0 - x_0 \cos\beta_m \sin\beta_0 + (y_m - y_0)\sin\beta_m \cos\beta_0}{\sin(\beta_n - \beta_m)} \\ y_{m0} = \frac{y_0 \sin\beta_m \cos\beta_0 - y_m \cos\beta_m \sin\beta_0 + (x_0 - x_m)\sin\beta_m \cos\beta_0}{\sin(\beta_m - \beta_0)} \end{cases} \tag{1.9}$$

The vertex coordinates of the target triangle can be determined by (1.10). The inner coordinates of the triangle are

$$x_{is} = \frac{Ax_{nm} + Bx_{n0} + Cx_{m0}}{A + B + C} \tag{1.10}$$

$$y_{is} = \frac{Ay_{nm} + By_{n0} + Cy_{m0}}{A + B + C} \tag{1.11}$$

The length of the triangles is

$$\begin{aligned} l_1 &= \sqrt{(x_{n0} - x_{m0})^2 + (y_{n0} - y_{m0})^2} \\ l_2 &= \sqrt{(x_{nm} - x_{m0})^2 + (y_{nm} - y_{m0})^2} \\ l_3 &= \sqrt{(x_{nm} - x_{n0})^2 + (y_{nm} - y_{n0})^2} \end{aligned} \tag{1.12}$$

In conclusion, we discussed the situation of UAV transmitting and receiving signals under special circumstances, and we calculated the precise positioning of UAV by establishing polar coordinate equations. Due to the existence of errors, the azimuth lines form a target triangle, and the intersection points formed by the two azimuth lines of the UAV are calculated, finally, the intersection point formed by the triangle bisector of the target is obtained, which is the precise positioning of the UAV.

1.3 The Establishment and Solution of Model

Assuming that FY00 and FY01 send a signal to a misplaced UAV (x_u, y_u), the angle is β_u, we establish the model equations:

$$\begin{cases} x_{01} = \frac{x_1 \sin\beta_0 \cos\beta_1 - x_0 \cos\beta_0 \sin\beta_1 + (y_0 - y_1)\sin\beta_0 \cos\beta_1}{\sin(\beta_0 - \beta_1)} \\ y_{01} = \frac{y_0 \sin\beta_0 \cos\beta_1 - y_1 \cos\beta_0 \sin\beta_1 + (x_1 - x_0)\sin\beta_0 \cos\beta_1}{\sin(\beta_0 - \beta_1)} \end{cases}$$

$$\begin{cases} x_{0u} = \frac{x_u \sin\beta_0 \cos\beta_u - x_0 \cos\beta_0 \sin\beta_u + (y_0 - y_u)\sin\beta_0 \cos\beta_u}{\sin(\beta_0 - \beta_1)} \\ y_{0u} = \frac{y_0 \sin\beta_0 \cos\beta_u - y_u \cos\beta_0 \sin\beta_u + (x_u - x_0)\sin\beta_0 \cos\beta_u}{\sin(\beta_0 - \beta_u)} \end{cases}$$

$$\begin{cases} x_{1u} = \frac{x_u \sin\beta_1 \cos\beta_u - x_1 \cos\beta_u \sin\beta_1 + (y_1 - y_u)\sin\beta_1 \cos\beta_u}{\sin(\beta_0 - \beta_1)} \\ y_{1u} = \frac{y_1 \sin\beta_1 \cos\beta_u - y_u \cos\beta_1 \sin\beta_u + (x_u - x_1)\sin\beta_1 \cos\beta_u}{\sin(\beta_1 - \beta_u)} \end{cases} \tag{1.13}$$

$$\begin{cases} x_{01} = x_{0u} = x_{1u} \\ y_{01} = y_{0u} = y_{1u} \end{cases}$$

To solve the equations, we analyze coefficient matrix

$$A = \begin{bmatrix} \frac{\sin\beta_0\cos\beta_u}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_0\cos\beta_u}{\sin(\beta_0-\beta_1)} \\ \frac{-\cos\beta_u\sin\beta_1}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_1\cos\beta_u}{\sin(\beta_0-\beta_1)} \end{bmatrix}$$

The matrix determinant $|A| \neq 0$. If the solution of linear equations exists, the UAV can locate target UAV effectively. According to the law of exhaustion, we suppose that there are two position deviations of UAVs in transmitting signal, so the position deviation subjected to the same point of constraint:

$$\text{s.t.} \begin{cases} x_{01} = x_{0u1} = x_{1u1} = x_{0u2} = x_{1u2} \\ y_{01} = y_{0u1} = y_{1u1} = y_{0u2} = y_{1u2} \end{cases} \tag{1.14}$$

We obtained a nonlinear equation describing the position relationship between UAVs:

$$\begin{cases} x_{01} = \frac{x_1\sin\beta_0\cos\beta_1 - x_0\cos\beta_0\sin\beta_1 + (y_0-y_1)\sin\beta_0\cos\beta_1}{\sin(\beta_0-\beta_1)} \\ y_{01} = \frac{y_0\sin\beta_0\cos\beta_1 - y_1\cos\beta_0\sin\beta_1 + (x_1-x_0)\sin\beta_0\cos\beta_1}{\sin(\beta_0-\beta_1)} \end{cases}$$

$$\begin{cases} x_{0u1} = \frac{x_{u1}\sin\beta_0\cos\beta_{u1} - x_0\cos\beta_0\sin\beta_{u1} + (y_0-y_{u1})\sin\beta_0\cos\beta_{u1}}{\sin(\beta_0-\beta_1)} \\ y_{0u1} = \frac{y_0\sin\beta_0\cos\beta_{u1} - y_{u1}\cos\beta_0\sin\beta_{u1} + (x_{u1}-x_0)\sin\beta_0\cos\beta_{u1}}{\sin(\beta_0-\beta_{u1})} \end{cases}$$

$$\begin{cases} x_{1u1} = \frac{x_{u1}\sin\beta_1\cos\beta_{u1} - x_1\cos\beta_{u1}\sin\beta_1 + (y_1-y_{u1})\sin\beta_1\cos\beta_{u1}}{\sin(\beta_0-\beta_1)} \\ y_{1u1} = \frac{y_1\sin\beta_1\cos\beta_{u1} - y_{u1}\cos\beta_{u1}\sin\beta_1 + (x_{u1}-x_1)\sin\beta_1\cos\beta_{u1}}{\sin(\beta_1-\beta_{u1})} \end{cases}$$

$$\begin{cases} x_{0u2} = \frac{x_{u2}\sin\beta_0\cos\beta_{u2} - x_0\cos\beta_0\sin\beta_{u2} + (y_0-y_{u2})\sin\beta_0\cos\beta_{u2}}{\sin(\beta_0-\beta_1)} \\ y_{0u2} = \frac{y_0\sin\beta_0\cos\beta_{u2} - y_{u2}\cos\beta_0\sin\beta_{u2} + (x_{u2}-x_0)\sin\beta_0\cos\beta_{u2}}{\sin(\beta_0-\beta_{u2})} \end{cases}$$

$$\begin{cases} x_{1u2} = \frac{x_{u2}\sin\beta_1\cos\beta_{u2} - x_1\cos\beta_{u2}\sin\beta_1 + (y_1-y_{u2})\sin\beta_1\cos\beta_{u2}}{\sin(\beta_0-\beta_1)} \\ y_{1u2} = \frac{y_1\sin\beta_1\cos\beta_{u2} - y_{u2}\cos\beta_{u2}\sin\beta_1 + (x_{u2}-x_1)\sin\beta_1\cos\beta_{u2}}{\sin(\beta_1-\beta_{u2})} \end{cases} \tag{1.15}$$

In transmitting electromagnetic wave signal, we established linear equations matrix based on (1.15) between the UAVs

$$B = \begin{bmatrix} \frac{\sin\beta_0\cos\beta_{u1}}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_0\cos\beta_{u1}}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_{u1}\cos\beta_{u2}}{\sin(\beta_0-\beta_1)} \\ \frac{-\cos\beta_{u1}\sin\beta_1}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_1\cos\beta_{u1}}{\sin(\beta_0-\beta_1)} & \frac{\sin\beta_{u2}\cos\beta_{u1}}{\sin(\beta_0-\beta_1)} \\ \frac{\sin\beta_0\cos\beta_{u2}}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_0\cos\beta_{u2}}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_{u2}\cos\beta_{u1}}{\sin(\beta_0-\beta_1)} \\ \frac{-\cos\beta_{u2}\sin\beta_1}{\sin(\beta_0-\beta_1)} & \frac{-\sin\beta_1\cos\beta_{u2}}{\sin(\beta_0-\beta_1)} & \frac{\sin\beta_{u1}\cos\beta_{u2}}{\sin(\beta_0-\beta_1)} \end{bmatrix}$$

The matrix determinant $|B| \neq 0$ implies that (1.15) are linear equations, and the coordinates of the targets can be obtained. It can be concluded that the two UAVs can be used as the signal transmitter to locate other UAVs effectively. Given

the initial position parameters of UAV, we can solve the linear equations of system to get the accurate positioning of UAV by MATLAB. The function code is $x = f\text{solve}(\text{fun}, x0)$, where fun equation to be solved, $x0$ is the initial value of calculation and x is the solution.

1.3.1 Multi-objective Dynamic Scheduling Model

In multi-objective dynamic scheduling model, we take the times of unmanned scheduling and the shortest distance from the received signal UAV to the ideal position as the optimization objectives, and establish the multi-objective UAV scheduling optimization model.

Objective Function

In each UAV adjustment scheme, we set the objective function according to the optimal objective with the minimum distance between the adjusted UAV and the ideal position:

$$\min z = \min N_i^{(k)} \times S^{(k)} \tag{1.16}$$

where $N_i^{(k)}$ is the number of scheduled UAVs and $S^{(k)}$ is the scheduling distance.

Constraint Conditions

1. Each UAV receives a signal in a dispatch only

$$\sum_{i=1}^{9} N_{ni} = 1 \tag{1.17}$$

2. After receiving the signal, each UAV changes position to schedule

$$f_i^{(k+1)} = f_i^{(k)} + x_i^{(k)}(1 - f_i^{(k)}) \tag{1.18}$$

3. The dispatching distance of UAV cannot exceed its next dispatching distance

$$\sum_{k=1}^{k} \sum_{i=1}^{9} x_i^{(k)}(s_{\text{i}}^{(k)} + s_{\text{Ti}}^{(k)} \times f_i^{(k)}) \leq 9 \times S_T \tag{1.19}$$

4. The $k + 1$th scheduling point $p^{(k+1)}$ is the end of kth scheduling point $q^{(\text{k})}$

$$q^{(k)} = p^{(k+1)} \tag{1.20}$$

5. The number of UAVs on the circle increased by 1

$$N_{ni}^{(k+1)} = N_{ni}^{(k)} + \sum_{i=1}^{9} \left(f_i^{(k)} + x_i^{(k)} \right) \tag{1.21}$$

6. Before the first dispatch, there is one UAVs on circle

$$N_{n0} = 1 \tag{1.22}$$

Therefore, we established the scheduling model of UAVs based on the shortest deviation angle from the ideal distance

$$\min z = \min N_i^{(k)} \times S^{(k)}$$

$$\text{s.t.} \begin{cases} k = 1, 2, 3 \ldots, i = 1, 2, \ldots 9 \\ \max(N_m^{(k)}, T_i^{(k)}) = \min T_i^{(k)} \\ \sum\limits_{i=1}^{9} N_{ni} = 1 \\ f_i^{(k+1)} = f_i^{(k)} + x_i^{(k)}(1 - f_i^{(k)}) \\ \sum\limits_{k=1}^{k} \sum\limits_{i=1}^{9} x_i^{(k)}(s_{\mathrm{i}}^{(k)} + s_{\mathrm{Ti}}^{(k)} \times f_i^{(k)}) \le 9 \times S_T \\ q^{(\mathrm{k})} = p^{(\mathrm{k+1})} \\ N_{ni}^{(k+1)} = N_{ni}^{(k)} + \sum\limits_{i=1}^{9} (f_i^{(k)} + x_i^{(k)}) \\ N_{n0} = 1 \end{cases} \tag{1.23}$$

1.3.2 The Solution of Model

Considering the angle of the UAV to reach ideal position, we obtained the heuristic information of the distance of UAV to reach the ideal position, and designed the objective function of the deviation angle

$$S^{(k)} = \min(s_{p(k)q(k)}^{(k)} + s_i^{(k)} N_i^{(k)}) \tag{1.24}$$

In the dynamic optimization model of UAV formation, we establish the model taking optimal flying angle as the objective function.

(1) The UAV has completed the scheduled tasks

For the scheduled UAVs, taking the optimal deviation angle as the objective function, we establish the model:

$$\theta_1 = \min \sum_{m=1}^{M} \sum_{k=1}^{V} \sum_{z=1}^{i} F_{mk} Y_{kz} d_z \tag{1.25}$$

F_{mk} is the number of UAVs, Y_{kz} is the size of the circle, and d_z is the distance between adjacent UAVs.

(2) The UAVs have not completed scheduled tasks

If the UAV has not received the signal or moved to the target position, we take optimal flying angle as objective function, and the shortest flying distance as the constraint condition:

$$\theta_2 = \sum_{m=1}^{\mathrm{m}} \sum_{k=0}^{\mathrm{I}_{mk}} \sum_{z=0}^{\mathrm{I}_{mk}} F_{mk} Y_{kz} d_z L_{mk} + \sum_{n=1}^{M} \left(\sum_{k=l_{mk}}^{V} \sum_{z=l_{mn}}^{V} F_{mk} Y_{kz} d_z l_{mk} + d_m l_m \right) \tag{1.26}$$

where θ_1 is the angle that drone approaches the target and θ_2 is the angle which the drone approaches the target:

$$\theta = \theta_1 + \theta_2 \tag{1.27}$$

We take (1.26) pluses (1.27) as the function of the optimal flying angle in dynamic scheduling optimization model:

$$\text{s.t.} \sum_{k=0,z=0}^{V} \sum_{k \neq z}^{V} P_{mk} O_{mk} d_z \leq \theta \tag{1.28}$$

In this case, the objective function is a dynamic optimization model, so we can solve the model of adjustment scheme for UAV with step acceleration method. First, we select the UAVs in good position except FY00 and FY01 as the launch source according to the initial data, then use the signals from UAVs to determine the locations of other UAVs. The positions of UVAs are adjusted according to the position information obtained. After an adjustment of position, we compare the positions of the seven UAVs other than FY00 and FY01, and select the UAVs with better positions as the launch source again. Repeating the above steps, if the final distribution of the UAVs has no deviation, all UAVs on the circle move to the exact position.

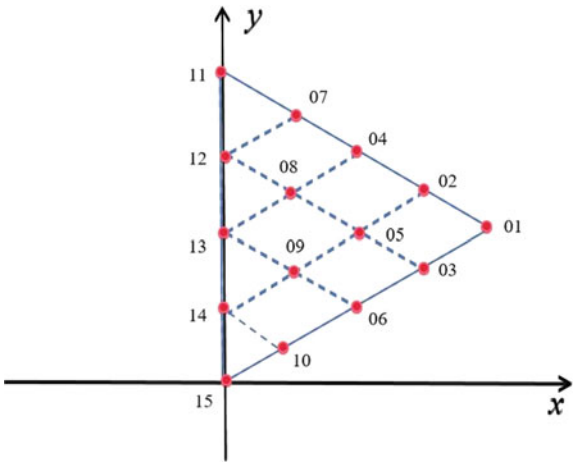


Fig. 1.7 Distribution of tapered formation UAVs at the same height

The UAVs Reach the Same Height

The distribution of tapered formation UAVs at the same height is shown in Fig. 1.7.

The coordinates of UAV can be calculated combining (7)

$$P_{\mathrm{FY0p}}\left(\frac{x_a+x_b+x_c}{3},\frac{y_a+y_b+y_c}{3},\frac{z_a+z_b+z_c}{3}\right)$$

where

$$\begin{aligned}
x_{\mathrm{FY0p}} &= \frac{x_b\sin\beta_a\cos\beta_b - x_a\cos\beta_a\sin\beta_b + (y_a-y_b)\sin\beta_a\cos\beta_b}{3\sin(\beta_a-\beta_b)} \\
&+ \frac{x_c\sin\beta_a\cos\beta_c - x_a\cos\beta_a\sin\beta_c + (y_a-y_c)\sin\beta_a\cos\beta_c}{3\sin(\beta_a-\beta_b)} \\
&+ \frac{x_b\sin\beta_b\cos\beta_c - x_c\cos\beta_b\sin\beta_c + (y_b-y_c)\sin\beta_b\cos\beta_c}{3\sin(\beta_a-\beta_b)} \\
y_{\mathrm{FY0p}} &= \frac{y_a\sin\beta_a\cos\beta_b - y_b\cos\beta_a\sin\beta_b + (x_b-x_a)\sin\beta_a\cos\beta_b}{3\sin(\beta_a-\beta_b)} \\
&+ \frac{y_a\sin\beta_a\cos\beta_c - y_c\cos\beta_a\sin\beta_a + (x_c-x_a)\sin\beta_a\cos\beta_c}{3\sin(\beta_a-\beta_c)} \\
&+ \frac{y_c\sin\beta_b\cos\beta_c - y_b\cos\beta_b\sin\beta_c + (x_c-x_b)\sin\beta_b\cos\beta_c}{3\sin(\beta_b-\beta_c)} \\
z_{\mathrm{FY0p}} &= H
\end{aligned} \tag{1.29}$$

$P_{\mathrm{FY0k_1}}$ is close to P_{FY0p}, and the coordinates of $P_{\mathrm{FY0k_1}}$,$P_{\mathrm{FY0k_2}}$, $P_{\mathrm{FY0k_3}}$ are (x_1, y_1, z_1),(x_2, y_2, z_2), (x_3, y_3, z_3), in which $Z_1 = Z_2 = Z_3 = H$. If

$P_{\mathrm{FY0k_1}}, P_{\mathrm{FY0k_2}}, P_{\mathrm{FY0k_3}}$ are non-collinear, we can determine the position of plane, otherwise we search the UAV position coordinates, and find the non-collinear points.

UAVs Are Flying at Different Height

When the formation UAVs are flying at different altitudes, we adopt the multi-target dynamic scheduling model, and calculate the stochastic reconfiguration optimization combining the simulated annealing algorithm [9]. According to the simulated annealing algorithm, when the annealing temperature is high and the annealing temperature drop is slow sufficiently, the solution of system converges to the global optimal solution. The steps of simulated annealing algorithm are as follows.

Step 1. Initial Solution

We choose random initial solution x_0 to generate the iterative algorithm at the initial simulated annealing temperature T_k $(k = 0)$.

Step 2. The Objective Function

When the temperature fluctuates around T_k, we continue the following operation until T_k reaches a stable equilibrium state. The difference of the solution $\left|f(x) - f(x')\right| = \Delta f$ is solved. If $\triangle < 0$, x' is feasible, otherwise, x' is accepted according to $\min(1, \exp(-\Delta f / T_k)) > \mathrm{rand}[0, 1]$.

Step 3. Annealing Cooling

We select the annealing coefficient C to annealing cooling. Among them $T_{k+1} = C_{T_k}, k = k + 1, C \in (0, 1)$, it should be noted that the initial annealing temperature should be high, and annealing temperature drops slow sufficiently.

Step 4. The End of Annealing

We judge that the simulated annealing is finished according to the termination annealing temperature. If the annealing temperature satisfies the convergence condition, the simulated annealing process ends, otherwise, returns to step 2.

The annealing temperature determines the optimal the solution of function, and the system accepts the non-optimal solution through the spatial probability. The simulated annealing algorithm can jump out of the local extremum and converge continuously, then get the global optimal solution. The multi-objective dynamic scheduling model is

$$\begin{cases} f_i^{(k+1)} = f_i^{(k)} + x_i^{(k)}(1 - f_i^{(k)}) \\ \sum_{k=1}^{k} \sum_{i=1}^{15} x_i^{(k)}(s_i^{(k)} + s_i^{(k)} \times f_i^{(k)}) \le 15S \\ q^{(k)} = p^{(k+1)} \end{cases} \tag{1.30}$$

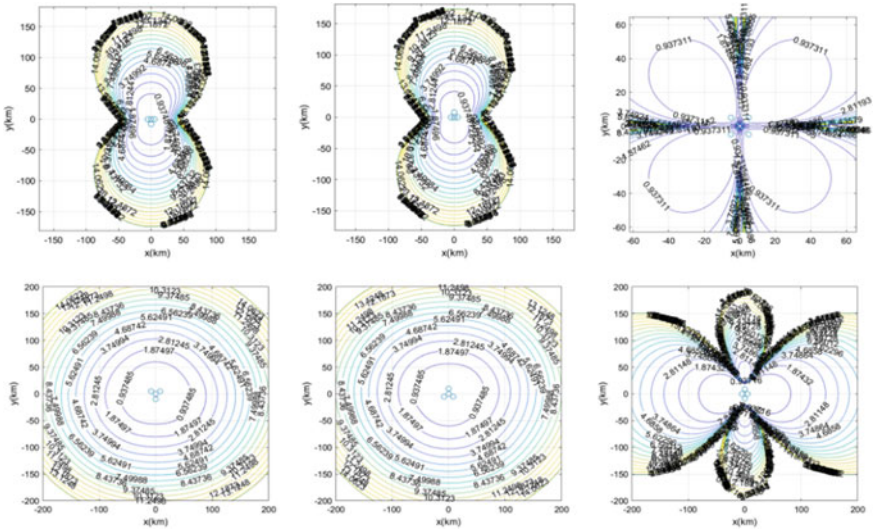


Fig. 1.8 Positioning accuracy under different formation shapes of UVAs

Scheduling Scheme of UAV Formation in Different Shapes

In actual, there are many UAV formations, such as t-type, y-type, diamond-type, and so on. Therefore, many simulation stations are used to simulate, and the positioning accuracy of various stations is compared according to the GDOP. According to the GDOP, the positioning accuracy of *Y* and inverted *Y* formations is higher than other formations. The GDOP of two UAV formations is more uniform than other UAV formations, so it is suitable for large-scale formation of UAVs. At the same time, the formation of the UAV can maintain strong positioning accuracy when it is disturbed by electromagnetic interference.

In Fig. 1.8, we present the comparison of positioning accuracy of bearings-only UAVs with different formation shapes, and give the specific adjustment scheme according to the dynamic scheduling model.

1.4 Conclusions

In this paper, the formation of a cluster of UAVs during formation flying is discussed, and a suitable positioning model is constructed by using multi-objective dynamic programming. Taking the regular and simple circular formation scenario as the starting point, a multi-objective dynamic optimization model based on bearings-only is established on the premise of knowing part of the UAV's position information and number, and the adaptive adjustment scheme in the real scene is discussed.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grants No. 12105365.

References

1. Han, G.J., Zhang, C.Y., Liu, T.Q., et al.: A multi-anchor nodes collaborative localization algorithm for underwater acoustic sensor networks. Wirel. Commun. Mob. Comput. **16**(6), 682–702 (2016)
2. Huang, Y., Fan, Y.: An optimized localization algorithm for mobile WSN based on the connectivity of anchors. Chin. J. Sens. Actuators **30**(12), 1925–1931 (2017)
3. Chang, L.J., Liu, M.Y., Zhang, L.C., et al.: A localization method for underwater wireless sensor networks based on modified particle swarm optimization algorithms. J. Northwestern Polytech. Univ. **33**(4), 648–654 (2017)
4. Chen, Y.R., Wan, J.H., Su, Z.Y., et al.: Study on movement path planning algorithm of auxiliary locating beacon node. Adv. Eng. Sci. **49**(2), 160–168 (2017)
5. Wang, H., Wang, K.: Optimal anchor deployment for target monitoring in wireless sensor net works. Comput. Eng. Des. **39**(7), 1807–1812 (2018)
6. Huang, B.Q., Du, Q.Z., Long, H.: Path planning algorithm of mobile anchor node for WSN. J. Yunnan Univ. (Nat. Sci. Ed.) **40**(1), 29–35 (2018)
7. Cheng, X.H., Qi, Y.: Indoor indicator path planning algorithm based on grid method. J. Chin. Inertial Technol. **26**(2), 236–240 (2018)
8. Zhang, M.: Research on image rectification and stitching of UAV laser line scanning 3D imaging system (2018)
9. Zhang, B., Yi, K., Hu, L.: Application of simulated annealing algorithm in path optimization. Proc. Comput. Sci. **17**(1), 95–101 (2004)

Chapter 2
Extraction and Fusion of Geographic Information from Multi-source Remote Sensing Images Based on Artificial Intelligence

Zirui Wang

Abstract Multi-source RS image fusion technology is mainly a data processing technology to organize and correlate the image data of the same scene under different imaging modes through specific calculation rules, and then obtain more accurate, perfect and rich information of comprehensive images. The information contained in remote sensing (RS) images of different sensors is imprecise, uncertain and fuzzy to varying degrees, so the fusion method used in fusing these information must solve these problems. The way to solve the problem of image registration is to solve the problem of relative correction of images. On the basis of analyzing and discussing the principle, hierarchy, structure and characteristics of multi-source RS image data fusion, this article puts forward the extraction and fusion technology of multi-source RS image geographic information combined with artificial intelligence (AI) algorithm. Compared with K-means classification method, this classification fusion method can effectively reduce the uncertain information in the classification process and improve the classification accuracy. The results verify the feasibility of the extraction and fusion method of geographic information from multi-source RS images proposed in this article in practical application.

2.1 Introduction

With the growth of modern RS technology, various Earth observation satellites continuously provide RS images with different spatial resolution, temporal resolution and spectral resolution. This multi-platform, multi-temporal, multi-spectral and multi-resolution RS image data is coming in at an alarming rate, forming a multi-source image pyramid in the same area [1]. As far as image information is concerned,

Z. Wang (✉)
School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, China
e-mail: 2239741578@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_2

there are multi-spectral images, visible light images, infrared images and synthetic aperture radar images, and there are also some materials about other attribute information [2]. Making full use of information from different sources to classify RS images will inevitably improve the classification accuracy. In RS images, the larger the values of spectral resolution and spatial resolution, the stronger the ability to distinguish different spectral features and the clearer the edges of the scene [3]. With the continuous growth of modern RS technology, the quality requirements of RS images in environmental detection, precision agriculture, urban planning and other fields are increasing [4]. The traditional RS image registration is completed by selecting control points. Generally, this method can achieve good results in the application of RS images with the same sensor, but if it is used in images with different imaging characteristics, there is a big error in registration [5]. This is due to the terrain differences caused by different imaging mechanisms of different images, which makes it difficult to select control points with the same name. On the other hand, the method of selecting control points requires a lot of manual intervention, so it is time-consuming and laborious [6]. Multi-source RS image fusion technology is mainly a data processing technology to organize and correlate the image data of the same scene under different imaging modes through specific calculation rules, and then obtain a more accurate, perfect and rich comprehensive image [7]. The biggest advantage of RS technology is that it can obtain a wide range of observation data in a short time, and display these data to the public through images and other means. Multi-modal data fusion technology represented by multi-spectral images and panchromatic images is the most widely used in multi-source RS image fusion processing [8]. The main purpose of this multi-modal data fusion is: aiming at the characteristics of low spatial resolution and high spectral resolution of multi-spectral images and high spatial resolution and low spectral resolution of panchromatic images, and the fusion algorithm is used to maximize and improve the spectral resolution and spatial resolution of comprehensive images, so that the spatial geometric features and spectral features can reach an optimal balance [9, 10]. Spatio-temporal fusion technology of optical RS data can generate fusion results with high temporal resolution and high spatial resolution corresponding to the scene image time through low spatial resolution and high spatial resolution image sequences and low spatial resolution images with predicted time [11]. On the basis of analyzing and discussing the principle, hierarchy, structure and characteristics of multi-source RS image data fusion, this article puts forward the extraction and fusion technology of multi-source RS image geographic information combined with AI algorithm.

2.2 Methodology

2.2.1 Basic Theory of Multi-source Data Fusion

RS image map is a kind of map which uses RS images and certain map symbols to represent the geographical spatial distribution and environmental conditions of cartographic objects. The drawing content of RS image map is mainly composed of images, and then some map symbols are used to mark it on the images, which is more conducive to expressing or explaining the drawing objects. Image map is richer in ground information than ordinary map, which has distinct contents and clear and easy-to-read drawings, which can not only show the specific content of the map, but also show all the information on the ground in a three-dimensional way [12]. Multi-source images include a lot of redundant information, and the fusion results make the system more fault-tolerant, reliable and robust, which reduces uncertainty and provides higher detection accuracy. At the same time, it can also eliminate the influence of irrelevant noise and improve detection ability. Geographical information is regional, multi-dimensional and dynamic, which plays an important role in various fields and brings convenience to people's life and work.

The production of geographic information data has the characteristics of long cycle, high cost, large engineering quantity and complex technology, which requires researchers to improve the technology of extracting information to obtain geographic information efficiently and intelligently, and further serve the growth of mankind. More complementary information can be obtained from images obtained from different types of sensors. The result of fusion makes various data optimized and integrated, providing more and richer information. Through feature-level image fusion, relevant feature information can be mined in the original image, the credibility of feature information can be increased, false features can be eliminated and new composite features can be established. Maps and RS images can show all kinds of geographical information in detail, and they can show all kinds of geographical changes in real time, making the geographical information obtained by people more accurate and three dimensional. The dynamic monitoring process of RS geographic information is shown in Fig. 2.1.

For feature extraction from different images, it is necessary to fuse them according to the same type of features on their respective images, so that useful image features can be extracted with high confidence. Feature-level fusion is suitable for the situation that it is impossible to combine multi-source data at pixel level, and it is simple and practical in many cases. Decision-level fusion is a high-level fusion. According to the requirements of the application, it first classifies the images, determines the characteristic images in each category and then carries out fusion processing. Distinguish the important targets contained in various images from other targets, carry out accurate and rapid inspection, increase the recognition and effectively extract the required geographic information. Generally, we can use microscopic feature extraction, dynamic change detection and other methods. The layered processing of ground objects can make full use of the characteristics of various ground objects in different bands and get good results.

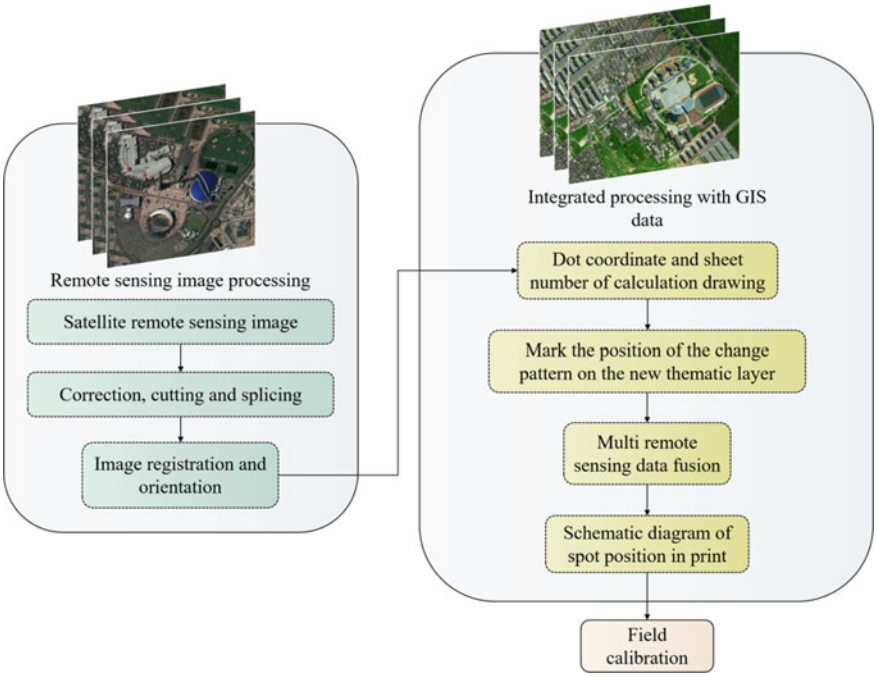


Fig. 2.1 Dynamic monitoring process of RS geographic information

2.2.2 Geographic Information Extraction and Fusion Algorithm of RS Image

Multi-source remote sensing image fusion is a composite model structure, which integrates the information provided by remote sensing image data sources of different sensors to obtain high-quality image information, and at the same time eliminates the redundancy and contradiction between multi-sensor information, complements them, reduces their uncertainty and ambiguity, so as to enhance the clarity of information in the image, improve the accuracy, reliability and utilization rate of interpretation and form a relatively complete and consistent information description of the target. The purpose of remote sensing data fusion is to synthesize the multi-band information of a single sensor or the information provided by different types of sensors, improve the timeliness and reliability of remote sensing information extraction and improve the efficiency of data use [13]. For multi-source RS images, registration refers to the optimal superposition of two or more image data from the same target area, at the same time or at different times, from different perspectives and obtained by the same or different sensors in the same coordinate system. Because the image matching accuracy of feature points is relatively high, feature extraction must be carried out before image matching.

Features are divided into point features and line features. Although point features are good, they often encounter noise points, so line features must be supplemented for identification. Feature-level fusion refers to the feature extraction of registered data, then the correlation processing, so that each sensor can get the feature vectors of the same target, and finally these feature vectors are fused for image classification or target recognition. Its advantage is that considerable information compression is realized, and the features provided are directly related to decision analysis. The data in X_j can be described by linear regression model:

$$x(t) = f\left(t, \theta_j\right) + \varepsilon_j(t), \quad t \in \left\{t_{j,1}, \ldots, t_{j,n_j}\right\} \tag{2.1}$$

The linear regression model corresponding to data segment X_j is:

$$f\left(t, \theta_j\right) = a_j t + b_j \tag{2.2}$$

The model parameter vector is:

$$\theta_j = \left[a_j, b_j\right]^{\mathrm{T}} \tag{2.3}$$

The linear regression model parameter a_j is called the trend characteristic value of X_j. The first data element $x\left(t_{j+1,1}\right)$ of X_{j+1} is called the dividing point of X_j.

Multi-source RS image data fusion covers a wide range, which obviously emphasizes the structure and foundation of RS image data fusion, rather than the processing technology and method itself as usually emphasized. Signal-level fusion refers to mixing signals from different sensors in some form to produce fused information with better quality and reliability. The function of pixel level fusion is to increase useful information components in the image in order to improve the effects of processing such as segmentation and feature extraction. Feature-level fusion enables the extraction of useful image features with high confidence. Symbol-level fusion allows information from multiple sources to be effectively utilized at the highest abstract level. Feature-level image fusion can not only increase the possibility of extracting feature information from images, but also obtain some useful composite features.

Features are abstracted and extracted from image pixel information. Typical feature information used for fusion includes edges, corners, textures, similar brightness regions and so on. The advantage of feature-level fusion is that it realizes considerable information compression and is convenient for real-time processing. Because the extracted features are directly related to decision analysis, the fusion results can give the feature information needed for decision analysis to the maximum extent.

Overlapping several image blocks whose image layout distance is less than a given threshold to form an image layout matrix Z, obtaining a set of processed image layout blocks Y through 3D transformation of the image layout matrix Z, and finally performing weighted average on all image layout blocks in each group Y to obtain a preliminary estimated image layout Y:

$$Y_b = \left(\sum\sum \omega_h Y\right) / \left(\sum\sum \omega_h X\right) \tag{2.4}$$

where X represents the characteristic function of an image layout block in the interval [0,1], and the image layout block set Y and the weight function ω_h are expressed as:

$$Y = T_{3d}^{-1}\{h[T_{3d}(Z)]\} \tag{2.5}$$

$$\omega_h = \begin{cases} 1/\left(\sigma^2 N_h\right) & \text{if } N_h \\ 1 & \text{else} \end{cases} \tag{2.6}$$

where $h[T_{3d}(Z)]$ stands for 3D transformation of image layout, T_{3d}^{-1} stands for inverse 3D transformation form of image layout, d stands for spatial dimension and h stands for hard threshold shrinkage coefficient.

Where σ^2 stands for zero mean Gaussian noise variance and N_h stands for nonzero coefficient retained after hard threshold filtering. The image layout Y_b is further grouped and matched to obtain the 3D matrix Y_w of RS image layout, and the original noise image layout block is matched to obtain the 3D matrix Z_w according to the corresponding coordinate information of Y_w. Performing inverse 3D transformation on the processed image layout matrix to obtain the processed image layout block set Y_{wi}, and finally performing weighted average on the spatial RS image layout blocks in Y_{wi} to obtain the final estimated RS image layout:

$$Y_f = \sum\sum \omega_{wi} Y_{wi} / \sum\sum \omega_{wi} X \tag{2.7}$$

The weight function ω_{wi} of the RS image layout block set Y_{wi} is given by the following formula:

$$\omega_{wi} = \sigma^{-2} \| W \|_2^{-2} \tag{2.8}$$

where W stands for empirical Wiener filtering shrinkage coefficient.

Remote sensing image fusion is one of the effective ways to solve the problem of multi-source mass data enrichment, which will help to enhance the ability of multi-data analysis and environmental dynamic detection, improve the timeliness and reliability of remote sensing information extraction, effectively improve the utilization rate of data, provide a good foundation for large-scale remote sensing application research and make full use of the remote sensing data obtained by spending a lot of money, that is, it has the characteristics of redundancy, complementarity, time limit and low cost.

2.3 Result Analysis and Discussion

Extracting geographic information from map images or RS images belongs to the problem of computer image understanding. The simulation operating system is Windows 11, the processor is Core i7 13700k, the graphics card is RTX 3070, the memory is 16 GB and the hard disk capacity is 1 TB. In our experiment and evaluation, we selected four polarimetric SAR image data, one from the spaceborne system (RADARSAT-2 of Canadian Space Agency) and the other from the airborne system (NASA/JPL CALTECH ARSAR). Image understanding is to use computer system to help explain the meaning of images, so as to realize the interpretation of the objective world by using image information. In order to complete a visual task, it is necessary to determine what information needs to be obtained from the objective world through image acquisition, what information needs to be extracted from the image through image processing and analysis and what information needs to be further processed to obtain the required explanation. In the training stage, according to the chromaticity information, different types of regions are artificially selected as training samples of the single subset in each information source identification framework and the gray average and variance of pixels in the selected different types of regions are used as the characteristics of obtaining the basic probability distribution function of each single subset. Figure 2.2 shows the influence of RS platform geographic information extraction and fusion on evenly distributed datasets. Figure 2.3 shows the influence of geographic information extraction and fusion on real datasets.

Because the node algorithm is simplified and compressed sensing is used to process the transmitted data, the data transmission amount of this scheme is greatly reduced, so the energy consumption is also low. The most important step in the integration process is to transform all or part of the organizational standards and organizational meanings of multi-source heterogeneous data based on a unified data

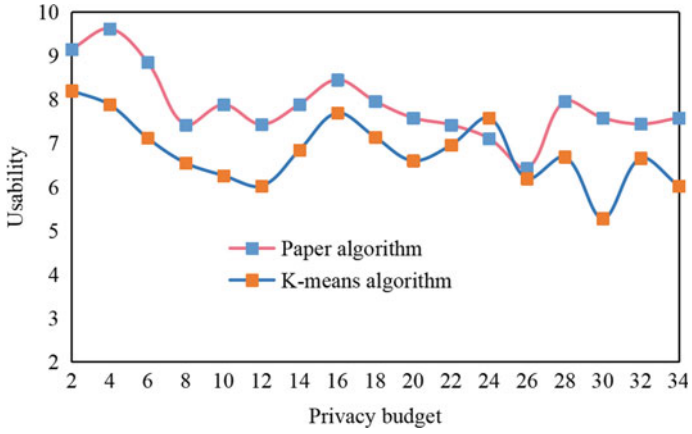


Fig. 2.2 Influence of geographic information extraction and fusion on evenly distributed datasets

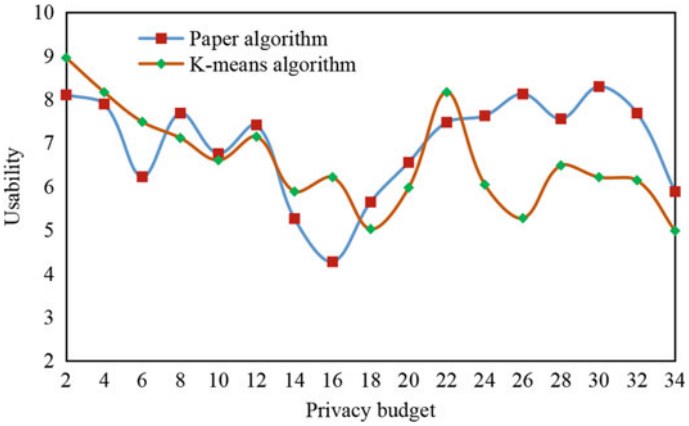


Fig. 2.3 Influence of geographic information extraction and fusion on real datasets

model, so that the integrated dataset has a global unified data model externally and the heterogeneous data are fully compatible with each other internally.

For the fusion of high spatial resolution images and low spatial resolution multispectral images, the purpose is to obtain multispectral images with enhanced spatial resolution. In fact, enhancing the resolution of multispectral images by fusion will inevitably lead to more or less changes in the spectral characteristics of multispectral images. By training the designed information fusion model with discrete geographic data, better network weights can be obtained. Then, substituting the obtained network weights into this model can become the basic model of geographic information extraction and fusion. Compare the output data of the information fusion model with the real geographic data, as shown in Fig. 2.4.

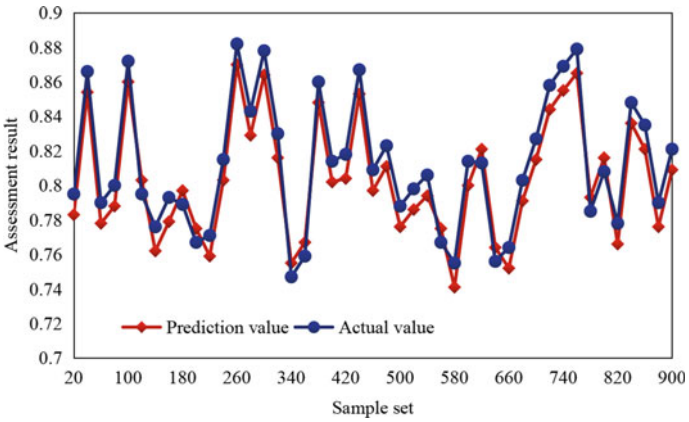


Fig. 2.4 Learning results of information fusion model

It is not difficult to see that the learning results of multi-source RS image geographic information fusion model are convergent and can approximate the original data well. Filter the rules according to the time sequence constraints satisfied by the attributes before and after the rules, and get the time sequence rules. Because the echo process of SAR sensor signal transmission is quite complicated, there are inevitably many error sources. The gain and error generated by radar antenna, transmitter and receiver, imaging processor and other parts will cause signal distortion, and radar images cannot accurately reflect the echo characteristics of ground objects. The radiometric calibration of SAR data can correct the gain and error of the whole SAR system from signal transmission to image generation by establishing the relationship between radar image and target cross-sectional area and backscattering coefficient, and convert the original data received by the sensor into backscattering coefficient.

For the general data fusion process, with the improvement of the fusion level, the higher the abstraction of data, the lower the homogeneity of each sensor, the higher the unity of data representation, the greater the data conversion, and the higher the fault tolerance of the system, but the lower the fusion level. The more detailed information is saved by fusion, but the processing capacity of fusion data increases, which requires high registration accuracy between the data used in fusion, and the dependence of fusion method on data source and its characteristics increases, which makes it difficult to give a general method of data fusion and has poor fault tolerance. There are two main factors that affect the output of information fusion model, namely whether the learning ability of the model is efficient and whether it has excellent generalization ability, and the input and output variables in the model will also affect the implementation effect of the model. Comparing this model with the traditional K-means algorithm, the result is shown in Fig. 2.5.

As can be seen from Fig. 2.5, after many iterations, the MAE of this method is obviously superior to the traditional K-means algorithm, and the error is reduced by 31.66%. The results show that the improved information fusion strategy enhances the robustness of the model and the rationality of the initial weight threshold setting,

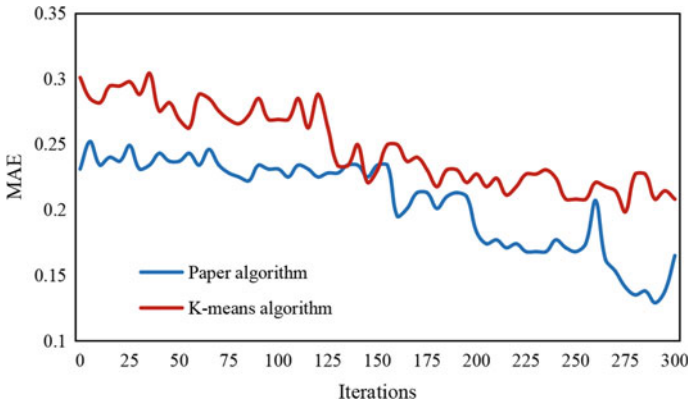


Fig. 2.5 Comparison of MAE

and effectively improves the classification performance of the model. The relative registration and fusion of multiple different images can not only provide a large range of high-resolution multi-spectral digital orthophoto images, but also provide a good basic data layer for geographic information systems. The data with low spatial resolution and high spectral resolution can be fused with the data with high spatial resolution and low spectral resolution by image fusion technology to obtain an image with the advantages of both data.

2.4 Conclusions

Multi-source RS image fusion technology is mainly a data processing technology to organize and correlate the image data of the same scene under different imaging modes through specific calculation rules, and then obtain more accurate, perfect and rich information of comprehensive images. When different sensors take RS images of the same terrain area, they will get many RS images, all of which have their own advantages and disadvantages. Complementary combination design fusion rules generate multi-source RS images, so as to achieve a fused image with better vision, clearer target, good redundancy and high complementarity and strong structure. Multi-source images include a lot of redundant information, and the fusion results make the system more fault-tolerant, reliable and robust, which reduces uncertainty and provides higher detection accuracy. On the basis of analyzing and discussing the principle, hierarchy, structure and characteristics of multi-source RS image data fusion, this article puts forward the extraction and fusion technology of multi-source RS image geographic information combined with AI algorithm. After many iterations, the MAE of this method is obviously better than the traditional K-means algorithm, and the error is reduced by 31.66%. The results show that the improved information fusion strategy enhances the robustness of the model and the rationality of the initial weight threshold setting, and effectively improves the classification performance of the model.

There are still many methods to measure the statistical characteristics of images, and the calculation indexes to judge the quality of fusion results have not been unified. Therefore, when constructing fractional differential, the measurement index selected in this paper has some limitations, and the selection of statistical features can be further studied in combination with the information difference between multi-spectral images and panchromatic images. The application of multi-spectral remote sensing data GIS in resource sharing and interoperability also needs further and deeper research.

References

1. Guan, X., Shen, H., Gan, W., et al.: A 33-year NPP monitoring study in Southwest China by the fusion of multi-source remote sensing and station data. Remote Sens. **9**(10), 1082 (2017)
2. Li, X., Zhang, H., Yu, J., et al.: Spatial–temporal analysis of urban ecological comfort index derived from remote sensing data: a case study of Hefei, China. J. Appl. Remote Sens. **15**(4), 042403 (2021)
3. Hiroki, M., Chikako, N., Iwan, R., et al.: Monitoring of an Indonesian tropical wetland by machine learning-based data fusion of passive and active microwave sensors. Remote Sens. **10**(8), 1235 (2018)
4. Ansari, A., Danyali, H., Helfroush, M.S.: HS remote sensing image restoration using fusion with MS images by EM algorithm. IET Signal Proc. **11**(1), 95–103 (2017)
5. Li, X., Long, J., Zhang, M., et al.: Coniferous plantations growing stock volume estimation using advanced remote sensing algorithms and various fused data. Remote Sens. **13**(17), 3468 (2021)
6. Han, Y., Liu, Y., Hong, Z., et al.: Sea ice image classification based on heterogeneous data fusion and deep learning. Remote Sens. **13**(4), 592 (2021)
7. Ren, J., Yang, W., Yang, X., et al.: Optimization of fusion method for GF-2 satellite remote sensing images based on the classification effect. Earth Sci. Res. J. **23**(2), 163–169 (2019)
8. Elmannai, H., Salhi, A., Hamdi, M., et al.: Rule-based classification framework for remote sensing data. J. Appl. Remote Sens. **13**(1), 1 (2019)
9. Douaoui, A., Nicolas, H., Walter, C.: Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. Geoderma **134**(1–2), 217–230 (2017)
10. Kussul, N., Lavreniuk, M., Skakun, S., et al.: Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. **2017**(99), 1–5 (2017)
11. Khodadadzadeh, M., Li, J., Prasad, S., et al.: Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. **8**(6), 2971–2983 (2017)
12. Sukawattanavijit, C., Jie, C., Zhang, H.: GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data. IEEE Geosci. Remote Sens. Lett. **14**(3), 284–288 (2017)
13. Mondal, A., Khare, D., Kundu, S., et al.: Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. Egypt. J. Remote Sens. Space Sci. **20**(1), 61–70 (2017)

Chapter 3
Architecture Simulation of Unmanned Lunar Exploration Platform Based on Modelica

Bo Ni, Yan Qu, Zhihui Liu, and Lifeng Xi

Abstract Faced to the multi-system collaboration of lunar exploration platform, the multi-level and multi-domain simulation model is built, based on the Modelica simulation language. The simulation model realizes the mechanism of the cooperative operation process, support the virtual verification of mobile detection, space communication, and energy supply. Based on the simulation model, the test case is carried out to verify the rationality and feasibility of the platform scheme.

3.1 Introduction

The unmanned lunar exploration platform is a typical large-scale complex system engineering. Compared with the traditional lunar probe [1], the unmanned lunar exploration platform has stronger multi-domain coupling, greater technical difficulty, longer mission cycle, and greater operation, maintenance, and coordination challenges. In addition, at the early phase of the demonstration and scheme design phase, it is needed to adopt the new generation of digital technology, construct the multi-task and multi-domain virtual simulation model, evaluate platform technology ability, in order to reduce repeated physical test and improve the design efficiency.

System-level multi-domain unified modeling and simulation technology provides methods and technical support for forward innovation design and full digital simulation verification of complex large-scale systems. Since the beginning of the twenty-first century, the system-level design and simulation technologies represented by model-based systems engineering (MBSE) and Modelica are becoming the core content of the new generation of digital technology reform, which provides support for the forward innovation design and full digital simulation verification of large

B. Ni (✉) · L. Xi
School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai, China
e-mail: niboone@126.com

Y. Qu · Z. Liu
Suzhou Tongyuan Software and Control Tech. Co. Ltd., Suzhou, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_3

complex systems. National Aeronautics and Space Administration (NASA), German Aerospace Center (DLR), and other enterprises have carried out MBSE and system simulation [2–8]. Domestic aerospace industry of the various research institutes is also carried out a lot of work, defines the space system oriented MBSE model system, and carried out based on Modelica system-level simulation work in various fields [9–13]. Five system models, including probe system, carrier rocket system, ground measurement and control system and launch site system, have been constructed and unified modeling and simulation has been carried out.

Faced up to the collaborative demonstration requirements of lunar unmanned exploration platform, based on the MWorks system modeling and simulation software platform, the lunar unmanned exploration platform system simulation model is constructed, and the simulation is taken for energy supply, mobile detection, surface transport, space communication, and command center space modules. Finally, one simulation case is taken to analyze the specified mission of the scene.

3.2 Overall Model Architecture

3.2.1 System Composition of Unmanned Lunar Exploration Platform

The unmanned lunar exploration platform includes four functional modules:

Mobile detection module has the function of carrying out mobile detection on the lunar surface according to the control command. It can complete the lunar surface movement through the lunar rover and other platforms. On the other hand, it can carry out lunar surface detection and complete the transmission of payload information.

The energy supply module provides energy supply for other modules, with solar power generation, mobile power supply, energy supply, and other functions;

The lunar surface communication module supports the lunar surface communication and the communication transfer between the lunar surface module and the relay communication satellite.

The command center has the function of mission control and resource scheduling with the cooperation of multiple devices within a certain range of the lunar surface.

Unmanned lunar exploration platforms is able to perform a variety of complex missions, such as scientific exploration, resource development and utilization, and lunar takeoff and landing. In each mission, the control center configures the task system flexibly, determines the relationship between energy supply and communication interface integration, defines the task command and load curve, drives the mobile detection module to complete the detection task, and returns the detection information. The unmanned lunar exploration platforms is a large-scale project in long-term operation. Not only can the platform module be constructed through multiple launch missions, but also new modules can be continuously launched to the platform to expand the scale and function of the platform.

3.2.2 System Simulation Model Architecture

The Modelica language and system modeling technology is object-oriented, and its simulation model architecture has the characteristics of top-down decomposition, which can be bottom-up integration to complete virtual simulation verification oriented to specific simulation tasks. According to the system structure and mission characteristics of the unmanned lunar exploration platform, the model base can be divided into three categories: platform scheme, function module, and mission flow. The mission flow model mainly refers to several typical mission modes of the unmanned lunar exploration platform. The functional module model is the main module of the unmanned lunar exploration platform including the information interface with the external system, while the platform scheme model is used to characterize the integration relationship between the modules of the platform.

The platform model library, which is constructed as Fig. 3.1, consists of relay communication satellite, lunar surface communication module, solar power generation device, energy supply module, mobile detection module, etc.

According to the model base architecture shown as shown in Fig. 3.1, simulation verification of system composition, system mechanism and mission operation logic can be carried out, including lunar takeoff and landing, lunar orbit transfer, scientific exploration, resource development and utilization, and deep space exploration transfer, etc.

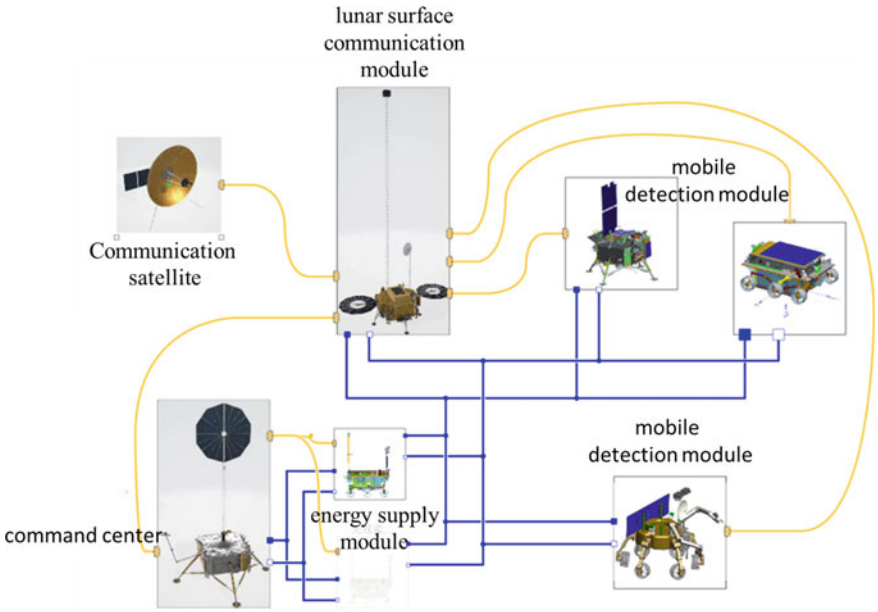


Fig. 3.1 Platform architecture

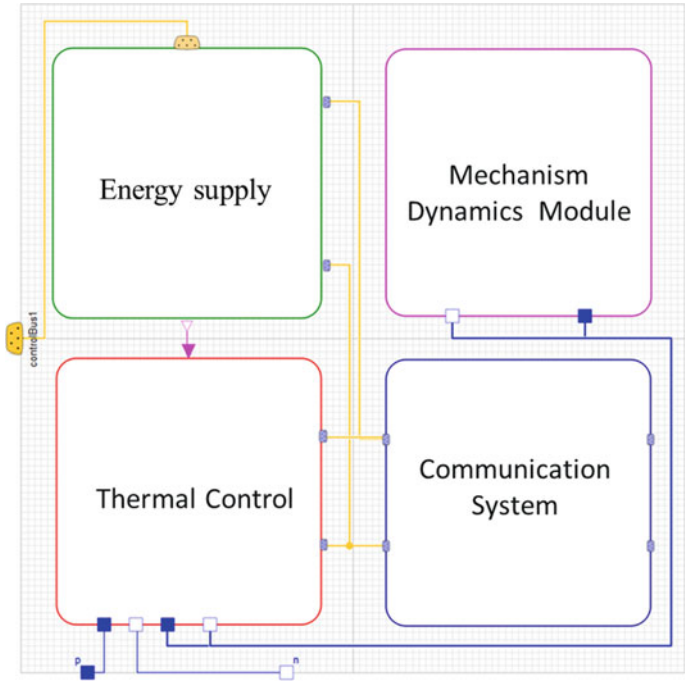


Fig. 3.2 General model architecture of multi-module

According to the system simulation requirements of different tasks and different configurations, the platform scheme templates are defined and can be configured in energy and communication aspects, including standard general interface model, so as to support the flexible configuration of the module models.

The functional module model is the basic component of the above model architecture, and it can still be further subdivided. To ensure that each functional module can operate independently within the lunar surface environment, it can be divided into four types of functions: dynamics, thermal control, power supply and distribution, and information communication, which is shown in Fig. 3.2.

3.2.3 Operation Process

Based on the model library of unmanned lunar exploration platform, system simulation models can be constructed with different schemes, different environments, and different parameters for the specified mission scenarios.

Its operation process is shown as follows:

System solution model integration: According to the system composition scheme, energy state, task state, and health state of each module, the component model is integrated into the overall system model based on the standardized module interface.
Definition of system operating environment: According to the platform design scheme and the exploration requirements of scientific targets, the position coordinates of the platform on the lunar surface, the lunar surface environment, sunshine and gravity acceleration, and other parameters are defined.
System task model configuration: According to the specific task process, define the running route of the mobile detection module, the detection task process, and the energy supply strategy to meet the task completion requirements, and complete the detection information feedback transmission mode and rate.
Technical capability verification: According to the requirements of the technical indicators of the platform, the technical capabilities of the platform in terms of mobile range, power generation and charging efficiency, command and control capability, lunar communication bandwidth are analyzed, and the technical capability evaluation of the platform is completed.

For the diversity of missions and environments of the unmanned lunar exploration platform, such as launch, exploration and test, and aiming at the actual state of long-term unmanned operation and maintenance of the system, the parameterized reconstruction of the platform architecture and mission scheme simulation model can be realized based on the configurable and extensible interface and architecture model.

3.3 Multi-domain Modeling Method

3.3.1 Common Function Modeling of Multi-module

As mentioned above, each platform contains functional components such as mechanism dynamics, thermal control, power supply and distribution, and information and communication. For general operation, energy control, information and communication requirements, the functional components of each module have similar modeling methods.

As is shown in Fig. 3.3, the thermal control model includes heat exchange equipment, medium, pipeline, sensor, example model, temperature control valve and other models. Taking sunshine radiation as the heat source, the models of electric heating, body heat transfer, heat dissipation, and heat control system are defined to describe the thermodynamic performance of the system.

Power supply model includes power supply system, power distribution system, load, environment, etc., including model for power supply equipment is divided into

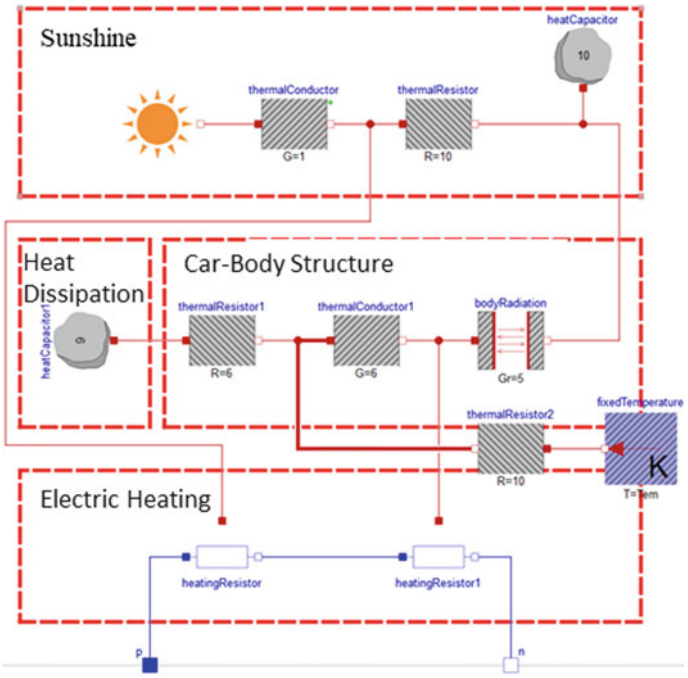


Fig. 3.3 Heat-control model

the solar cell array, battery, shunt regulator, charging conditions, discharge regulator, the main error amplifier, DC-DC converter, and so on various components model, support the realization of each module to charge and discharge, the analysis and calculation of power flow.

As is shown in Fig. 3.4, information communication model contains functions such as signal pretreatment, signal receiving and signal emission as well as the communication interface, signal preprocessing function according to system load encoding format, communication protocol for decoding, signal receiving, emission is used to select communication goals, used to define the communication link, and through the information bus interface, complete the definition of communication link and data transmission.

3.3.2 Mobile Detection Module Modeling

The mobile detection module contains two functions: movement and detection, and its composition structure is shown in Fig. 3.5. It receives movement commands and controls the module to move to the specified position, and then completes the scientific detection work in the specified form. The detection results are feedback to

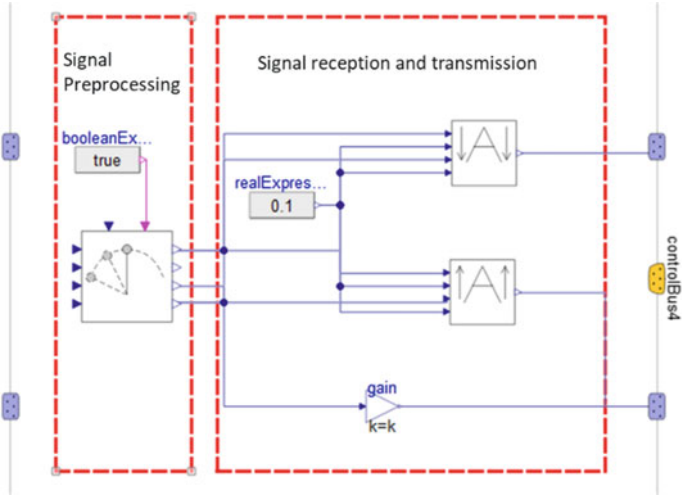


Fig. 3.4 Information communication model

the communication module in the form of digital signals and transmitted back to the ground measurement and control system through relay communication satellites.

(1) Mobile function modeling

The mobile function model includes motor drive module, rover chassis dynamics, steering control module, and covers vehicle sub-system models such as power system, transmission system, braking system, and control system. In order to facilitate the independent simulation work, the road model is also integrated into the mobile function model. Pavement model inherits the moon detection platform mobile wheel soil function model of module, including soil and all kinds of rigid wheel model components, include all kinds of surface driving scenarios test model of soil and the wheel, support for lunar rover, and the interaction between robot and soil, support for surface driving the traction performance through the analysis and calculation. The mobile function model can be used to study the motion performance of the mobile detection module, including vehicle control stability, suspension characteristics, traveling speed, load power, etc., under the specific geomorphology and acceleration conditions of the lunar surface.

(2) Probe function modeling

The detection function model is actually the payload system of the unmanned lunar exploration platform, which specifically includes the instruments and equipment loaded on the platform to directly realize the specific tasks to be completed by the scientific exploration target, including space exploration equipment, antennas and transponders.

Facing the moon no detection platform the overall performance of the verification platform, detailed model of specific detection equipment will have little impact on

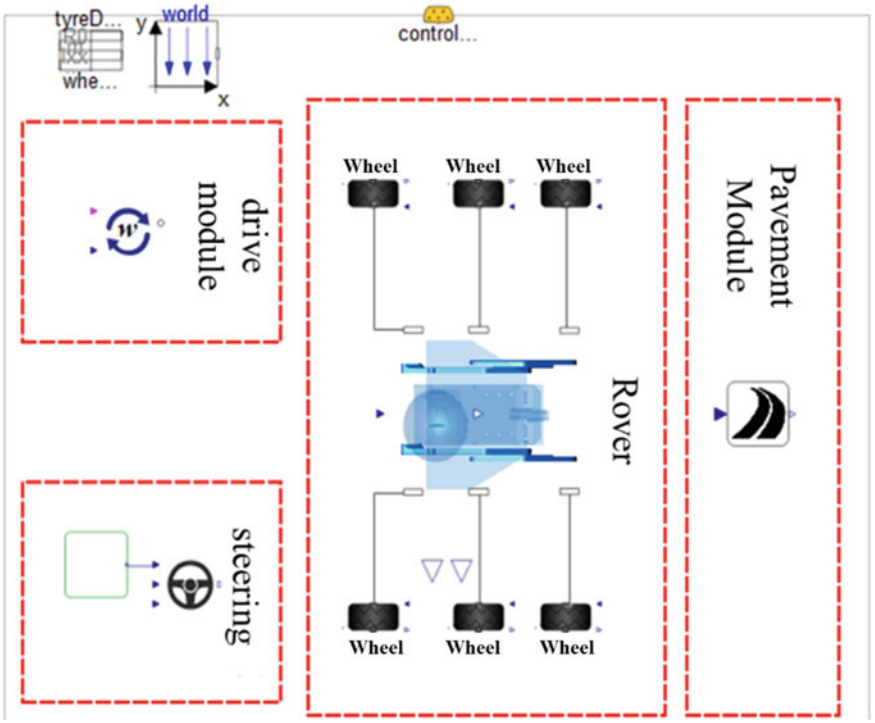


Fig. 3.5 Mobile detection module model

the verification results, so it can be simplified into energy and the information load, in addition, also including amplifiers and antenna model, receiving signals from the detection of function module, and the amplifier after processing, the signal is forwarded through the transmitting antenna.

Considering the granularity of the model, the detection function model realizes signal receiving, processing, frequency control and signal gain, and transmits the signal to the lunar surface communication module through the antenna and then to the relay communication satellite.

3.3.3 Energy Supply Module Modeling

The energy supply system is a sub-system of power generation of each module of the platform, and also a power source system of the lunar surface. It is composed of environmental model, solar battery array, battery model, power controller model, load model, and other sub-models to provide power for other modules, which is shown in Fig. 3.6. The power system mainly includes: environment, solar array, battery, power controller, load, telemetry command model, and so on.

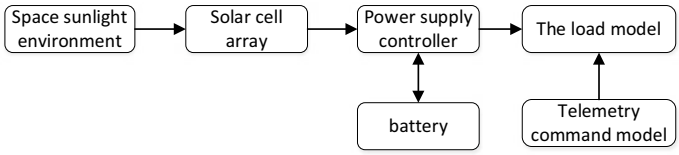


Fig. 3.6 Energy supply module model

In order to satisfy the simulation verification of power supply and distribution performance under different schemes, this paper constructs the immovable energy station model and the movable energy vehicle model. The simulation model takes both wired and wireless charging functions into account, and the model differences include charging power, interface connection mode, correlation checking form, number of concurrent charging devices, and other parameters or mechanism models.

3.3.4 Lunar Communication Module Modeling

The lunar surface communication module supports the data communication with the lunar surface movement detection module. Similar to the satellite communication link model, it can be defined as uplink and downlink, as Fig. 3.7. For the position relationship between the mobile detection module and the lunar surface communication module, the antenna direction and pitch angle can be adjusted. Combined with the load transmission demand of the mobile detection module, the load transmission power can be adjusted dynamically.

The scientific test data obtained by the unmanned lunar exploration platform on the lunar surface need to be transmitted back to the ground application system through relay satellites. In order to ensure the integrity of the simulation model, the relay satellite model needs to be established for simulating the lunar surface communication link or the lunar earth communication link.

The relay star model includes relay star control model, thermal control system model, power supply and distribution system model, information communication model, and other models. The relay star control model is used to realize coordination. The state machine model of Modelica is used to realize dynamic evaluation of relay satellite position, state, data communication rate, and load changes.

3.3.5 Modeling of Command Center Module

The command center module is used to realize the command and control of the mobile detection module, the energy supply module, and the lunar communication

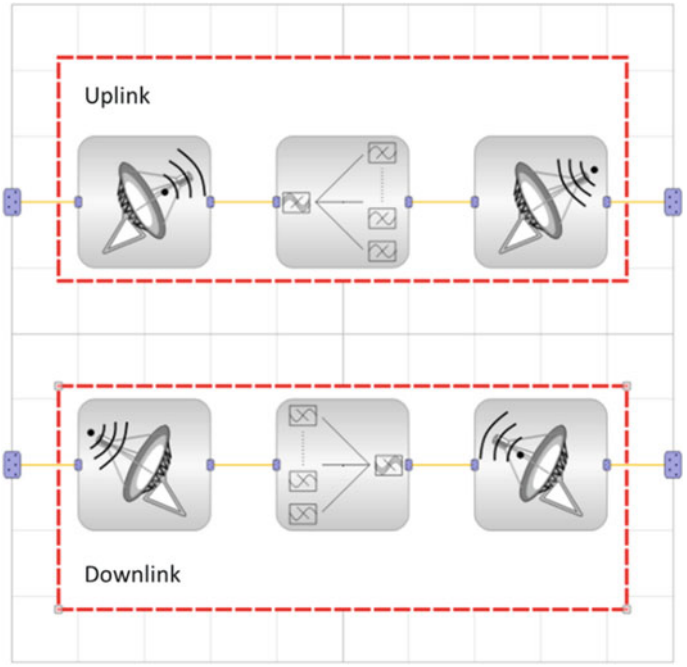


Fig. 3.7 Two place communication link model

module. The model includes the ground mission command analysis, the lunar platform operation process planning, the mission system construction, the control of each function module, and other functions.

Each mechanism model of the command center module is modeled in the form of state machine, and the system state and response mechanism of each module of the lunar unmanned exploration platform are defined, respectively. The system mission process planning can build a state machine logic model based on the technical capability of the lunar unmanned exploration platform. The control algorithm model of each function module can be combined with the acquisition, analysis and control algorithm modeling of each module state to achieve the control of each function module.

The typical status of each module is given as Table 3.1:

The core of the command center module is the task planning model, which can be realized by the above state machine model or completed by the external planning algorithm as Fig. 3.8. The task process is transformed into the position parameters and operation instructions of each module for task process control.

Table 3.1 System status parameters

No.	Function module	State parameter	Executable tasks
1	Mobile detection	Standby	ALL
2		Lack of electricity	NONE
3		Charging	NONE
4		Full recharge	ALL
5		Fault	Part OF
6	Lunar surface communication	Free	ALL
7		Start sending	NONE
8		Received	NONE
9		Establishing a lunar link	ALL
10		Establishing the moon-earth link	ALL
11		Congestion	NONE
12	Energy supply	Charging	Part OF
13		Full recharge	NONE
14		Discharge	Part OF
15		Congestion	NONE

```
parameter Integer Year = 20261018
parameter Modelica.SIunits.Angle Longitude = 2.16097941335678
parameter Modelica.SIunits.Angle Latitude = -1.5499745488186
parameter Modelica.SIunits.Temperature Tem = 323.15
parameter Modelica.SIunits.Mass mass_center = 1000
parameter Modelica.SIunits.Mass mass_tower = 200
parameter Integer patrolpath_rover = 1
    choice = 1 ,
    choice = 2 ,
    choice = 3 ));
parameter Modelica.SIunits.Mass mass_rover = 200
parameter Modelica.SIunits.Mass mass_energy = 200
parameter Integer wiremode = 2
parameter Integer patrolpath_robot = 1
    choice = 1 ,
    choice = 2 ,
    choice = 3 ));
parameter Modelica.SIunits.Mass mass_robot = 150
parameter Modelica.SIunits.Mass mass_satellite = 200
output Real route_rover[2] = {detector.dynamics.xyz_XSQ[1], detector.dynamics.xyz_XSQ[2]};
output Real Rem_rover = detector.electrical.battery.FD ;
output Real route_robot[2] = {collector_gc.dynamics.xyz[1], collector_gc.dynamics.xyz[2]} ;
output Real Rem_robot = collector_gc.electrical.battery.FD ;
output Real Rem_center = comCenter1_1.electrical.battery.FD ;
output Real Rem_tower = communicationTower.electrical.battery.FD ;
```

Fig. 3.8 Part of task model code

3.4 Mission-Oriented Integrated Simulation Applications

3.4.1 Example Description

According to the positioning of the unmanned lunar exploration platform, this paper determines the specific application scenario, defines the construction position of the unmanned lunar exploration platform, the relative position of the earth and moon, and the sun and moon, determines the task sequence, and completes the specific mission simulation and verification of the unmanned lunar exploration platform.

It can be divided into the following five stages:

S1 stage: The relay star receives the detection command and forwards it to the lunar communication module;
S2 stage: The lunar communication module transmits the command to the command center;
S3 stage: The command center conducts comprehensive calculation and planning for the task process, and decompositions it into the command sequence of detection, charging, and communication of the mobile detection module.
S4 Stage: The mobile detection module receives the control command from the command center, moves to the designated position, and completes other predefined mobile and detection tasks;
S5 Stage: Complete specific operations such as charging, moving, detection, and communication. The in situ resource development robot is launched to carry out in situ development tasks in the suitable area and can demonstrate various communication modes and energy supply modes.

3.4.2 Simulation Model Integration

On the basis of the system model, the model architecture and input parameters of the system simulation are determined, the configuration parameters and interface relations of the task system are established, the parameter changes of the system operating environment are described, and the parameter definitions can be instantiated for the configuration of the simulation model of the overall system and each module are provided, the system evaluation model is established, and the system operating indicators are defined, as is shown in Fig. 3.9. Through the simulation results, the function, performance, and task execution ability of the system are evaluated, and the virtual verification and evaluation of the task system are completed.

Mission simulation parameters are given in Table 3.2:

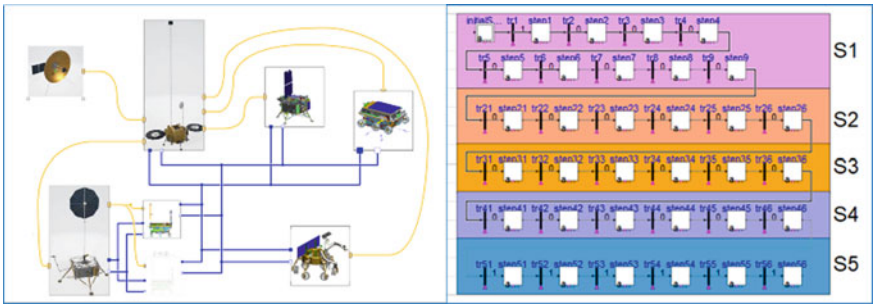


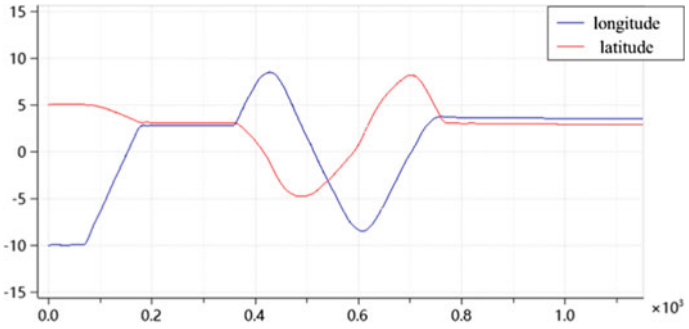
Fig. 3.9 Integrated simulation models with state machines

Table 3.2 Part of simulation parameters

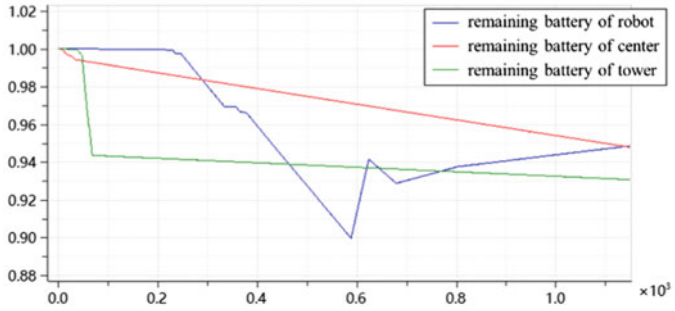
No.	Parameter name	Value
1	Moon position longitude	123.815
2	Moon position latitude	− 88.807
3	Ambient temperature	50
4	Mass	200 kg
5	Charging method	Wireless
6	Energy supply battery capacity	20,000
7	Power generation	20 kW
8	Motion detection peak power consumption	900W
9	Maximum movement speed	400 m/H
10	Maximum climbing angle	30°

3.4.3 Simulation Solution and Visualization Analysis

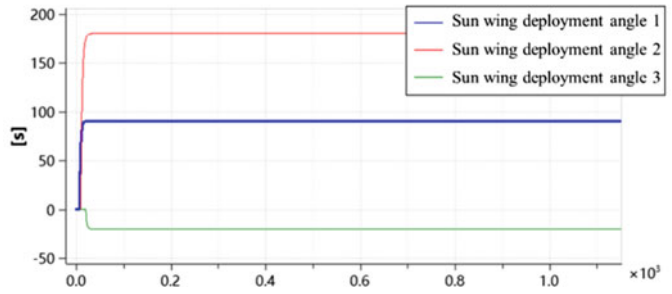
The above model and parameters are used to carry out the simulation solution of the platform mobile detection task, obtain the simulation results of communication link, antenna pointing, module power, communication data amount and other parameters, identify the operating state of the system, and verify the feasibility of the system scheme. The simulation results are shown as Fig. 3.10.



(a) Rover position



(b) Remaining battery



(c) Three deployment angles of sun-wing

Fig. 3.10 Simulation result of mobile exploring

3.5 Conclusion

Facing to the virtual verification requirements of the international lunar scientific research station system, aimed to the characteristics of multi-disciplinary mechanism coupling and multi-level system integration, the system simulation model library is

built based on Modelica language, which can support the collaborative verification between multiple modules. Further, configurable model architecture is built to realize the rapid integration for different tasks. It can provide reference data for system function performance analysis, technical index evaluation and scheme weighing, and improve the scientificity and comprehensiveness of task demonstration.

Acknowledgements This paper is prospered by National Key Research and Development Plan (SQ2019YFE20137), National Defense Basic Scientific Research Program (JCKY2020903B001), Civil Space Technology Pre-research Project (D020101).

References

1. Pei, Z.Y., Liu, J.Z., Wang, Q., et al.: Overview of lunar exploration and International Lunar Research Station (in Chinese). Chin. Sci. Bull. **65**, 2577–2586 (2020)
2. Jia, C.X., Wang, L.F.: Challenges and development suggestions of the model based systems engineering in China. J. Syst. Sci. **24**(04), 100–104 (2016)
3. Seal, D.: The Model-Based Engineering Diamond (2020)
4. Deng, Y.C., Mao, Y.X., Lu, Z.A., Xia, Q.W.: Analysis of the development of model-based systems engineering application. Sci. Technol. Rev. **37**(07), 49–54 (2019)
5. Haskins, C., Forsberg, K., Krueger, M., et al.: Systems Engineering Handbook (2006)
6. Lykins, H., Friedenthal, S., Meilich, A.: Adapting UML for an Object Oriented Systems Engineering Method (OOSEM). Wiley Online Library (2000)
7. Aleksandraviciene, A., Morkevicius, A.: MagicGrid Book of Knowledge. Vitae Litera, Kaunas, Lithuania (2018)
8. Hoffmann, H.: Harmony/SE: A SysML based systems engineering process. Innovation (2008)
9. Brower, E.W., Delp, C., Karban, R., et al.: OpenCAE case study: Europa lander concept. In: Model-Based Systems Engineering Products in the OpenCAE Model-Based Engineering Environment with Europa Lander as a Case Study (2019)
10. Delp, C.: Open model-based engineering environments (2019)
11. Bandecchi, M.: Concurrent engineering at ESA: from the concurrent design facility to a distributed virtual facility (2007)
12. Liu, J.Z., Pei, Z.Y., Yu, G.B., et al.: Space engineering multi-state holographic model and its applications. J. Astronaut. **40**(05), 535–542 (2019)
13. Zhang, B.N., Qi, F.R., Xing, T., et al.: Model based development method of manned spacecraft: research and practices. Acta Aeronautica Astronautica Sinica **41**(7) (2020)

Chapter 4
Trajectory Optimization Control System of Intelligent Robot Based on Improved Particle Swarm Optimization Algorithm

Ziyu Hu

Abstract Trajectory optimization is a hot topic in the field of intelligent robots, whose task is to plan the optimal motion trajectory that passes through a specified point and satisfies constraints such as velocity and acceleration based on a given target trajectory point. In the time optimization problem of robots, particle swarm optimization (PSO) has been widely applied in the time optimization problem of robotic arms due to its simple structure and adjustable parameters. This article conducts research on the intelligent robot trajectory optimization control system based on PSO, and the results show that through experimental data of three joints, it can be found that the optimized trajectory motion time has been reduced by an average of about 50%, achieving the expected goal. Compared with before optimization, PSO has stronger global search ability, faster convergence speed, and good stability, which can effectively improve robot work efficiency and maintain smooth operation. By dynamically adjusting the value of the learning factor in the particle swarm optimization algorithm through PSO, the particle swarm can search for the optimal value in a short period of time in the early stage of iteration and can quickly and accurately converge to the optimal solution in the later stage of iteration.

4.1 Introduction

The application engineering research of intelligent robots has attracted worldwide attention. Outdoor small intelligent mobile robots are a kind of engineering service robots, which have broad application prospects and can be applied to environmental cleaning, agricultural and forestry plant protection, resource exploration and other occasions. The motion control of intelligent robots can be divided into three categories: point stabilization, path tracking and trajectory tracking. In trajectory tracking control, the tracking path is related to time, so it is the most complicated. At present,

Z. Hu (✉)
School of Economics and Management, Shanghai Institute of Technology, Shanghai 200235, China
e-mail: 1292107005@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_4

there are mainly control based on kinematic model, control based on dynamic model and intelligent control [1]. At present, there are two kinds of trajectory planning: one is to optimize the time, and choosing the motion time as the optimization object can improve the working efficiency of the robot; the other is to optimize the system energy and choose the system energy as the optimization object, which can reduce the energy loss of the robot and prolong the service life [2]. Trajectory optimization is a hot issue in the robot field. Its task is to plan the optimal trajectory that passes through the specified point and meets the constraints of speed and acceleration according to the given target trajectory point. In the robot time optimization problem, PSO is widely used for its simple structure and adjustable parameters [3]. In this paper, the trajectory optimization control system of intelligent robot is studied based on PSO, and the time is optimized by PSO to find out the optimal trajectory. Experiments show that when selecting the optimal position of particle swarm, the poor solution is accepted with a certain jump probability to jump out of the local extreme value, and after a certain number of iterations, if the optimal individual in the swarm is found to be not obviously optimized, particles with different concentrations are inhibited or promoted [4, 5]. Fully considering the dynamic characteristics of the mobile robot, it "retreats" to the whole system step by step. Based on the dynamic model of the mobile robot, a globally stable trajectory tracking control law is designed, which has engineering application value. By dynamically adjusting the value of learning factor in PSO, the particle swarm can search for the optimal value in a short time at the beginning of iteration and converge to the optimal solution quickly and accurately at the end of iteration [6].

4.2 Research on Trajectory Planning of Intelligent Robots Based on Improved Particle Swarm Optimization Algorithm

4.2.1 Trajectory Optimization Control System

Intelligent robots have various internal and external information sensors, such as vision, hearing, touch and smell. As shown in Fig. 4.1, intelligent robots not only have receptors, but also effectors as a means of acting on the surrounding environment. This is the muscle, also known as the stepper motor, which moves the hands, feet, long nose, antennae and so on. From this, it can also be seen that intelligent robots must have at least three elements: sensory elements, reaction elements and thinking elements.

The difference between intelligent robots and industrial robots is that they have the ability to perceive, recognize, reason and judge like humans do. You can modify the program within a certain range based on changes in external conditions, which means it can adapt to changes in external conditions and make corresponding adjustments to itself. However, the principles for modifying the program are predetermined by

Fig. 4.1 Intelligent robot

individuals. This type of primary intelligent robot has already possessed a certain level of intelligence. Although it does not yet have the ability to automatically plan, it has also begun to mature and reach a practical level. The trajectory planning is divided into two stages during implementation. The first stage is called path planning, which refers to the various paths that require the robot to move to a designated position. How to move to the designated position, such as through a straight line or arc to move to the designated position. The second stage is called trajectory tracking, which means that for a known path or trajectory, the robot cannot perform motion uniformly according to it, but uses approximation to complete the given trajectory [7]. However, trajectory planning has always been a research hotspot in the field of intelligent robot control [8]. In this regard, this article will conduct research on the trajectory optimization control of intelligent robots based on PSO. The basic PSO adopts a "speed position" search model to solve optimization problems.

Elementary particle swarm optimization algorithm has fast convergence performance in the initial stage. The performance of this controller depends on the selection of controller parameters, and it takes a lot of work to determine these parameters by repeated experiments [9]. This article designs an intelligent robot trajectory optimization control system based on PSO and uses an improved particle swarm optimization algorithm for parameter optimization calculation. The overall structure of the controller is shown in Fig. 4.2.

In the trajectory optimization control system of the improved particle swarm optimization algorithm, the motion time of robot joints is regarded as particles in the search space, and each particle has a position attribute and a speed attribute, and the

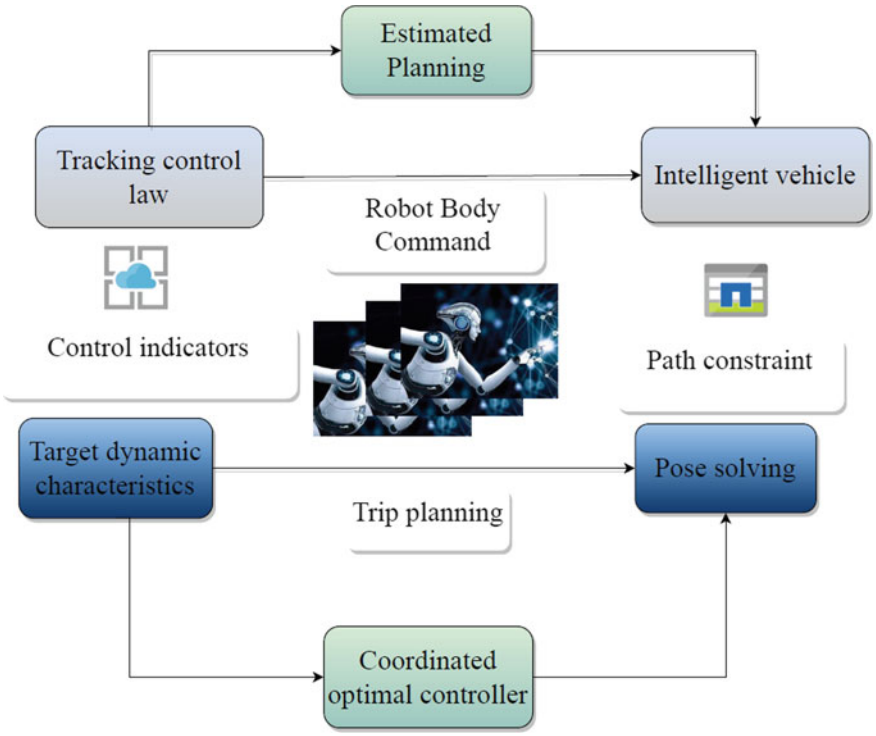


Fig. 4.2 Trajectory optimization control system

optimal time is sought by the fitness function [10]. The particle swarm adjusts the next evolutionary direction according to the individual optimal position and the global optimal position, but when it is local optimal, all particles will be influenced by it and move toward the local optimal position, which will lead to the rapid convergence of the particle swarm, resulting in local extremum or stagnation.

4.2.2 Optimization Goal Establishment

The entire particle swarm is a large whole, treating each particle as a separate small individual. Each particle randomly finds a location that may be the best within the feasible domain. The target point is determined as the global optimal solution, and the motion process of the particle swarm can be seen as the iterative evolution process of the algorithm. In particle swarm optimization, each particle exists as an independent individual, representing an optimization mechanism that seeks the optimal solution through mutual cooperation between particles. The PSO algorithm is an intelligent optimization algorithm that is very simple, easy to understand and does not require

complex calculations or excessive parameter adjustments. Its robustness is stable. Particle swarm optimization (PSO) algorithm is prone to falling into the defects of extreme values and declining particle swarm diversity during the iteration process for improvement, mainly through parameter improvement, adjusting particle states and integrating with other algorithms. Scholars often make improvements based on inertia weight coefficients.

When the intelligent robot is working, whether it meets the requirements is mainly measured by running time, energy consumption and pulsating impact. According to different iterative stages, the learning factor is dynamically adjusted to prevent the particle swarm from falling into local optimum in the initial rapid aggregation stage, and to ensure that the global optimal solution can be quickly found in the later search stage while searching in a large range [11]. Aiming at the optimization of intelligent robot's time pulse impact, the time-based objective function S_1 and the pulse impact-based objective S_2 are constructed:

$$\min S = T = \sum_{i=0}^{n-1} (t_{i+1} - t) \tag{4.1}$$

$$S_2 = \sum_{n=1}^{N} \sqrt{J_i^2 \mathrm{d}t} \tag{4.2}$$

Among them: T stands for the total running time of intelligent robot, which can be used to measure the running efficiency; the J represents the pulsating impact and measures the smoothness of the robot trajectory.

The objective function is normalized by setting the weight coefficient, and the fitness function is defined as:

$$f = \alpha_1 S_1 + \alpha_2 \beta S_2 \tag{4.3}$$

where α_1 and α_2 are weight coefficients and satisfy $\alpha_1 + \alpha_2 = 1$; β is an elastic adjustment factor, and β exists to balance the difference in order of magnitude between pulsating impact and action time.

In order to perform time optimal trajectory planning for articulated industrial robots, it is necessary to optimize particle parameters and find the optimal solutions for each particle, in order to obtain the optimal solution for the entire optimization process. The optimal value of the particle swarm has a significant impact on the behavior of the entire particle swarm. The idea of simulated annealing is used to select the optimal solution of the particle swarm [12, 13]. In the early stages of optimization, it is acceptable to accept inferior particles as particle swarm particles, maintaining particle diversity. In the later stage of optimization, only high-quality particles are accepted as the optimal particles, so that the optimization converges to the optimal value.

4.2.3 Particle Velocity and Position Updates

PSO is an algorithm based on traditional PSO, but the number of objects to be solved is different. After the target particle swarm is initialized, the particle characteristics in the particle swarm are represented by position, speed and fitness. If the robot needs to accurately track a given trajectory, then the controller must control the manipulator's executive end to move along this trajectory. However, because the robot is highly nonlinear and has complex constraints, Coriolis force and centrifugal force have great influence on the movement of the whole system at high speed. It is generally impossible for a robot to accurately track a given trajectory. So most of the trajectory tracking is realized by approximation. In order to ensure that the robot's end gripper can reach Cartesian space coordinates, the robot is solved by forward kinematics and the space coordinate position of its starting point and the space coordinate position of its ending point are obtained. Compared with the different optimal individual fitness produced by each iterative update, the global optimal solution G_{best} is determined, and the iterative expressions of the velocity and position of particle i are as follows:

$$\upsilon_i^{k+1} = \varphi_1 R_1\left(P - X_i^k\right) + \varphi_2 R_2\left(G - X_i^k\right) \tag{4.4}$$

Among them, υ_i^k represents the velocity of the i particle at the k iteration; X_i^k is the position after the k iteration. R_1 and R_2 are random numbers between (0, 1). P_1 and P_2 represent the optimal positions for particle fitness.

The selection of particle swarm fitness function is the key to the optimization results. Due to the existence of many different evaluation systems and performance indicators in the control process, such as stability, controllability, convergence speed, steady-state characteristics and dynamic characteristics. For robots with joints, corresponding trajectories are set in joint space. If more points are selected, the distance between adjacent two points will be smaller and the accuracy will be higher, but this also leads to the need for more inverse operations. The lines and arcs implemented in this section are implemented in this way. Therefore, different emphasis points in the optimization process will be reflected in the different characteristics presented on the tracking trajectory.

4.3 Simulation Experiment and Result Analysis

4.3.1 Simulation Conditions and Parameter Settings

In this paper, the improved particle swarm optimization algorithm, the standard particle swarm optimization algorithm, the inertia weight PSO algorithm and the compressibility factor PSO algorithm are used to carry out path planning simulation experiments through MATLAB. Due to multiple robots performing path planning

tasks in the same operating environment, there may be resource grabbing and collaborative cooperation between robots. Therefore, we need to allocate tasks reasonably, as well as prioritize node interaction, evaluation and decision-making.

During the experiment, we need to consider not only the conditions set in the particle swarm optimization algorithm, but also the normal conditions of industrial robots, which should not exceed the normal range of robots. PSO is an algorithm to ensure global search. In order to further improve the global search ability of the algorithm, improve the search performance for solving multimodal optimization problems and improve the ability of searching in a large range in the later stage of the algorithm. In this paper, the operation strategy of randomly selecting the best individual in the group is put forward to avoid the diversity of the group being too small, so as to improve the global search ability of the algorithm.

In each iteration, the particle itself will constantly adjust by tracking two extreme values. In the optimization process, firstly, the global optimal particle is obtained from twenty groups of particles by comparing fitness values. The particles are continuously updated through n iterations, and the optimal solution particle with constraints is calculated, which is the optimal time particle. So the optimal solution found from these particle neighbors is called local extremum. In the process of particle optimization, it is impossible for particles to always get a better solution only by tracking the global extremum or local extremum. The problem is solved satisfactorily when particles follow the global extreme value or local extreme value while following the individual extreme value.

4.3.2 Simulation Result

Starting from the active motion of the robot, an improved particle swarm optimization is used to solve the optimal detection point of the robot, so as to make the location and tracking of the trajectory target of the intelligent robot more accurate. In theory, the bottleneck of inaccurate distance measurement in existing positioning has been overcome, effectively solving the problem of inaccurate distance measurement in trajectory positioning accuracy. In this paper, the most appropriate weight coefficient will be selected according to the actual engineering needs to minimize the fitness function, so as to achieve the optimal solution of both. If collision needs to be avoided in the motion space of the robot, the best way is to first plan a trajectory that avoids obstacles. This trajectory is usually preplanned and stored in memory through teaching or calculation, and only needs to track the motion of this path during operation. But when the robot reaches its working position, the path of obstacles to avoid becomes short and difficult to control. This makes it very difficult to obtain these trajectories through teaching or calculation, especially in situations where the execution time is very short.

To demonstrate the stability of PSO, taking Joint 3 as an example, multiple optimizations were conducted before and after optimization. The optimization time convergence diagram based on PSO is shown in Figs. 4.3, 4.4 and 4.5. In the early

stage of the experimental process, the particle swarm can search for the optimal value in a short time, while also ensuring fast and accurate convergence to the optimal solution in the later stage.

Analyzing the data in Figs. 4.3, 4.4 and 4.5, it can be seen that there is a significant difference between before and after optimization. In the time convergence results of Joint 1, it takes 7 s to complete before optimization, and only about 4 s to complete after optimization. In the time convergence results of Joint 2, it takes 9 s to complete before optimization, and only about 5 s to complete after optimization. In the time

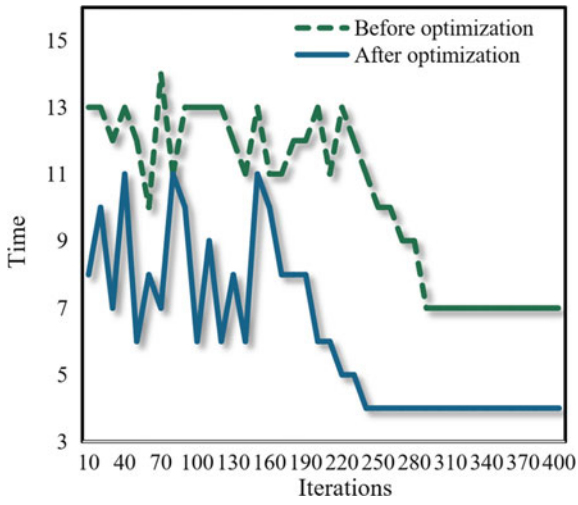


Fig. 4.3 Time convergence graph of Joint 1

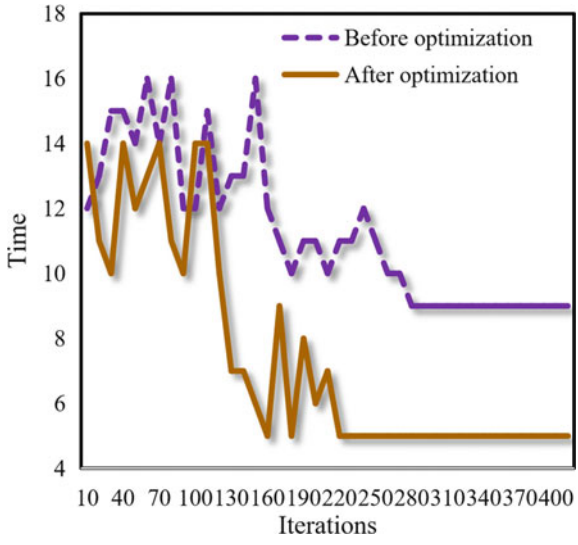


Fig. 4.4 Time convergence graph of Joint 2

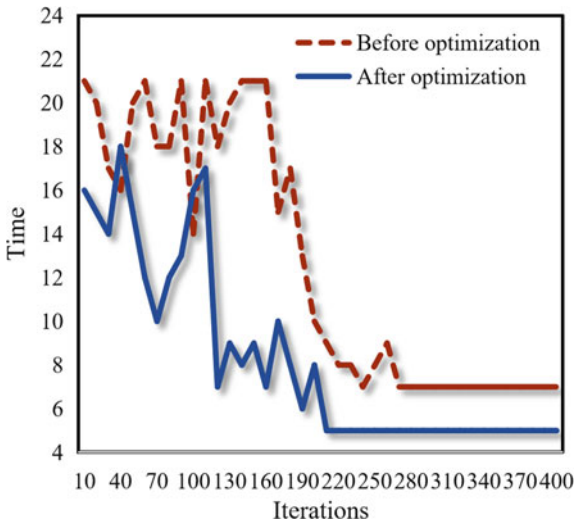


Fig. 4.5 Time convergence diagram of Joint 3

convergence results of Joint 3, it takes 7 s to complete before optimization, and only about 5 s to complete after optimization. Through the experimental data of the above three joints, it can be found that the optimized trajectory motion time has been reduced by an average of about 50%, achieving the expected goal.

4.4 Conclusions

In this paper, the trajectory of intelligent robot constructed by PSO is proposed. Combined with polynomial programming method, the optimal design of robot's time and pulse impact is carried out by improving hybrid PSO, which effectively shortens the movement time and has good stationarity, and effectively improves the robot's working performance. When the intelligent robot is working, whether it meets the requirements is mainly measured by running time, energy consumption and pulsating impact. According to different iterative stages, the learning factor is dynamically adjusted, so that the particle swarm can avoid falling into local optimum in the initial rapid aggregation stage, and the global optimal solution can be quickly found in the later search stage while searching in a large range. When designing the path, the execution end of the robot is often regarded as a point. But in fact, the size of the robot's hand is relatively large, and it moves in space with a certain volume during its movement, not a line. To avoid collision, it is necessary to ensure that the volume corresponding to this space trajectory does not intersect with obstacles. The results show that through the experimental data of three joints, it can be found that the optimized trajectory motion time is shortened by about 50% on average,

which achieves the expected purpose. The results show that compared with before optimization, PSO has stronger global search ability, faster convergence speed and good stability, which can effectively improve the working efficiency of the robot and maintain the running stability.

Acknowledgements National Key R&D Program Funded Project "Technology for Trading Property Rights of Scientific and Technological Achievements and Data Resources" (2017YFB1401100).

References

1. Zhou, L., Chen, K., Dong, H., Chi, S.K., Chen, Z.: An improved beetle swarm optimization algorithm for the intelligent navigation control of autonomous sailing robots. IEEE Access **9**, 5296–5311 (2020)
2. Yan, M., Yuan, H., Xu, J., Yu, Y., Jin, L.B.: Task allocation and route planning of multiple UAVs in a marine environment based on an improved particle swarm optimization algorithm. EURASIP J. Adv. Signal Process. **2021**, 94 (2021)
3. Ma, T., Liu, S., Xiao, H.: Location of natural gas leakage sources on offshore platform by a multi-robot system using particle swarm optimization algorithm. J. Nat. Gas Sci. Eng. **84**, 103636 (2020)
4. Xu, L., Song, B., Cao, M.: An improved particle swarm optimization algorithm with adaptive weighted delay velocity. Syst. Sci. Control Eng. **9**(1), 188–197 (2021)
5. Lei, J.H., Wan, B., Liu, M., et al.: The design of four-rotor human-simulated intelligent controller based on particle swarm optimization algorithm. Comput. Simul. **46**(14), 16–23 (2022)
6. Gao, Q.: Intelligent vehicle global path planning based on improved particle swarm optimization. Open Access Libr. J. **5**, e4587 (2018)
7. Mahmoodabadi, M.J., Taherkhorsandi, M.: Intelligent control of biped robots: optimal fuzzy tracking control via multi-objective particle swarm optimization and genetic algorithms. AUT J. Mech. Eng. **4**(2), 183–192 (2020)
8. Tao, Q., Sang, H., Guo, H., Wang, P.: Improved particle swarm optimization algorithm for AGV path planning. IEEE Access **9**, 33522–33531 (2021)
9. Wu, Q., Zhu, Z., Yan, X., Gong, W.Y.: An improved particle swarm optimization algorithm for AVO elastic parameter inversion problem. Concurr. Comput. Pract. Exp. **31**(9), e4987 (2019)
10. Selma, B., Chouraqui, S., Abouassa, H.: Optimal trajectory tracking control of unmanned aerial vehicle using ANFIS-IPSO system. Int. J. Inf. Technol. **12**, 383–395 (2020)
11. Guo, Y., Fang, X., Dong, Z., Mi, H.L.: Research on multi-sensor information fusion and intelligent optimization algorithm and related topics of mobile robots. EURASIP J. Adv. Signal Process. **2021**, 111 (2021)
12. Yu, Y., Xu, Y., Wang, F., Li, W.S., Mai, X.M., Wu, H.: Adsorption control of a pipeline robot based on improved PSO algorithm. Complex Intell. Syst. **7**, 1797–1803 (2021)
13. Wang, M.: Real-time path optimization of mobile robots based on improved genetic algorithm. Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng. **235**(5), 646–651 (2021)

Chapter 5
Design of Remote Sensing Image Processing Algorithm Based on Machine Learning

Shuying Liu

Abstract Remote sensing (RS) image classification is one of the most basic problems in RS image information processing. Its classification technology is the key technology in RS application system. In the practical application of RS image classification processing, a large number of training data are needed to obtain high-precision classification results, marking these training data requires a lot of manpower and material resources, and it is also time-consuming. Traditional hyperspectral image processing methods can only extract the shallow features of the image, but ignore the deep features of the image. In this paper, an RS image processing algorithm based on deep learning (DL) is proposed, which can fully mine the useful information of a large number of unlabeled samples in RS image classification, so as to expand a small number of labeled samples, enhance the classifier's ability, and improve the classification accuracy. Experiments show that the RS image processing algorithm in this paper has high efficiency and short response time, with an average of about 0.3 s; at the same time, the root mean square error (RMSE) of the algorithm is low, which can be stabilized at about 0.523. The research in this paper is of great practical value and practical significance for improving the classification accuracy of RS images with few labeled training samples.

5.1 Introduction

The sensing system carried on the satellite can obtain large-scale data in a short time, and the information is presented in the form of pictures or non-circular films; at the same time, it can replace human beings to go to places that are difficult to reach or in bad conditions for in-depth observation [1]. Hyperspectral RS usually adopts imaging spectrometer or non-imaging spectrometer, which covers a fixed band, and emits a large number of narrow-band electromagnetic waves from a long distance

S. Liu (✉)
Department of Computer, Xianyang Normal University, Xianyang 712000, China
e-mail: liushuying2009@163.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_5

to ground objects in a non-contact way, and obtains hyperspectral images through the interaction between electromagnetic waves and ground objects [2]. RS digital processing technology properly uses these characteristics to interpret the characteristics of the object obtained by the detector, then obtains the physical properties and spectral characteristics of the object, and effectively processes the image information digitally [3]. Different from other RS images, hyperspectral images have many bands, high spectral resolution, and rich ground information. The pixel values of different bands in the same position on the image form a vector, which represents the spectral information values of the pixel in different bands [4]. Because different ground objects have different reflection wavelengths, their corresponding spectral information is also different. Manual calibration of RS images is time-consuming and laborious, so it is necessary to use machine learning (ML) to automatically process images [5]. RS image classification is one of the most basic problems in RS image information processing. Its classification technology is the key technology in RS application system. The accuracy of RS image classification method directly affects the application and development of RS. The classification of RS image data is essentially a special problem of data processing and pattern recognition of spatial data. Usually, the ideas and methods corresponding to data processing and pattern recognition can be applied to the classification of RS image data [6]. In practical application, the spectral characteristics obtained by various sensors will be affected by scattering and atmospheric effects, which will make the observed spectral reflection value different from the theoretical measurement value [7]. A feasible way to solve this problem is to add spectral derivative characteristics to the spectral characteristics. Inspired by the structure of human brain, DL constructs a neural network model with multiple nonlinear mapping layers to extract data features. Compared with the traditional feature extraction method [8]. It does not depend on manual design and can automatically extract different levels of features, including low-level features such as lines and textures and complex and abstract high-level features. The extracted features have very strong generalization ability [9]. RS image classification is one of the most critical technologies in digital processing of RS images, and fast and high-precision RS image classification algorithm is the premise to realize various practical applications. In this paper, an RS image processing algorithm based on DL is proposed, which can fully mine the useful information of a large number of unlabeled samples in RS image classification, so as to expand a small number of labeled samples, enhance the ability of classifier, and improve the classification accuracy.

5.2 Methodology

5.2.1 Characterization of RS Data Characteristics

For RS images, spectral features and texture features are regarded as two views, respectively. However, for a small training sample, the feature attributes of the connection may be over-fitted, and because each group of features has a special statistical characteristic. As a new method, multi-view learning makes full use of redundant views of the same set of data to improve learning performance. Therefore, at present, multi-view has become a research hotspot. The key of ensemble learning is how to design a base classifier with strong generalization ability and great individual differences to form a multi-classification system. The accuracy of base classifiers and their diversity are two important factors that affect the integrated classification system [10]. The extraction of RS image texture features is to extract the feature quantity representing each pixel or spatial distribution from the image and transform the difference of texture spatial structure into the difference of feature space. Active learning independently selects the unlabeled samples with the richest information, requests expert labeling, adds these samples to the existing training data set, trains the classifier, and selects unlabeled samples again. Repeated iterations like this can maximize the accuracy of supervised learning in classifying unlabeled samples.

Traditional ML algorithms, such as support vector machine, discriminant analysis, and neural network, all connect multiple features into an independent set of feature vectors to train and learn. That is, a learner uses a set of feature sets as a training ensemble [11]. In practical application, different features of the research object can be obtained from different angles, and the feature set obtained from each angle or method of extracting features is regarded as a view, which constitutes a multi-view feature domain. Texture reflects some changes in surface color and gray level of an object. It not only represents the statistical information of image gray level, but also reflects the spatial distribution information and structural information of the image, which can better pay attention to the macro-structure and micro-structure of the image. Training samples are obtained by using labeled samples to train the supervised classifier. In each learning iteration, the classifier actively selects the most informative samples from unlabeled data according to the sampling strategy, submits them to experts for marking, and then adds them to the training set for the next iteration.

Hyperspectral imaging technology is an imaging set in multi-wavelength bands, that is, for the same position in space, imaging at different wavelengths will return a relative reflection intensity value, so for spectral imaging pictures, each pixel needs to be represented by a spectral feature vector, and each element in the vector represents the reflection intensity of the corresponding band. Consistent with semi-supervised classification, active learning also uses unlabeled samples to mine "valuable" samples and uses them to assist a small number of labeled training sets to improve the classification performance. In active learning, the classifier actively selects unlabeled samples with a large amount of information, adds them to the training set after

labeling, and updates the classifier, so as to obtain higher classification accuracy when there are fewer labeled samples.

5.2.2 RS Image Processing Algorithm

Generally speaking, there are many unrelated attributes in high-dimensional data, and the distances between the data are similar, which makes these data have no clusters at all in the full-dimensional space, so the traditional clustering method based on distance cannot be used to construct clusters. Aiming at the above problems, a subspace clustering algorithm combining feature selection and clustering is proposed [12]. In practical application, the research object can obtain descriptive information from multiple angles, thus forming different views. However, due to some limitations, it is impossible to get multiple views of research data directly. Usually, there is only one view that represents data. Firstly, the convolutional encoder is used to encode the input data and get its deep-seated potential feature representation. In order to make better use of shallow features such as edges and textures, the obtained shallow features and deep features reflecting shape categories are fused across layers, so as to enhance the performance of the model for complex images and get a feature map close to the input data more accurately and quickly. Finally, the parameters in the pre-training model are used to initialize the improved deep subspace clustering network and fine-tune the whole network. The flow of RS image enhancement algorithm is shown in Fig. 5.1.

In the process of RS image classification, there is no classifier that is optimal for all classification problems, and it is difficult to find the optimal classifier [13]. The existing classification algorithms are combined by appropriate methods, and the complementarity between different classifiers is used to improve the classification accuracy of RS images. In the process of automatic classification of RS data, all kinds of image distortions and unnecessary absorption particles have been corrected and removed in advance, which lays the foundation for correct classification of images.

The standardized optimal index set is used as the reference data column, and the standardized index value $(y_{i1}, y_{i2}, y_{i3}, \ldots, y_{im})$ $(i = 1, 2, 3, \ldots, n)$ is used as the compared data column. Then use the following formula to calculate the grey correlation coefficient:

$$\delta_i(j) = \frac{\min_i \min_j \left|s_j - y_{ij}\right| + \rho \max_i \max_j \left|s_j - y_{ij}\right|}{\left|s_j - y_{ij}\right| + \rho \max_i \max_j \left|s_j - y_{ij}\right|} \tag{5.1}$$

where $\delta_i(j)$ is the correlation coefficient between the j index of the i sample and the j optimal index value in the optimal index set; ρ is the resolution coefficient, which is generally taken as 0.5. So the grey assessment matrix is obtained:

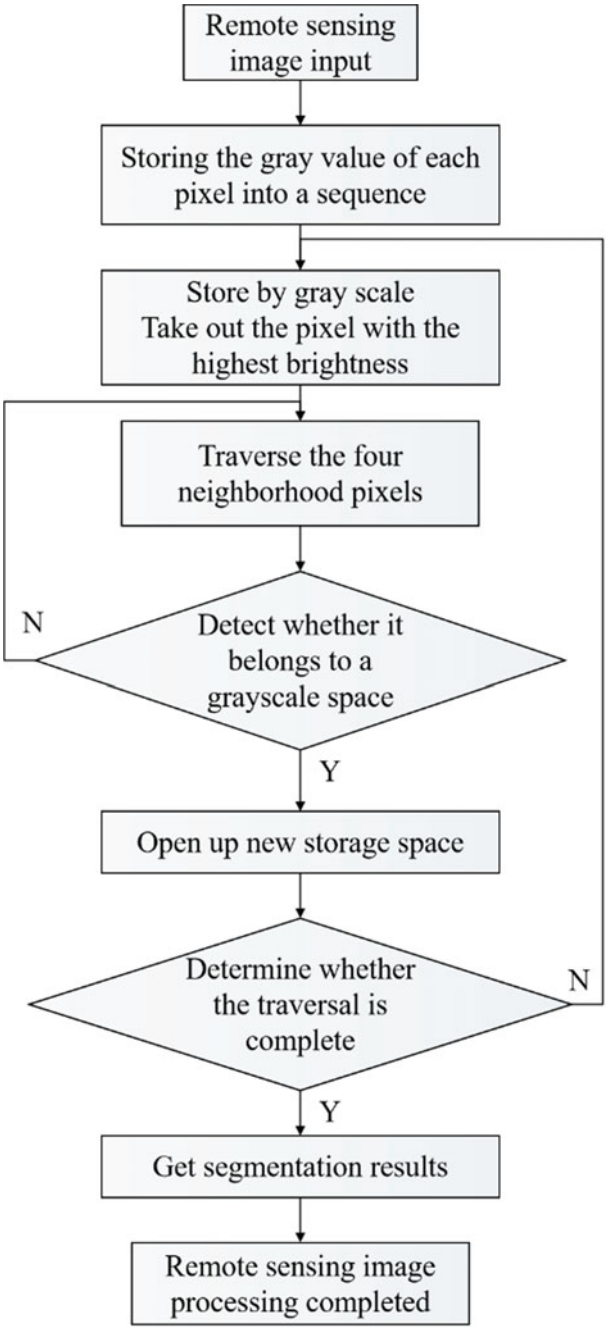


Fig. 5.1 Flowchart of RS image enhancement algorithm

$$E = \begin{bmatrix} \delta_1(1) & \delta_1(2) & \cdots & \delta_1(m) \\ \delta_2(1) & \delta_2(2) & \cdots & \delta_2(m) \\ \cdots & \cdots & \cdots & \cdots \\ \delta_n(1) & \delta_n(2) & \cdots & \delta_n(m) \end{bmatrix} \tag{5.2}$$

The correlation degree between comparison series and reference series is expressed by correlation degree, and the calculation formula is as follows:

$$\gamma_{0i} = \frac{1}{n} \sum_{k=1}^{n} \gamma_{0i}(k) \tag{5.3}$$

where $\gamma_{0i}(k)$ is the correlation degree, that is, the average of the correlation coefficients of the same factor.

The preliminary task of multi-view learning is to consider how to build multi-views that meet the needs from a single view. The characteristic of multi-view is to include redundant views. The construction of multi-view should not only get views with different attribute characteristics, but also consider that views can fully represent the research data and meet the conditions of ML assumption.

Given the original remote sensing image sample data set:

$$X = \{x_1, x_2, \ldots, x_N\} \tag{5.4}$$

x_i is the image sample set of class i, and x_{ij} is the j sample of class i. The expressions of the inter-class similarity measure function S_{B} and the intra-class similarity measure function S_{W} of the sample set are as follows:

$$S_{\mathrm{B}} = \sum_{i=1}^{c} p_i (m_i - m)(m_i - m)^{\mathrm{T}} \tag{5.5}$$

$$S_{\mathrm{W}} = \sum_{i=1}^{c} p_i \sum_{j=1}^{n_i} \left(x_{ij} - m_i\right)\left(x_{ij} - m_i\right)^{\mathrm{T}} \tag{5.6}$$

p_i represents the prior probability of class i samples, m_i is the class average of class i samples, m represents the average of the total sample set, and n_i is the number of class i samples.

Aiming at the classification task, ensemble learning obtains a group of classifiers through learning and combines these classifiers through some rules to obtain a ML algorithm with better classification effect than a single classifier. In order to obtain better classification results, it is required that the base classifiers participating in the integration are independent of each other. Although there are no acceptable formal methods and measures for the differences in ensemble classifiers, some effective heuristic mechanisms can be used to generate the differences in ensemble models when constructing ensemble classifiers.

5.3 Result Analysis and Discussion

In the integrated system based on neural network, random disturbance can be introduced into the learning process of each base classifier, and different base classifiers can be trained [14]. If the base classifier is sensitive to random changes, this method can effectively generate multiple different base classifiers. The attributes of training data are described by a set of features. Different feature subspaces constitute different training data views. According to the sample data in different views, the classifiers are trained and established, respectively, so as to obtain multiple classifier sets with differences. In this paper, three different hyperspectral remote sensing data sets (Botswana, KSC, Pavia University) are used for experimental analysis. The generation of base classifier mainly produces different data subsets by processing the original data set and trains the different base classifier on the different data subsets. Configure the grid platform through a software package provided by a software toolbox developed by the alliance. The selection of base classifiers means that a series of generated base classifiers will choose the best one to participate in the integrated classification according to a certain metric or strategy. Figure 5.2 shows the variation of data error with the noise in the training set.

Parallel classifiers combine the results of each base classifier at the same time. The key of this structure is to choose a representative result combination method. If the decision-making process is designed reasonably, the system can achieve a better performance. However, improper selection of decision-making combination

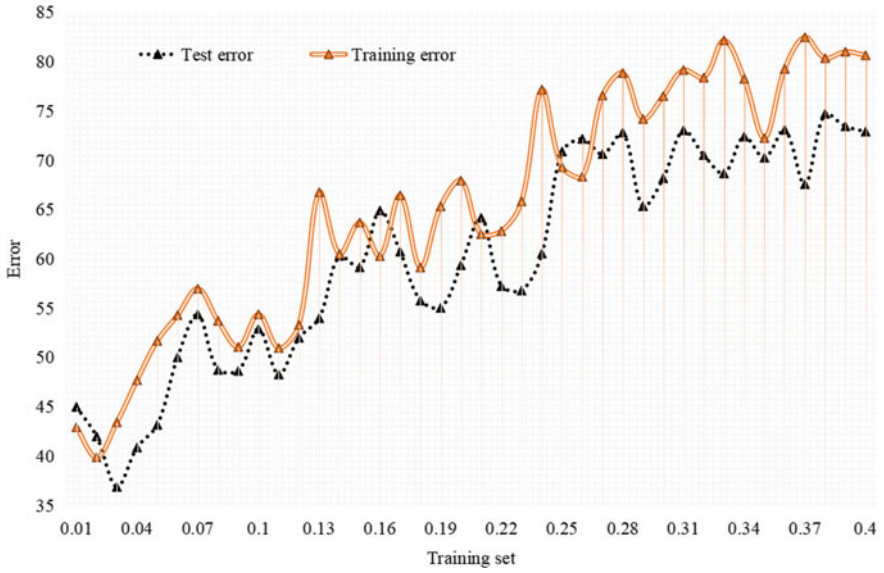


Fig. 5.2 Variation of data error with noise in training set

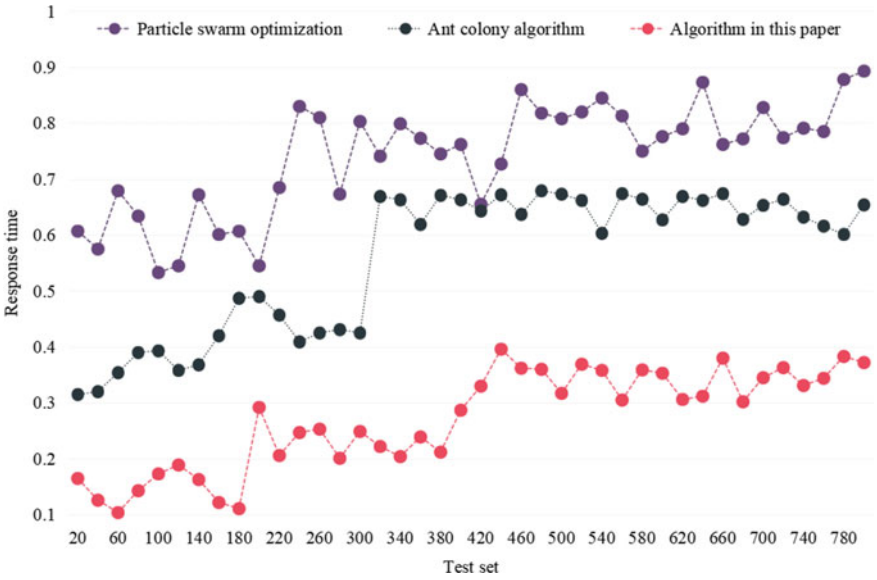


Fig. 5.3 Algorithm efficiency testing

methods may increase the influence of weak classifiers. The efficiency of RS image data processing is tested by the algorithm, and the result shown in Fig. 5.3 is obtained.

The experimental results show that the efficiency of the RS image processing algorithm in this paper is high, and the response time is short, with an average of about 0.3 s, which shows that its efficiency is at an excellent level.

In the training stage, the samples are submitted to each base classifier, and the obtained results reach a better training process after combined decision. The process of building a multi-decision is similar to that of building a multi-classifier topology. Constructing a reasonable topological structure can adjust the accuracy and error checking of the classification system. The RMSE test of the algorithm is shown in Fig. 5.4.

The results show that the RMSE of the RS image processing algorithm in this paper is low and can be stabilized at about 0.523. This shows that the algorithm has better performance. When the initialization of the algorithm is completed, the initial classification model is completed by training the training points in the original domain, and then the initial output space division and input space division are obtained. When the high confidence semi-marker points are selected and some training points in the original domain are deleted, the new training set will be used to retrain and adjust the original classification model to make it more consistent with the data distribution in the target domain.

The classification result of each classifier is passed to the next classifier step by step until the last classifier gets the final classification result. The main disadvantage of this structure is that the later classifier in this cascade chain is unable to correct

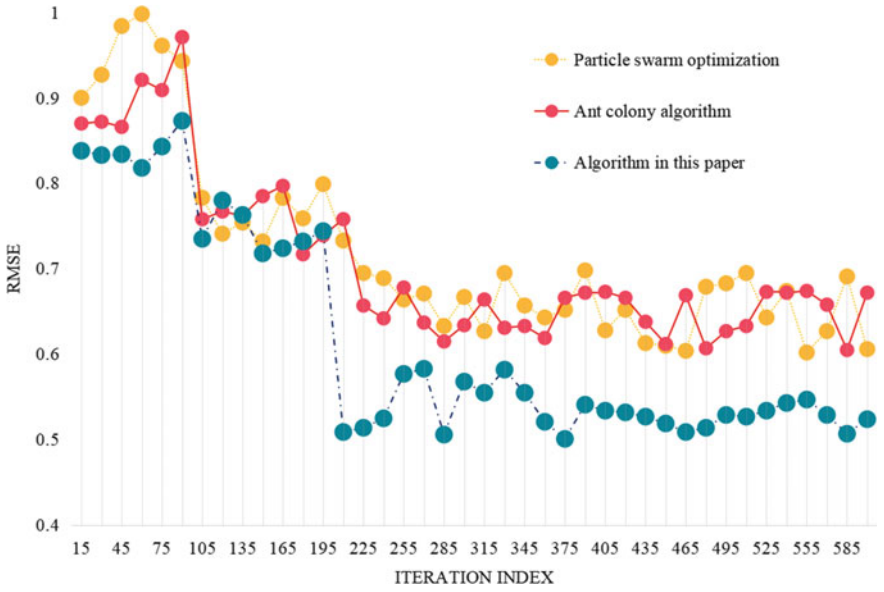


Fig. 5.4 RMSE test results of the algorithm

the wrong classification judgment of the previous classifier, and this error will affect the performance of the next classifier. In order to test whether the performance of RS image processing algorithm is improved, the performance of three different algorithms is compared, and the accuracy results are shown in Fig. 5.5.

In order to verify the effectiveness of the method proposed in this paper, this section carries out several tests. The experimental results show that the RS image processing algorithm based on ML has high efficiency and short response time, with an average of about 0.3 s: At the same time, the accuracy of the algorithm is high, the RMSE is low, and the RMSE can be stabilized at about 0.523. All the above results show that the proposed RS image processing algorithm has good performance and can better realize the mining and processing of RS image data.

5.4 Conclusions

The classification of RS image data is essentially a special problem of data processing and pattern recognition of spatial data. Usually, the ideas and methods corresponding to data processing and pattern recognition can be applied to the classification of RS image data. In this paper, an RS image processing algorithm based on DL is proposed, which can fully mine the useful information of a large number of unlabeled samples in RS image classification, so as to expand a small number of labeled samples, enhance the ability of classifier, and improve the classification accuracy. The experimental

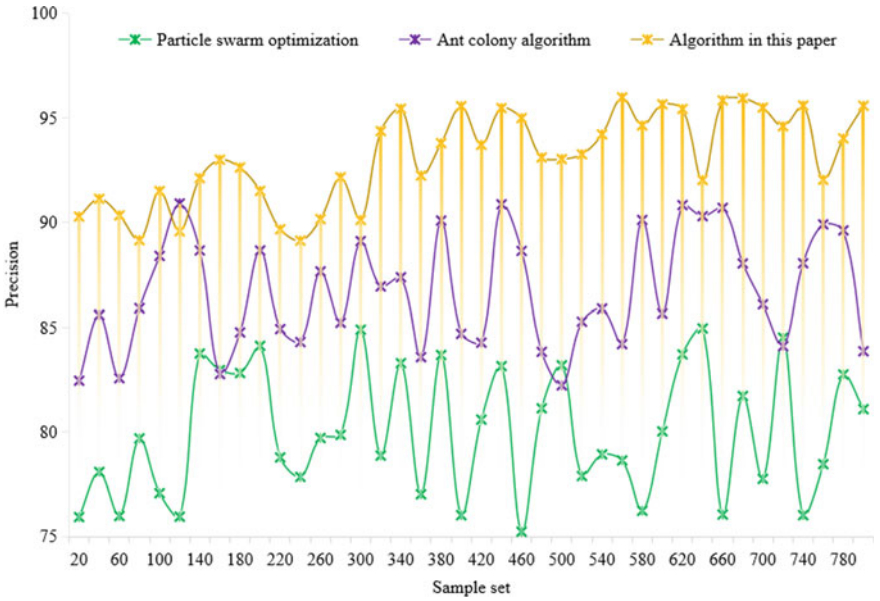


Fig. 5.5 Performance comparison of different algorithms

results show that the RS image processing algorithm based on ML has high efficiency and short response time, with an average of about 0.3 s: At the same time, the RMSE of the algorithm is low, which can be stabilized at about 0.523. This shows that the performance of the algorithm is excellent, which can better realize the mining and processing of RS image data and provide convenient services for related work. In the future application of ecological RS, it is necessary to be familiar with the application of ML in all kinds of problems involved and master the frontier technology of science and technology in time, which can benefit from every revolutionary breakthrough of information technology to the greatest extent and promote the innovation of RS application mode with ML, especially DL technology.

Acknowledgements (1) Foundation Item: Education and Teaching Reform Research Foundation of Xianyang Normal University (2019Y014); (2) Foundation Item: Xianyang Development Research Institute 2020 Annual Project (2020XFZ009).

References

1. Guo, M., Wang, S.: Remote sensing monitoring and ecological risk assessment of landscape patterning in the agro-pastoral ecotone of Northeast China. Complexity **2021**(2), 1–13 (2021)
2. Zhu, Z., Luo, Y., Qi, G., et al.: Remote sensing image defogging networks based on dual self-attention boost residual octave convolution. Remote Sens. **13**(16), 3104 (2021)

3. Zotin, A.G.: Fast algorithm of image enhancement based on multi-scale retinex. Int. J. Reason.-Based Intell. Syst. **12**(2), 106 (2020)
4. Liao, C., Hu, H., Li, H., et al.: Joint learning of contour and structure for boundary-preserved building extraction. Remote Sens. **13**(6), 1049 (2021)
5. Sun, L., Tang, C., Xu, M., et al.: Non-uniform illumination correction based on multi-scale Retinex in digital image correlation. Appl. Opt. **60**(19), 5599–5609 (2021)
6. Zhang, H., Wang, M., Wang, F., et al.: A novel squeeze-and-excitation W-Net for 2D and 3D building change detection with multi-source and multi-feature remote sensing data. Remote Sens. **13**(3), 440 (2021)
7. Hai, H., Li, P., Zou, N., et al.: End-to-end super-resolution for remote-sensing images using an improved multi-scale residual network. Remote Sens. **13**(4), 666 (2021)
8. Li, X., Zhang, H., Yu, J., et al.: Spatial–temporal analysis of urban ecological comfort index derived from remote sensing data: a case study of Hefei, China. J. Appl. Remote Sens. **15**(4), 042403 (2021)
9. Li, X., Long, J., Zhang, M., et al.: Coniferous plantations growing stock volume estimation using advanced remote sensing algorithms and various fused data. Remote Sens. **13**(17), 3468 (2021)
10. Han, Y., Liu, Y., Hong, Z., et al.: Sea ice image classification based on heterogeneous data fusion and deep learning. Remote Sens. **13**(4), 592 (2021)
11. Ren, J., Yang, W., Yang, X., et al.: Optimization of fusion method for GF-2 satellite remote sensing images based on the classification effect. Earth Sci. Res. J. **23**(2), 163–169 (2019)
12. Chen, D.Y., Peng, L., Li, W.C., et al.: Building extraction and number statistics in WUI areas based on UNet structure and ensemble learning. Remote Sens. **13**(6), 1172 (2021)
13. Carter, S., Herold, M., Jonckheere, I., et al.: Capacity development for use of remote sensing for REDD+ MRV using online and offline activities: impacts and lessons learned. Remote Sens. **13**(11), 2172 (2021)
14. Sepasgozar, S., Hui, F., Shirowzhan, S., et al.: Lean practices using building information modeling (BIM) and digital twinning for sustainable construction. Sustainability **13**(1), 161 (2020)

Chapter 6
Design of Hovering Orbit and Fuel Consumption Analysis for Spacecraft Considering J_2 Perturbation

Liang Zhang, Tao Song, Hao Ding, and Honghao Liu

Abstract Spacecraft hovering belongs to close range space operations and is mainly applied to on-orbit service of spacecraft. In on-orbit service, the distance between two hovering spacecraft is generally within the range of 0–10 m, whose numerical magnitude is relatively small compared to the orbital radius, so the influence of space perturbation must be considered. In order to improve the accuracy of hovering position, this paper introduces J_2 perturbation into the relative dynamics model of the spacecraft and derives high-precision hovering control equations. Furthermore, the influence of eccentricity and semimajor axis on the hovering control is obtained through an example. Then, using the fuel consumption calculation formula, the distribution of the fuel consumption of the hovering spacecraft is given, when the spacecraft is hovering at different positions in a fixed orbital period. Considering the limited carrying capacity of fuel, the fuel consumption at different hovering positions was analyzed for the same hovering distance, and the problem of determining the hovering position was solved to reduce fuel consumption.

6.1 Introduction

With the continuous advancement of people's exploration of space, it is necessary to extend the lifespan of spacecraft to ensure that they can perform space missions more stably and continuously in unknown space. On-orbit Service (OOS) technology is mainly applied to the maintenance, repair, and upgrade of spacecraft in operation, in order to extend the service life of the spacecraft [1–3]. Spacecraft hovering is a formation configuration with relatively stationary positions, and this fixed state characteristic can provide a stable working environment for on-orbit service, so that spacecraft can successfully complete space operations. Therefore, the control of hovering orbit is the key to achieving the design of spacecraft hovering orbit.

L. Zhang (✉) · T. Song · H. Ding · H. Liu
College of Mechanical and Architectural Engineering, Taishan University, Tai'an 271000, Shandong, China
e-mail: zhang_liang473@126.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_6

The study of spacecraft hovering orbits originated from the exploration of small celestial bodies. Scheeres conducted research on the hovering of spacecraft relative to spinning asteroids [4]. Based on the physical characteristics of small celestial bodies, Broschart and Scheeres defined two concepts of spacecraft hovering small celestial bodies [5]. Lu and Love proposed that the gravitational interaction between Earth and asteroids can enable spacecraft to orbit at a fixed position relative to the asteroid [6].

When the hovering target is a spacecraft, the hovering orbit is divided into circular orbit and elliptical orbit based on different operation orbits. From the perspective of dynamic modeling, Wang et al. [7] deduced the expression of control for the mission spacecraft to achieve hovering at a given position in an elliptical orbit of the target spacecraft. Zhang et al. [8] conducted in-depth research on the impact of different orbit parameters on velocity increment in hovering and explored the hovering feasibility without applying control within certain special parameter variation ranges. In view of the limited fuel carried by spacecraft, a hovering method for electrically charged spacecraft using the hybrid propulsion with conventional chemical propulsion and Lorentz force is proposed [9]. Considering the situation of spacecraft thruster failure, Huang et al. [10] established a dynamic model of underactuated hovering orbit and conducted a detailed analysis of the controllability of the system under underactuated conditions. Furthermore, Huang and Yan [11] proposed an adaptive reduced order observer for speed and parameter estimation in response to disturbance mismatch during underactuated conditions. Huang and Yan [12] designed a backstepping controller to obtain feasible hovering positions under saturated underactuated conditions.

Unlike the previous modeling methods for dealing with J_2 disturbances in spacecraft relative motion, this paper introduces J_2 perturbation as a known term into the hovering orbit dynamics model and obtains the corresponding spacecraft hovering control considering J_2 perturbation. Based on this, the variation law of hovering control under different orbit eccentricity and semimajor axis is obtained. Then, the fuel consumption formula is used to analyze the fuel consumption distribution of different hovering positions for the mission spacecraft in a fixed orbital period, and the selection method for determining the hovering position at the same distance is summarized.

6.2 Hovering Control Considering J_2 Perturbation

In Fig. 6.1, *O-XYZ* represents ECI (Earth Center Inertia) frame, and *M-xyz* represents the LVLH (Local Vertical Local Horizontal) frame. In *M-xyz*, *M* is the center of LVLH, *x* is in the radial direction, *z* is perpendicular to the plane of the target spacecraft's orbit, and *y* constitutes a Cartesian coordinate system.

The relative dynamics equation is expressed in *M-xyz*.

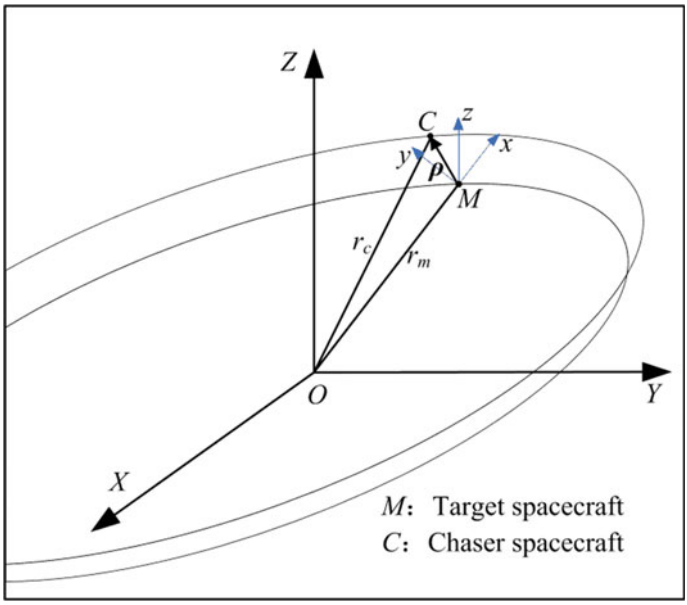


Fig. 6.1 Diagram of relative hovering operation of spacecraft

$$\ddot{\boldsymbol{\rho}} = -2\boldsymbol{w} \times \dot{\boldsymbol{\rho}} - \boldsymbol{w} \times (\boldsymbol{w} \times \boldsymbol{\rho}) - \dot{\boldsymbol{w}} \times \boldsymbol{\rho} + \frac{\mu}{r_{\mathrm{m}}^3}\boldsymbol{r}_{\mathrm{m}} - \frac{\mu}{r_{\mathrm{c}}^3}\boldsymbol{r}_{\mathrm{c}} + \boldsymbol{f}_{\mathrm{c}} - \boldsymbol{f}_{\mathrm{m}} + \boldsymbol{f}_{\mathrm{u}} \quad (6.1)$$

where

m	target spacecraft
c	chaser spacecraft
μ	gravitational constant
$\boldsymbol{w}$	orbital angular velocity of the target spacecraft
$\dot{\boldsymbol{w}}$	orbital angular acceleration of the target spacecraft
$\boldsymbol{f}_{\mathrm{c}}$	external perturbation of the mission spacecraft
$\boldsymbol{f}_{\mathrm{m}}$	external perturbation of the target spacecraft
$\boldsymbol{f}_{\mathrm{u}}$	control acceleration of the mission spacecraft
$\boldsymbol{\rho} = \begin{bmatrix} x & y & z \end{bmatrix}^{\mathrm{T}}$	the position vector of the mission spacecraft.

Obviously, Eq. (6.1) is a nonlinear model considering the external perturbation. Based on the relative stationary state characteristics of hovering spacecraft, the state $\boldsymbol{\rho}$, $\dot{\boldsymbol{\rho}}$ and $\ddot{\boldsymbol{\rho}}$ are as follows:

$$\begin{cases} \boldsymbol{\rho} = \text{const (given)} \\ \dot{\boldsymbol{\rho}} = \ddot{\boldsymbol{\rho}} = \boldsymbol{0} \end{cases} \quad (6.2)$$

Substituting the relative stationary state characteristics Eq. (6.2) into Eq. (6.1) and taking J_2 gravitational perturbation into account. $\boldsymbol{f}_{uJ2}$ is used to denote the required

control acceleration considering J_2 perturbation, and the expression is

$$\boldsymbol{f}_{uJ2} = \boldsymbol{w}_{J2} \times (\boldsymbol{w}_{J2} \times \boldsymbol{\rho}) + \dot{\boldsymbol{w}}_{J2} \times \boldsymbol{\rho} + \frac{\mu}{r_{cJ2}^3}\boldsymbol{r}_{cJ2} - \frac{\mu}{r_{mJ2}^3}\boldsymbol{r}_{mJ2} + \boldsymbol{f}_{mJ2} - \boldsymbol{f}_{cJ2} \tag{6.3}$$

where J_2 denotes the relevant orbital parameters under J_2 perturbation influence. $\boldsymbol{w}$ and $\dot{\boldsymbol{w}}$ are given as

$$\begin{cases} \boldsymbol{w} = \left[w_{xJ2} \ w_{yJ2} \ w_{zJ2} \right]^{\mathrm{T}} \\ \dot{\boldsymbol{w}} = \left[\dot{w}_{xJ2} \ \dot{w}_{yJ2} \ \dot{w}_{zJ2} \right]^{\mathrm{T}} \end{cases} \tag{6.4}$$

Merging and simplifying $\boldsymbol{w}_{J2} \times (\boldsymbol{w}_{J2} \times \boldsymbol{\rho})$ and $\dot{\boldsymbol{w}}_{J2} \times \boldsymbol{\rho}$, yields

$$\boldsymbol{w}_{J2} \times (\boldsymbol{w}_{J2} \times \boldsymbol{\rho}) + \dot{\boldsymbol{w}}_{J2} \times \boldsymbol{\rho} = \begin{bmatrix} -w_{zJ2}^2 & -\dot{w}_{zJ2} & w_{xJ2}w_{zJ2} \\ \dot{w}_{zJ2} & -w_{xJ2}^2 - w_{zJ2}^2 & -\dot{w}_{xJ2} \\ w_{xJ2}w_{zJ2} & \dot{w}_{xJ2} & -w_{xJ2}^2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{6.5}$$

and

$$\boldsymbol{W}_{J2}\boldsymbol{\rho} = \boldsymbol{w}_{J2} \times (\boldsymbol{w}_{J2} \times \boldsymbol{\rho}) + \dot{\boldsymbol{w}}_{J2} \times \boldsymbol{\rho} \tag{6.6}$$

$\mathbf{S}$ denotes the transformation matrix (*O-XYZ* to *M-xyz*), and the J_2 perturbation of the mission spacecraft in the *M-xyz* can be converted to

$$\boldsymbol{f}_{cJ2} = \mathbf{S} \cdot \boldsymbol{f}_{cJ2}^{O} \tag{6.7}$$

The required control with consideration of the J_2 perturbation can be rewritten as

$$\boldsymbol{f}_{uJ2} = \boldsymbol{W}_{J2}\boldsymbol{\rho} + \frac{\mu}{r_{cJ2}^3}\boldsymbol{r}_{cJ2} - \frac{\mu}{r_{mJ2}^3}\boldsymbol{r}_{mJ2} + \boldsymbol{f}_{mJ2} - \mathbf{S} \cdot \boldsymbol{f}_{cJ2}^{O} \tag{6.8}$$

6.3 Numerical Example

The initial orbital elements of the target spacecraft in the calculation example are shown in Table 6.1.

Table 6.1 Initial orbital elements of the target spacecraft

Orbital element	Value
Semimajor axis (m)	7×10^6
Orbit inclination (°)	45
Right ascension of ascending node (°)	0
Argument of perigee (°)	0
True anomaly (°)	0

6.3.1 Variation of Spacecraft Hovering Control Under Different Orbital Elements

Firstly, the influence of eccentricity on the hovering control is analyzed, and the hovering position $\boldsymbol{\rho}$ of the mission spacecraft is set as (1000, 1000, 1000 m). The eccentricity e of the target spacecraft is taken as 0, 0.15, 0.3, 0.45, 0.6, 0.75, and other orbital elements are shown in Table 6.1.

Using Eq. (6.9), the variation of hovering control $\left|\boldsymbol{f}_{uJ2}\right|$ in one orbital period with different eccentricity is calculated, as shown in Fig. 6.2.

In Fig. 6.2, when the eccentricity e is 0, the hovering control remains unchanged throughout the orbital period, which means that the corresponding mission spacecraft's hovering orbit is also circular orbit, and its essence is that constant hovering control and earth gravity provide centripetal force to ensure the mission spacecraft's hovering state. When the eccentricity e gradually increases, the hovering control

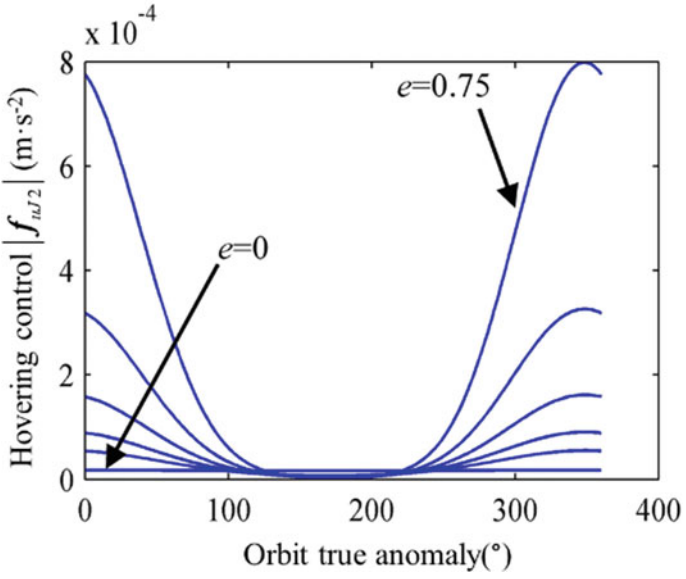


Fig. 6.2 Variation of hovering control with different eccentricity

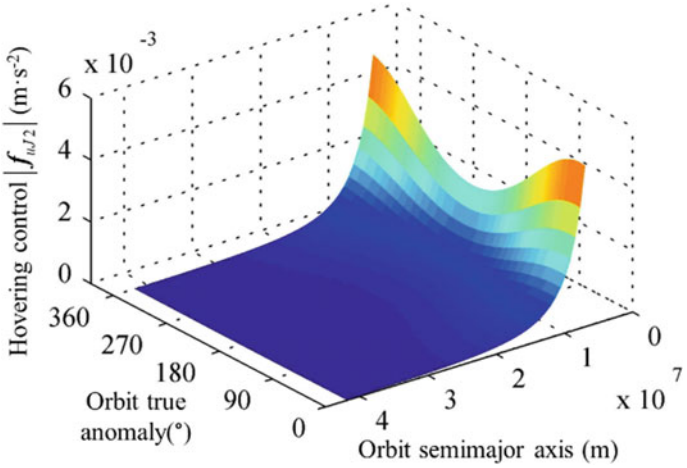


Fig. 6.3 Variation of hovering control with different orbit semimajor axes

near the perigee (when the true perigee is 0°) is large. This is because the semimajor axis a remains unchanged, and the orbital height of the target spacecraft gradually decreases and the angular velocity increases when it approaches the perigee. At this time, the mission spacecraft has to increase the hovering control force to ensure consistency with the orbital angular velocity of the target spacecraft. For the corresponding apogee (when the true near angle is 180°), the required hovering control force is small.

Set the variation range of the semimajor axis a to $7 \times 10^6 \sim 4.2 \times 10^7$ m, the hovering position $\boldsymbol{\rho}$ of the mission spacecraft is also (1000, 1000, 1000 m), the eccentricity is 0.1, and other orbital elements are listed in Table 6.1. The variation of hovering control with different semimajor axes is calculated by using Eq. (6.9), as shown in Fig. 6.3.

From Fig. 6.3, as the semimajor axis a increases, the required hovering control force for the mission spacecraft shows a gradually decreasing trend. When the semimajor axis is 7×10^6 m, the minimum hovering control force in orbital period is 2.6673×10^{-3} m s^{-2}; and the semimajor axis increases to 2×10^7 m, the maximum hovering control force is 2.2882×10^{-4} m s^{-2}, in comparison, the hovering control force decreases by 91.65%. It can be seen that the hovering control force decreases with the increase of semimajor axis throughout the orbital period, and the curve fluctuation of the entire orbital period gradually flattens out. If the hovering distance is constant at this time, the centripetal force of the mission spacecraft in orbit decreases, so the required hovering control force is reduced.

6.3.2 Research on Fuel Consumption in Hovering Orbits

This section mainly analyzes the fuel consumption of the mission spacecraft during the specified mission time. Set the original mass m_0 of the mission spacecraft to 900 kg. In view of the advantage of high specific impulse of electric propulsion, electric thruster is selected to provide continuous control force, and the specific impulse I_{sp} is 29,000 m/s (describe the amount of propellant with mass).

The orbital elements of the target spacecraft are shown in Table 6.1, and the eccentricity e is 0.1. The specified task time is one orbital period, and the range of the true anomaly is $0 \sim 2\pi$, then the fuel consumption formula is as follows Eq. (6.9).

$$\begin{cases} \Delta v = \int_0^{2\pi} \left| f_{uJ2} \right| \mathrm{d}\theta \\ \Delta m = m_0 \left[1 - \exp\left(-\dfrac{\Delta v}{I_{sp}} \right) \right] \end{cases} \tag{6.9}$$

where Δv denotes the speed increment, Δm denotes the reduced mass of spacecraft equal to fuel consumption.

Firstly, the fuel consumption of different positions for mission spacecraft hovering is calculated in the x–y plane. The range of values for the hovering positions are $x, y \in [-1000\ 1000]$, $z = 0$. Using Eq. (6.9) for calculation, the fuel distribution is given in Fig. 6.4.

From Fig. 6.4a, the fuel consumption gradually increases to 0.7 kg with the increase of x from 0 to 1000 m. As shown in Fig. 6.4b, when $x = 0$ and y ranges from 0 to 1000, the fuel consumption is less than 0.1 kg, indicating that the fuel consumption is less affected by the distance in the y direction. Figure 6.4c further demonstrates that fuel consumption is mainly affected by the position distance in the x direction. Assuming that the hovering distance remains constant at 1000 m, that is, the position distance in the x and y directions meets the condition $\sqrt{x^2 + y^2} = 1000$. The dashed circle in Fig. 6.4c represents the hovering position under assumed conditions, where the specific positions of A, B, C, and D are $(-1000, 0, 0)$ m, $(0, 1000, 0)$ m, $(1000, 0, 0)$ m, and $(0, -1000, 0)$ m, respectively. It is obvious that the fuel consumption at points A and C is the highest. As the hovering position changes with the arrow towards B and D, fuel consumption gradually decreases until it reaches its minimum at points B and D. This indicates that when the mission spacecraft is set in the x–y plane and there are no specific requirements for the hovering position, a larger hovering distance in the y-direction should be selected to reduce fuel consumption.

Similarly, using Eq. (6.9), the distribution of fuel consumption of mission spacecraft is calculated in the y–z plane. The range of values for the hovering position is, $x = 0$, $y, z \in [-1000\ 1000]$ in Fig. 6.5.

Figure 6.5 shows the distribution of fuel consumption in the y–z plane. It is evident that compared to the y direction, the fuel consumption increases faster with the increase of the hovering distance in the z direction. In the y–z plane, the hovering

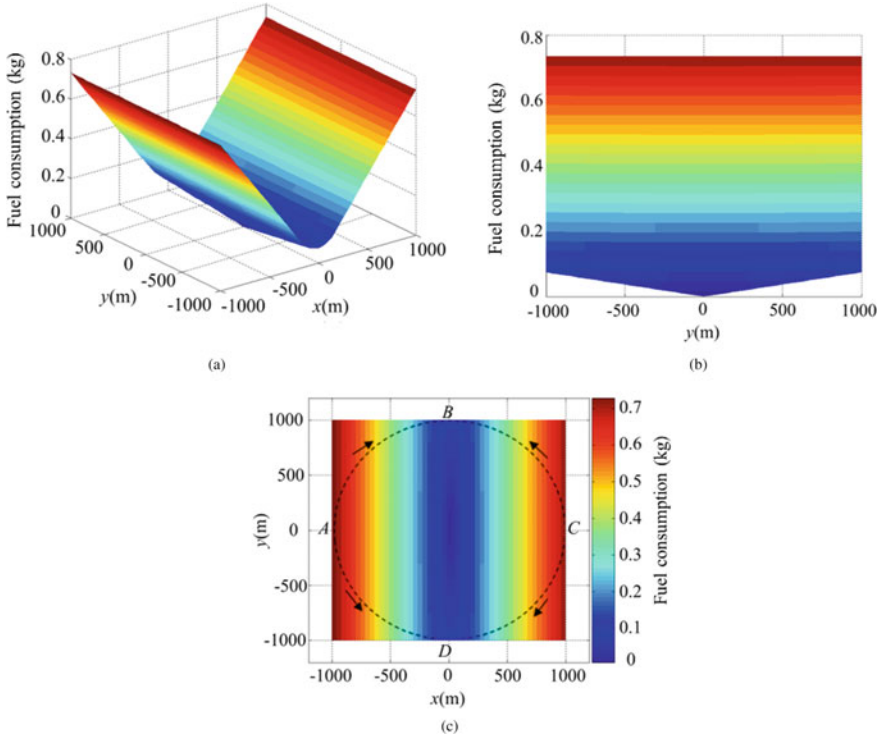


Fig. 6.4 Fuel consumption distribution at different hovering positions in the x–y plane

distance of the y and z directions meets the condition $\sqrt{y^2 + z^2} = 1000$. The dashed circle in Fig. 6.5c represents the hovering position under this condition, where the specific positions of A, B, C, and D are $(0, -1000, 0)$ m, $(0, 0, 1000)$ m, $(0, 1000, 0)$ m, and $(0, -1000, 0)$ m, respectively. Obviously, the fuel consumption at points B and D is the highest. As the hovering position changes with the arrow toward A and C, the fuel consumption gradually decreases until it reaches its minimum at points A and C, indicating that when there are no specific requirements for the specific hovering position, a larger hovering distance in the y direction should be selected to reduce fuel consumption.

Further comparing Figs. 6.4c and 6.5c, it is not difficult to find that under different plane conditions with the same hovering distance, the fuel consumption in the z direction is smaller than that in the x direction. Therefore, the following conclusion is drawn, when the hovering distance is given and the hovering position is uncertain, using reducing fuel consumption as a selection criterion, the first choice is to take a larger distance in the y direction, second in the z direction, and finally in the x direction.

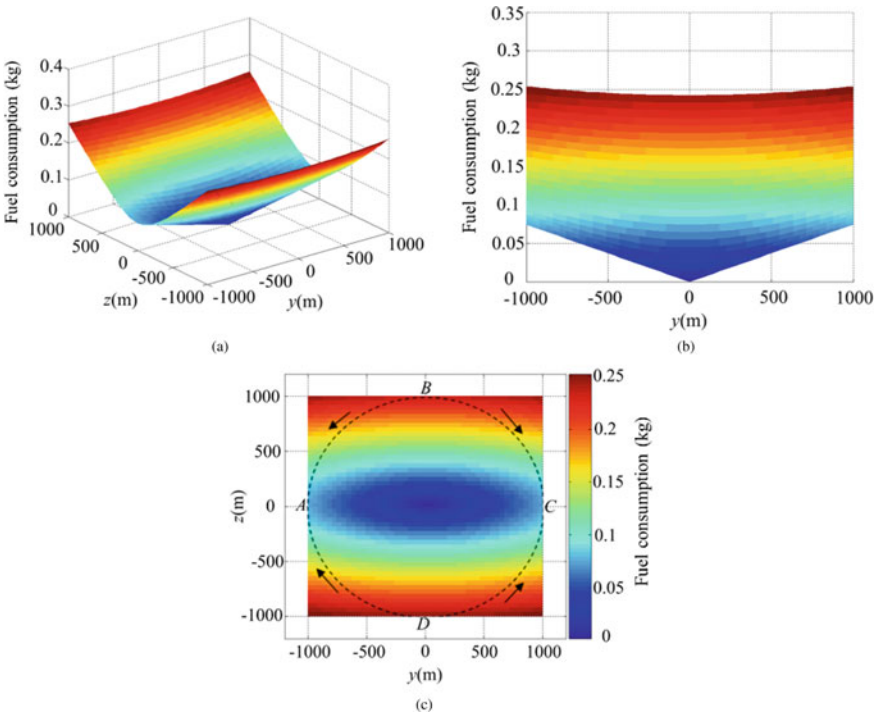


Fig. 6.5 Fuel consumption distribution at different hovering positions in the y–z plane

6.4 Conclusions

This chapter introduces J_2 perturbation into the relative dynamics model of spacecraft and obtains the hovering control equation; the influence of orbit parameters on hovering control is analyzed, mainly focusing on eccentricity and semimajor axis; then the fuel consumption within the determined range is calculated in the x–y plane and y–z plane, and specific fuel distribution cloud maps were provided. The conclusions are as follows:

(1) When the semimajor axis remains unchanged and the eccentricity e gradually increases, the hovering control near the perigee gradually increases; and the eccentricity e is constant, the hovering control decreases with the increase of semimajor axis throughout the orbit period.

(2) When the hovering distance of the mission spacecraft is determined, and there is no specific requirements for the hovering position, in order to reduce fuel consumption, the first choice is to take a larger distance in the y direction, second in the z direction, and finally in the x direction.

(3) Based on the method proposed in this article, the hovering position is selected to achieve the goal of minimizing fuel consumption, prolonging the time of on-orbit service, and providing time guarantee for the successful execution of on-orbit service.

Acknowledgements This study was supported by Ph.D. research startup foundation of University (No. Y-01-2022004) and Tai'an City Science and Technology Innovation Development Project (No. 2022GX042).

References

1. Long, A.M., Richards, M.G., Hastings, D.E.: On-orbit servicing: a new value proposition for satellite design and operation. AIAA J. Spacecr. Rockets **44**(4), 964–976 (2007)
2. Ellery, A., Kreisel, J., Sommer, B.: The case for robotic on-orbit servicing of spacecraft: spacecraft reliability is a myth. Acta Astronaut. **63**(5), 632–648 (2008)
3. Sellmaier, F., Boge, T., Spurmann, J.: On-orbit servicing missions: challenges and solutions for spacecraft operations. In: SpaceOps 2010 Conference, vol. 2159 (2010)
4. Scheeres, D.J.: Stability of hovering orbits around small bodies. In: AIAA Spaceflight Mechanics Meeting, 99-159 (1999)
5. Broschart, S., Scheeres, D.J.: Control of hovering spacecraft near small bodies: application to asteroid 25143 Itokawa. J. Guid. Control Dyn. **28**(2), 343–354 (2005)
6. Lu, E.T., Love, S.G.: Gravitational tractor for towing asteroids. Nature **438**, 177–178 (2005)
7. Wang, G.B., Zheng, W., Meng, Y.H., et al.: Research on hovering control scheme to non-circular orbit. Sci. China Technol. Sci. **54**(11), 2974–2980 (2011)
8. Zhang, J., Zhao, S., Yang, Y.: Characteristic analysis for elliptical orbit hovering based on relative dynamics. IEEE Trans. Aerosp. Electron. Syst. **49**(4), 2742–2750 (2013)
9. Huang, X., Yan, Y., Zhou, Y., et al.: Sliding mode control for Lorentz-augmented spacecraft hovering around elliptic orbits. Acta Astronaut. **103**, 257–268 (2014)
10. Huang, X., Yan, Y., Zhou, Y.: Nonlinear control of underactuated spacecraft hovering. J. Guid. Control Dyn. **39**(3), 685–694 (2016)
11. Huang, X., Yan, Y.: Output feedback control of underactuated spacecraft hovering in circular orbit with radial or in-track controller failure. IEEE Trans. Ind. Electron. **63**(9), 5569–5581 (2016)
12. Huang, X., Yan, Y.: Saturated backstepping control of underactuated spacecraft hovering for formation flights. IEEE Trans. Aerosp. Electron. Syst. **53**(4), 1988–2000 (2017)

Chapter 7
3D Scene Modeling and Real-Time Infrared Simulation Technology Based on Artificial Intelligence Algorithm

Huayan Zhu

Abstract With the continuous advancement of information technology and the increasing demand for real-time infrared simulation, establishing a high-performance real-time infrared simulation system is the primary task of battlefield information support system and the key factor to determine the outcome of the war. 3D infrared scene simulation technology is widely used in military-oriented fields such as performance test and evaluation of weapon imaging system, infrared remote sensing and mapping. In this chapter, driven by Artificial Intelligence (AI) technology, an infrared 3D scene modeling algorithm based on improved Convolutional Neural Network (CNN) is proposed. The calculation of material infrared image, the establishment of infrared image database and the real-time infrared scene modeling based on global 3D scene are preliminarily realized, and its modeling performance is simulated through experiments. The results show that after many iterations, the accuracy of the improved CNN is better than that of the traditional CNN, reaching more than 95%, and the error is also significantly reduced. The infrared image generated by this method is basically consistent with the real infrared image, which provides an effective way to prepare infrared reference map in real time.

H. Zhu (✉)
Shanghai Institute of Technology, Shanghai 200235, China
e-mail: 1525736145@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_7

7.1 Introduction

3D target recognition is of great practical significance in the military. At present, the research methods of 3D target recognition can be roughly divided into three types: based on visual computing theory, active vision and model-based recognition. Because the weapon system developed by using infrared imaging technology has a series of advantages, such as working day and night, strong anti-interference ability, high accuracy and strong flexibility, it has attracted the attention of all countries and developed vigorously [1]. Infrared imaging simulation technology has been widely used in military fields such as night navigation, optical remote sensing, precise guidance and target detection [2]. Based on the theoretical model of infrared radiation calculation, the infrared simulation technology using computer graphics and virtual reality technology has the main characteristics of flexibility, high precision, strong anti-interference ability and improving the efficiency of imaging system [3]. Because infrared imaging equipment is expensive and the production process is particularly complicated, obtaining different infrared target characteristics is a key problem to be solved urgently in infrared imaging technology [4]. In fact, although a lot of resources are invested, the captured infrared images still cannot meet the requirements of different scenes and weather conditions. Because the field test can get the real evaluation results, it is an accurate method to complete the dynamic test, but because of its shortcomings such as high cost, long cycle, high risk and large resource consumption, it cannot meet the technical research and progress requirements in the process of weapon development, thus restricting the effectiveness of its evaluation [5, 6].

With the development of computer simulation technology, infrared images under different conditions can be obtained by simulation-based methods, and the testing and verification of infrared equipment can be completed on the computer platform. The principle of infrared band imaging is different from that of visible band imaging. In infrared imaging, not only the sun-related radiation but also the thermal radiation of the object itself needs to be considered [7]. Based on the theory of visual computing, it is needed to reconstruct the 3D shape of an object by using stereoscopic images, image sequences or clues such as shadows and textures. In the future digital battlefield, whether it is military training, tactical drills, effectiveness evaluation or tactical demonstration, real-time infrared simulation scenes may be needed to support and guarantee [8]. Designing a purely digital or semi-physical computer photoelectric virtual simulation framework, establishing an infrared battlefield environment, simulating the movement of targets, interference, background, detectors and the dynamic change process of radiation energy, and constructing a radiation transmission model that conforms to the infrared imaging principle have become an indispensable and important content in the development of infrared imaging weapon systems [9]. The basic calculation model of target recognition has the characteristics of purpose-driven or purpose-guided. In this article, driven by AI technology,

an infrared 3D scene modeling algorithm based on improved CNN is proposed, and the calculation of material infrared image, the establishment of infrared image database and the real-time infrared scene modeling based on global 3D scenes are preliminarily realized.

7.2 Methodology

7.2.1 Application of AI in Infrared Image Processing

Any object in nature, such as the atmosphere, land or buildings, will have the characteristics of emitting, absorbing or transmitting electromagnetic waves when the temperature is higher than absolute zero. At present, most infrared scene simulations are based on local scene modeling, which is limited in scope, difficult to expand, single in experimental data and lacking of sufficient demonstration [10]. In infrared scene simulation, the color imbalance of visible light remote sensing images reduces the accuracy of automatic segmentation and classification of ground objects, and the temperature of the same ground object will jump when the temperature is modulated by the short-wave reflectivity characteristics of visible light remote sensing images, which needs to be dealt with. The system provides an indirect, efficient and unified interface for high-level geographic scheduling and display, and supports massive infrared scene data superposition and scheduling [11]. In the experiment, the processed infrared model is organized into a target scene and superimposed on the designated position on the earth. Infrared imaging link refers to the process that ground objects receive external radiation, and at the same time superimpose their own thermal radiation through self-reflection, and transmit upwards. After atmospheric attenuation, the infrared imaging system converts the radiation into digital images before reaching the lens of the infrared imaging system.

When obtaining visible light remote sensing images, whether for satellite images or aerial images, due to the influence of external atmosphere and illumination, the inhomogeneity of optical lenses, shooting time and other factors, the inhomogeneity of brightness and contrast of ground objects in images and the hue difference between images will be caused. In an image sequence, the scale of the target can't remain the same, maybe the target will go from far to near, from near to far, or both. In the process of going from far to near, the target becomes larger and the resolution becomes clearer, and the information becomes more and more abundant. Obviously, the contribution of the target to recognition is different under different scale resolution conditions [12]. Before constructing the full-link mathematical model of infrared imaging, it is needed to analyze the transmission path of energy between light source, atmosphere, scene and infrared imaging system, and consider the influence of various factors on infrared imaging, so as to make the constructed model more accurate and help to improve the fidelity of infrared simulation images.

Targets with small scale and large resolution provide more information, so they make great contributions to recognition; however, the target with large scale and small resolution provides less information, so it makes little contribution to recognition. In order to remove the phenomenon of uneven illumination, contrast and color difference in visible remote sensing images and realize the consistency of the overall color of the images, it is needed to balance the colors of the images. Infrared imaging link refers to the process that ground objects receive external radiation, and at the same time, they superimpose their own thermal radiation through self-reflection, transmit upwards, and reach the infrared imaging system through atmospheric attenuation, and the infrared imaging system converts the radiation into digital images.

7.2.2 3D Scene Modeling Algorithm of Infrared Image

With the increasing scale of the scene and the refinement of the model, the proportion of visible patches in all patches will gradually decrease at a specific point of view. At this time, most invisible parts in the scene can be cut out by visibility calculation, and redundant calculation of invisible parts can be avoided in the subsequent radiation field energy calculation. Image even light and color processing is a processing method to solve the color imbalance of visible light remote sensing images, so that the processed images can not only meet the subjective evaluation of human vision, but also achieve the effects of consistent brightness, moderate contrast and consistent tone between images, and optimize the mosaic effect of remote sensing images [13]. It is impossible for the posture of 3D target to change suddenly in a short time. In order to apply this dynamic constraint of target movement to target recognition, it is needed to correctly estimate the posture of the target at each moment, and then find the corresponding transfer matrix value, that is, the transfer cost between adjacent frames, according to the space transfer matrix of the target characteristic view. By increasing the high-frequency components in the image in frequency domain, strengthening the reflection information, reducing the low-frequency components in the image and weakening the ambient illumination information, the dynamic range of the brightness of the image can be reduced and the contrast degree can be enhanced, so as to achieve the effect of enhancing the characteristic information while weakening the illumination change. The dynamic fusion method of infrared image features based on CNN is shown in Fig. 7.1.

It is assumed that the scene of the dual-band infrared image is a rectangular bin $d_S = d_x d_y$ perpendicular to the main optical axis of the lens, its radiation brightness is described by L, and its observation distance is described by l, and its image $d_{S'}$ is also a rectangular bin. If the optical system satisfies the Abbe sine condition, then:

$$n_h \sin\theta_0 = n'h' \sin\theta'_0 \tag{7.1}$$

n and n' are used to describe the refractive index of infrared image scene space and corresponding image space. h and h' are used to describe the height and image height

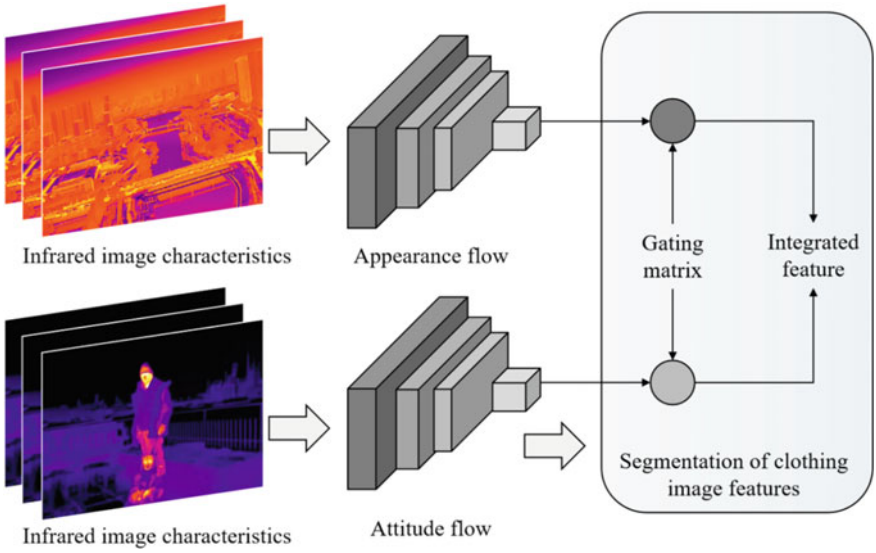


Fig. 7.1 Dynamic fusion of infrared image features

of infrared scenes. θ_0 and θ_0' are used to describe the angle between the infrared image scene space and the corresponding image space ray and the main optical axis.

The radiation power radiated from d_S to the solid corner element of the entrance pupil d_Ω $\left(d_\Omega = \sin\theta_0 d_{\theta_0} d_\phi\right)$ of the optical system can be obtained by the following formula:

$$d_P = L(\theta, \phi) d_x d_y \sin\theta_0 \cos\theta_0 d_{\theta_0} d_\phi \tag{7.2}$$

If the system has no loss in the whole process, the power needs to be calculated from the solid corner element d_Ω' $\left(d_\Omega' = \sin\theta_0' d_{\theta_0}' d_\phi'\right)$ and the image space element $d_{S'} = d_{x'} d_{y'}$.

Select a segment of the original polygon, and set the straight line on one side of the ideal rectangle corresponding to this segment as $\Delta t p_t u_t(y)$. If the straight line is set parallel to the corresponding edge of the minimum area circumscribed rectangle, the straight line is divided into $1 + \Delta t u_t\left(y'\right)$. The formula for the sum of square errors between the segment and the straight line in the original polygon is:

$$P_{t+\Delta t} = p_t + \Delta t p_t u_t(y) \tag{7.3}$$

$$P_{t+\Delta t} = p_t(1 + \Delta t u_t(y)) + p_t\left(1 + \Delta t u_t\left(y'\right)\right) \tag{7.4}$$

$P_{t+\Delta t}$ finds the first-order partial derivative of y. When the first-order partial derivative is 0, the error is the smallest, so:

$$s_{t+\Delta t}(y) = \frac{p_{t+\Delta t}}{P_{t+\Delta t}} = \frac{p_t + \Delta t p_t(y)}{p_t(1 + \Delta t u_t(y)) + p_t(1 + \Delta t u_t(y'))} \tag{7.5}$$

$$s_{t+\Delta t}(y) = \frac{p_{t+\Delta t}}{P_{t+\Delta t}} = \frac{s_t(y)(1 + \Delta t u_t(y))}{s_t(y)(1 + \Delta t u_t(y)) + s_t(y')(1 + \Delta t u_t(y'))} \tag{7.6}$$

The fitted straight line equation for this segment is:

$$s_{t+\Delta t}(y) - s_t(y) = s_t(y) \frac{\Delta t u_t(y) - \Delta t \overline{u}_t^p}{1 + \Delta t u_t} \tag{7.7}$$

The fitting linear equations of the other three segments are obtained by the same method.

After geometric modeling of 3D scene, it is needed to build texture database and material database for it, and realize the correlation between geometric model and texture database and material database, so as to facilitate the calculation of subsequent radiation energy field modules. The information of the target is related to the resolution of the image, that is, the scale. Under the condition of no zoom, the more pixels the target image occupies, that is, the higher the resolution, the smaller the scale and the richer the information. In the subsequent radiation energy field calculation module, the geometric data and material parameters of ground objects need to be used, so it is needed to consider the correlation between material parameters of objects and geometric models. It is needed to design a reasonable data organization structure in order to speed up the search of material parameters and improve the speed of system simulation in the subsequent calculation of radiation field energy.

7.3 Result Analysis and Discussion

Because of the variety of geography, climate, weather, materials and other factors in infrared scene simulation, in order to meet the real-time requirements, it is needed to establish an infrared mapping image database with different materials. The database adopts relational database structure, and is classified according to the characteristics of images. The classification principle is to divide according to infrared band, surface temperature and atmospheric environment. When designing a 3D real-time simulation platform for dual-band infrared image scene, the design of software is very important. In order to reduce the difficulty of platform development, it is needed to analyze the software functions, gather related functions together, and divide the software functions by module encapsulation. The 3D real-time simulation platform of dual-band infrared image scene should not only use the input interface to collect the infrared image of the infrared detector, but also use the output interface to play back the infrared image to the integrated processor. The simulation operating system is Windows 11, the processor is Core i7 13700k, the graphics card is RTX 3060Ti,

the memory is 16 GB, and the hard disk capacity is 1 TB. By assigning a material number to each pixel in the texture map, a texture map with the same size as the texture map is generated. The accuracy of image feature extraction of different algorithms in infrared image 3D modeling is shown in Figs. 7.2 and 7.3.

The final conclusion of the model depends on the sample set to some extent. Therefore, the selection of sample sampling technology is very important for successfully establishing a suitable infrared image feature extraction model. The parameters such as simulation time, aerosol mode and visibility are obtained, and then the atmospheric elements database is queried to obtain solar irradiance and atmospheric downward

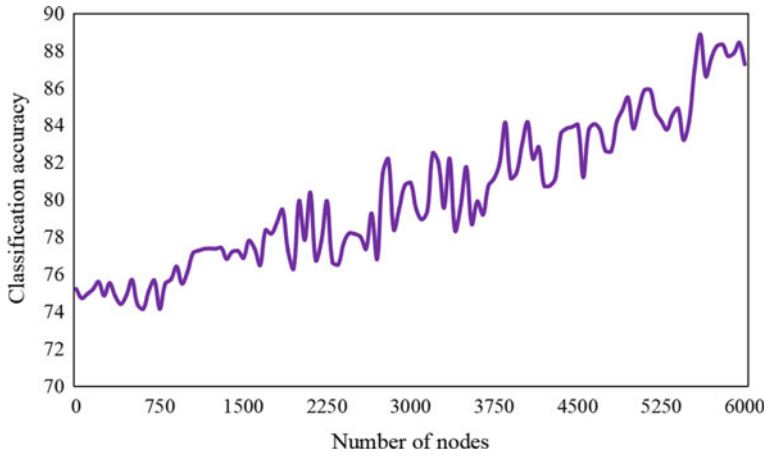


Fig. 7.2 Improving the accuracy of CNN feature extraction

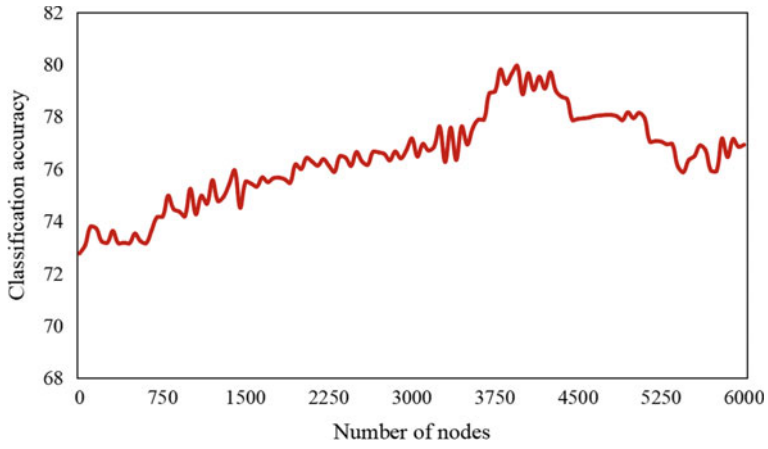


Fig. 7.3 Accuracy of traditional CNN feature extraction

transmittance parameters outside the atmosphere, and the calculation results of 3D scene and direct sunlight visibility are read in.

It does not require that the radiation of the simulation result is exactly the same as that of the actual scene, but it pursues the equivalent effect of the two under the observation of the infrared imaging system. Therefore, to simulate the infrared imaging by computer, we must start from the aspects of simulating the geometric characteristics of the scene, the infrared radiation characteristics and the effects of the atmosphere and the imaging system on the infrared radiation of the scene. Compare the recall and MAE of infrared image feature extraction model based on improved CNN with traditional CNN, and the results are shown in Figs. 7.4 and 7.5.

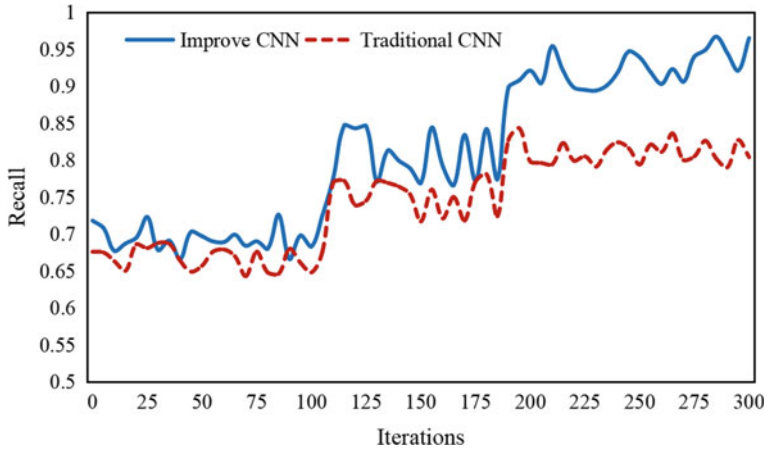


Fig. 7.4 Comparison of recall rates

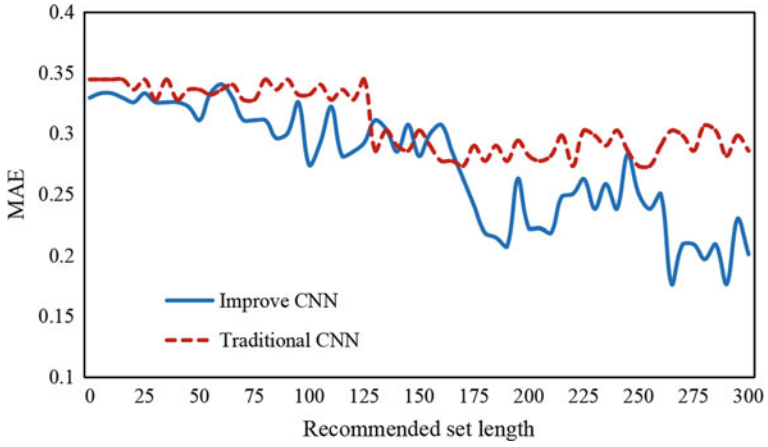


Fig. 7.5 Comparison of MAE

There are many related images in the image, and the images with high similarity rank high in the retrieval output, which can realize good retrieval expectation. As can be seen from Figs. 7.4 and 7.5, after many iterations, the accuracy of the improved CNN is better than that of the traditional CNN, reaching more than 95%, and the error is also significantly reduced. Through the improvement of this article, the convergence speed of the improved CNN parameters is faster, and finally, the modeling accuracy is higher. This method obtains ideal infrared image feature recognition results, and the recognition accuracy is higher than other feature extraction methods. Although the radiation of the generated infrared image is not exactly the same as that of the actual scene, they have equivalent effects under the observation of the infrared imaging system, which realizes the omni-directional, multi-angle and scientific management and control of tasks, simulation design, implementation coordination, efficiency evaluation and intelligent decision-making in real-time infrared simulation design.

7.4 Conclusion

Infrared imaging simulation technology has been widely used in military fields such as night navigation, optical remote sensing, precision guidance and target detection. Because the field test can get the real evaluation results, it is an accurate method to complete the dynamic test, but because of its shortcomings such as high cost, long cycle, high risk and large resource consumption, it cannot meet the technical research and progress requirements in the process of weapon development, thus restricting the effectiveness of its evaluation. Driven by AI technology, this article proposes an infrared 3D scene modeling algorithm based on improved CNN. After many iterations, the accuracy of the improved CNN is better than that of the traditional CNN, reaching more than 95%, and the error is also significantly reduced. Through the improvement of this article, the convergence speed of the improved CNN parameters is faster, and finally the modeling accuracy is higher. Although the radiation of the generated infrared image is not exactly the same as that of the actual scene, they have equivalent effects under the observation of the infrared imaging system, which realizes the omni-directional, multi-angle and scientific management and control of the real-time infrared simulation design. The infrared simulation system in this article simulates the scene on the ground, and the function of the system can be further improved and expanded in the future, so that it can support the scene simulation of the ocean and the sky and further strengthen the universality of the infrared simulation system.

References

1. Decker, R.S., Shademan, A., Opfermann, J.D., et al.: Biocompatible near-infrared three-dimensional tracking system. IEEE Trans. Biomed. Eng. **64**(99), 549–556 (2017)
2. Qiu, Y., Yang, Y., Valenzuela, C., et al.: Near-infrared light-driven three-dimensional soft photonic crystals loaded with upconversion nanoparticles. Adv. Opt. Mater. **2022**(9), 10 (2022)
3. Takahashi, S., Kimura, E., Ishida, T., et al.: Fabrication of three-dimensional photonic crystals for near-infrared light by micro-manipulation technique under optical microscope observation. Appl. Phys. Express **15**(1), 015001 (2022)
4. Ma, T., Inagaki, T., Tsuchikawa, S.: Three-dimensional grain angle measurement of softwood (Hinoki cypress) using near infrared spatially and spectrally resolved imaging (NIR-SSRI). Holzforschung **73**(9), 817–826 (2019)
5. Lv, J., Hong, B., Tan, Y., et al.: Mid-infrared waveguiding in three-dimensional microstructured optical waveguides fabricated by femtosecond-laser writing and phosphoric acid etching. Photonics Res. **8**(03), 39–44 (2020)
6. Luo, B., Wang, T., Zhang, F., et al.: Interdigital capacitive sensor for cable insulation defect detection: three-dimensional modeling, design, and experimental test. J. Sens. **2021**(6), 1–10 (2021)
7. Qiu, G.Y., Wang, B., Li, T., et al.: Estimation of the transpiration of urban shrubs using the modified three-dimensional three-temperature model and infrared remote sensing. J. Hydrol. **594**(3), 125940 (2020)
8. Shirazi, Y.N., Esmaeli, A., Tavakoli, M.B., et al.: Improving three-dimensional near-infrared imaging systems for breast cancer diagnosis. IETE J. Res. **2021**(4), 1–9 (2021)
9. Ajates, J.G., Aldana, J., Feng, C., et al.: Three-dimensional beam-splitting transitions and numerical modelling of direct-laser-written near-infrared $LiNbO_3$ cladding waveguides. Opt. Mater. Express **8**(7), 1890 (2018)
10. Katırcıoğlu, F., Çay, Y., Cingiz, Z., et al.: Infrared image enhancement model based on gravitational force and lateral inhibition networks. Infrared Phys. Technol. **100**(11), 15–27 (2019)
11. Lin, Y., Zhu, H., Peng, W., et al.: The three dimensional microspectroscopic tomography with synchrotron radiation infrared raster scanning method. Infrared Phys. Technol. **114**(3), 103649 (2021)
12. Li, R., Qiu, L., Meng, Z., et al.: A fast method for preparing a large diameter, three-dimensional photonic crystal infrared stealth material. Optik **180**(10), 894–899 (2019)
13. Li, J., Zhang, Y., Di, D., et al.: The influence of sub-footprint cloudiness on three-dimensional horizontal wind from geostationary hyperspectral infrared sounder observations. Geophys. Res. Lett. **2022**(11), 49 (2022)

Chapter 8
Simulation of Vehicle-Scheduling Model in Logistics Distribution Center Based on Artificial Intelligence Algorithm

Xiuping Zhu

Abstract The vehicle-scheduling problem of logistics distribution includes the optimization of goods collection route, goods loading and delivery route, which is the key to the optimization of logistics distribution system. The distribution vehicle scheduling problem is a NP (Non-deterministic Polynomial)-complete problem, which is usually solved by some heuristic algorithms and intelligent optimization algorithms. In this paper, the vehicle scheduling model of logistics distribution center is established based on AI (Artificial Intelligence) algorithm. The vehicle scheduling model of logistics distribution center is a very practical model, and the traditional model has the weakness of low solution accuracy or high consumption of computing resources and storage resources. Finally, an example is analyzed. From the calculation results of two examples, it is obvious that the average calculation time of the AI algorithm proposed in this chapter is much lower than that of GA (Genetic Algorithm). The average calculation time required in this paper is only 23 s, while GA takes 68 s. It can be seen that the algorithm proposed in this paper can effectively solve the logistics distribution vehicle scheduling problem. It can be seen that the algorithm in this paper effectively solves the optimal scheduling problem of logistics distribution vehicles with multiple distribution centers, multiple models and hard time windows.

8.1 Introduction

E-commerce has promoted the rapid development of the logistics industry, and with online shopping spreading to every corner of the world, the scale of logistics distribution is also constantly expanding. Logistics distribution optimization is mainly an optimization problem of distribution vehicle scheduling. It is a complex combinatorial optimization problem, which needs to consider multiple constraints

X. Zhu (✉)
Shanghai Institute of Technology, Shanghai 200235, China
e-mail: 2689096736@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_8

and objectives and belongs to NP hard problem. Vehicle scheduling problems can be further divided into single-source point scheduling problems and multi-source point scheduling problems [1, 2]. The single source point scheduling problem refers to scheduling vehicles from only one source point to deliver goods to multiple demand points. The multi-source scheduling problem refers to scheduling vehicles from multiple source points to deliver goods to multiple demand points [3]. The vehicle-scheduling model of logistics distribution centers is a highly practical model, and traditional models have weaknesses such as low solving accuracy or high consumption of computing and storage resources. How to schedule existing vehicle resources more efficiently and quickly to maximize the effects brought by resources has become an urgent issue [4]. The scheduling problem of logistics distribution vehicles includes optimization of cargo collection routes, cargo loading and delivery routes, which is the key to optimizing the logistics distribution system. The delivery vehicle scheduling problem is an NP complete problem, usually solved using some heuristic algorithms and intelligent optimization algorithms. The key to using AI algorithm to solve logistics vehicle scheduling problems is to solve particle encoding and decoding methods for scheduling problems. AI algorithm is different from other evolutionary algorithms. AI algorithm has an explicit computing model, which is easy to program and simple to implement. Once proposed, it has been widely used, especially when solving continuous function optimization problems, it shows good optimization performance [5]. This article establishes a vehicle scheduling model for logistics distribution centers, which has the characteristics of high computational efficiency and moderate solution accuracy.

8.2 Model Building

8.2.1 Model Hypothesis

Reasonable vehicle scheduling in logistics distribution center is based on the overall goal of logistics distribution center, using systematic principles and methods, making full use of all kinds of scheduling, selecting reasonable routes and means of transportation, and organizing goods transportation activities with the path, the least links, the fastest speed and the least labor consumption. This chapter will introduce AI-based algorithm to build and study the vehicle scheduling model of logistics distribution center. The left half of the AI algorithm can mainly determine the prerequisite for the application of the rule, and the right half describes the actions taken or conclusions drawn by applying the rule [6, 7]. After a rule meets the preconditions of application, the database can be operated to make it change.

The problem to be solved in vehicle smoothness is that vehicles start from the logistics distribution center to complete some distribution tasks. When the amount of each task is small, such as less than the capacity of one vehicle, in order to improve the utilization rate of vehicles, several vehicles can be arranged to perform several

transportation tasks, and how to arrange the route of vehicles can meet the needs of each task and minimize the total cost [8, 9]. The weight of each car is certain, so it is required to arrange the car route reasonably to make the total distance. In order to facilitate the modeling, the following assumptions need to be made on the model:

(1) The supply of goods in each distribution center is sufficient;
(2) The volume and rated load of vehicles are the same, the number of vehicles in each distribution center is limited, and the total number of vehicles is sufficient;
(3) Goods without special loading requirements, that is, all goods can be mixed together;
(4) The demand of each demand point must be met, and can only be delivered by a car;
(5) The distance between distribution center and customers and between customers is known and fixed;
(6) Only consider the delivery process between the distribution center and customers.

This paper studies a basic problem of vehicle scheduling, that is, the route optimization of transport vehicles under the constraints of cargo demand, delivery volume and vehicle capacity. On the premise of meeting the needs of users, the target values, such as distance, time, cost and number of vehicles, are optimized to a certain extent, and finally the minimum consumption of logistics distribution costs is achieved [10]. Among them, the openness is reflected in: the vehicles that have completed the distribution task can not return to the original parking lot; the multi-objective is reflected in: the optimization objective can be different single objectives or a combination of multiple objectives.

8.2.2 Modeling

Generally speaking, meeting customer requirements for transportation time, short transportation mileage, low transportation costs, and optimizing a single goal, which cannot meet the overall benefits [11]. The AI algorithm evolves towards a more optimal solution through repeated operations until the optimization convergence conditions are met. The AI algorithm is a repetitive search process, but this process is not simply a repetitive search, but a search with "memory". The algorithm itself prevents the search from evolving towards a lower region [12]. Use AI algorithm to identify interference prone points in logistics distribution roads and optimize vehicle paths. As shown in Fig. 8.1.

The increase in delivery frequency not only leads to an increase in the starting and total shipping costs, but also increases the ancillary activities of transportation, resulting in a decrease in various technical and economic indicators. Set up a distribution center with a total of vehicles participating in the scheduling, and the transportation distance from the customer area to is. Based on this, establish the following mathematical model:

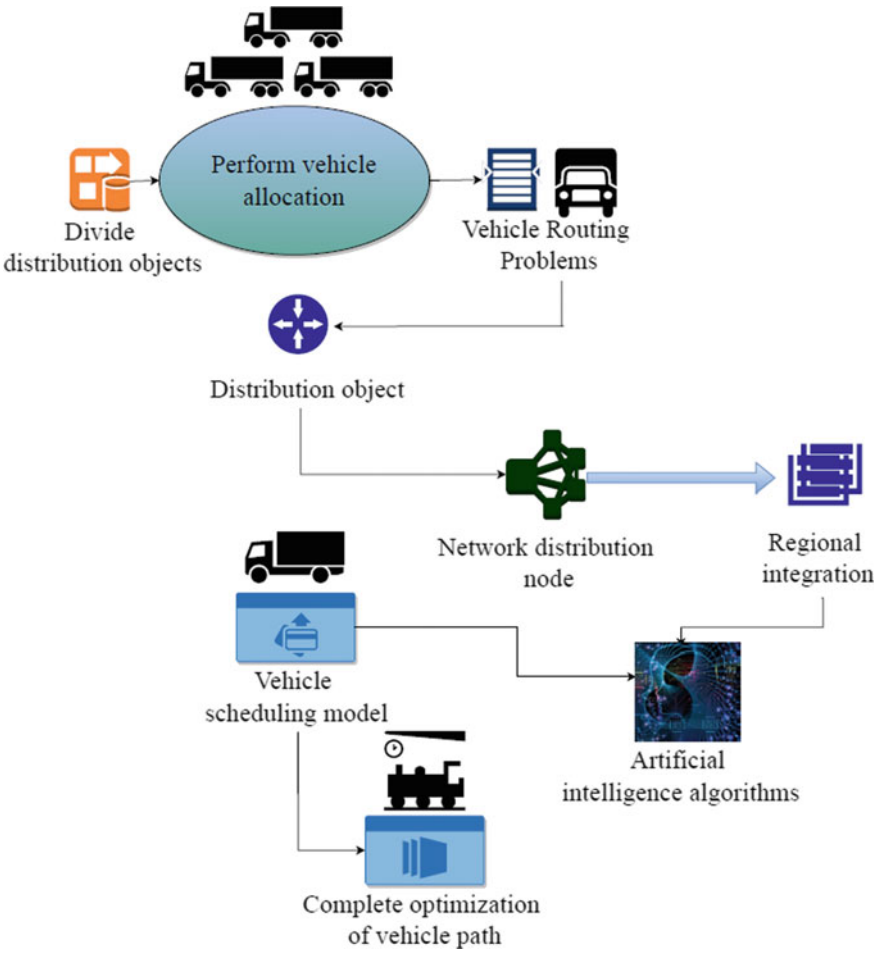


Fig. 8.1 Logistics distribution center vehicle scheduling model

$$\min F = \sum_{k=1}^{m} \sum_{i=0}^{n} c_k x_{ij} d_{ij} \tag{8.1}$$

Constraints:

$$\sum_{i=0}^{n} \sum_{k=1}^{m} x_{ijk} = \begin{cases} m & i = 0 \\ 1 & i = 1, 2, 3, \ldots, n \end{cases} \tag{8.2}$$

$$\sum_{i=0}^{n} \sum_{k=1}^{m} x_{ijk} = \begin{cases} m & j = 0 \\ 1 & i = 1, 2, 3, \ldots, n \end{cases} \tag{8.3}$$

$k = 1$ in Formula (8.1) represents the customer point that each vehicle is responsible for; $i = 0$ represents the total volume of goods demand; c_k represents the maximum volume of the vehicle; The freight volume represented by x_{ij}; d_{ij} stands for all tasks away from vehicle distribution; The x_{ijk} in Formulas (8.2) and (8.3) indicates that each customer can only be delivered by one car, and there must be another customer point i for customer point j, from customer point KK to customer point j.

How to arrange the route of the vehicle can not only meet the needs of each task, but also minimize the total cost selecting reasonable routes and means of transportation, and organizing goods transportation activities with the path, the least links, the fastest speed and the least labor consumption because of their scattered distribution, large distribution volume and fixed location [13]. If the decoded scheduling scheme is not feasible, it violates the constraints, such as the total demand of each receiving point on a certain path exceeds the capacity of delivery vehicles on this path, or there are vehicles in the scheduling scheme that are not assigned to receiving points.

8.3 Optimization Objectives for Logistics Delivery Vehicle Scheduling

Generally speaking, meeting customer requirements for transportation time, short transportation mileage, low transportation costs. After customer regionalization integration and vehicle allocation, the customer area to be delivered to each vehicle is fixed. After integrating distribution objects and allocating vehicles, the network nodes or customer areas that each vehicle needs to deliver are fixed [14]. An important issue in the routing problem of vehicles between fixed areas is the interference of road interference on logistics distribution optimization caused by frequent occurrence points. Calculate the distance between each demand point and each distribution center, classify and group according to the principle of proximity, that is, classify each demand point into the closest distribution center. Determine whether the total volume of goods required by all demand points within each group exceeds the total volume of all vehicles, or whether the total weight of goods required by demand points exceeds the total volume weight of all vehicles. Consider all distribution centers without assigned vehicles and calculate their geographical center of gravity. If there is a supersaturated group, priority should be given to assigning vehicles within the supersaturated group to perform distribution tasks; otherwise, assign vehicles to the group corresponding to the distribution center with the farthest distance from the center of gravity to perform the distribution task [15].

The problem of vehicle routing between fixed areas is transformed into a pure problem of freight forwarders. However, previous research on vehicle scheduling and distribution routes has mostly focused on optimizing a single objective, which cannot meet the overall benefits. The single source point scheduling problem refers to scheduling vehicles from only one source point to deliver goods to multiple demand points. The multi-source scheduling problem refers to scheduling vehicles from

multiple source points to deliver goods to multiple demand points. The vehicle-scheduling model of logistics distribution centers is a highly practical model, and traditional models have weaknesses such as low solving accuracy or high consumption of computing and storage resources. This chapter proposes the following optimization objectives.

(1) Minimum empty load rate of delivery vehicles

Can achieve the maximum load utilization rate of delivery vehicles. This can complete specific delivery tasks with the minimum number of vehicles and improve vehicle utilization.

(2) Deliver according to customer requirements to improve logistics service quality

Being able to meet customer requirements can enhance the reputation of the logistics distribution center, retain the company's customers, and is the fundamental guarantee for the logistics distribution center to achieve efficiency.

(3) Reasonably arrange logistics delivery to achieve the lowest delivery cost

The basic requirement for improving economic efficiency is to minimize costs, which can be another important goal second only to time accuracy.

Based on road condition information and vehicle information that can participate in scheduling, calculate the cost savings between each point pair, and then construct a line. When constructing a connecting point pair, check and constrain the maximum allowable advance time or maximum allowable delay time of points on the line to meet the requirements of hard time windows. Determine the customers served by each vehicle, develop vehicle driving routes, and display vehicle route graphics.

8.4 Simulation Experiment

Reasonable vehicle scheduling in logistics distribution center is based on the overall goal of logistics distribution center, using systematic principles and methods, making full use of all kinds of scheduling, selecting reasonable routes and means of transportation, and organizing goods transportation activities with the path, the least links, the fastest speed and the least labor consumption. Due to the constraints in solving the model, such as the number of vehicles starting from a parking lot must be the same as the number of vehicles returning to this parking lot, and each vehicle only corresponds to a starting parking lot and an ending parking lot.

Computer processing is the core of the system, mainly according to the realization of mathematical models and algorithms in the computer, and according to the given distribution users, distribution goods, collection points and vehicle conditions, the route arrangement and travel time arrangement of the dispatched vehicles and each vehicle are determined. In order to test the effectiveness of this algorithm, the vehicle scheduling model of logistics distribution center established in this paper is used to solve the situation consisting of one parking lot and six demand points. The load

Table 8.1 Coordinates and freight quality units of 6 demand points in 1 depot

Demand point	1	2	3
(x, y')	(− 4, 13)	(24, 35)	(− 7, 21)
gi	40	60	50
Demand point	4	5	6
(x, y')	(15, 38)	(− 24, 30)	(8, − 9)
gi	60	50	70

of each vehicle is 200 units and the coordinates of the parking lot are (0, 0). The specific data are shown in Table 8.1, where x and y represent the coordinates of demand points, relative to the parking lot, the quality unit of goods expressed by gi. As a result, the total transportation distance of vehicles is required to be the shortest.

The algorithm proposed in this paper and the genetic algorithm were used to calculate the example simultaneously, with a maximum iteration count of 290. The convergence speed of the algorithm proposed in this paper was compared with that of the genetic algorithm, and the experimental results are shown in Fig. 8.2. From Fig. 8.2, it can be seen that during the optimization process, the algorithm in this paper can reach the optimal value after 170 iterations, while GA requires 230 iterations, indicating that the algorithm in this paper can quickly converge to the optimal value.

Similarly, the above methods were used to test the number of optimal values and the average calculation time, and the experimental results are shown in Figs. 8.3 and 8.4.

From the calculation results of the above two examples, it can be seen that the AI algorithm proposed in this article has an optimal number of times in the optimal value test, which can reach 264 times, while GA only has 165 times. It is evident from the test results of the average computation time that the required average computation of

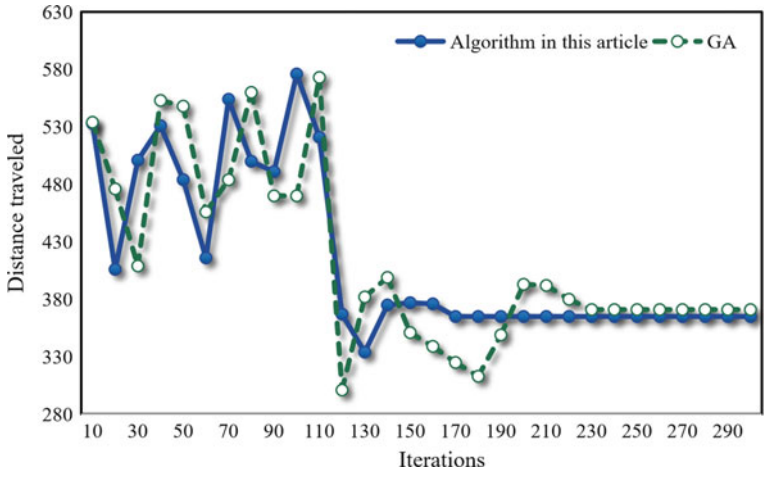


Fig. 8.2 Comparison of convergence rates

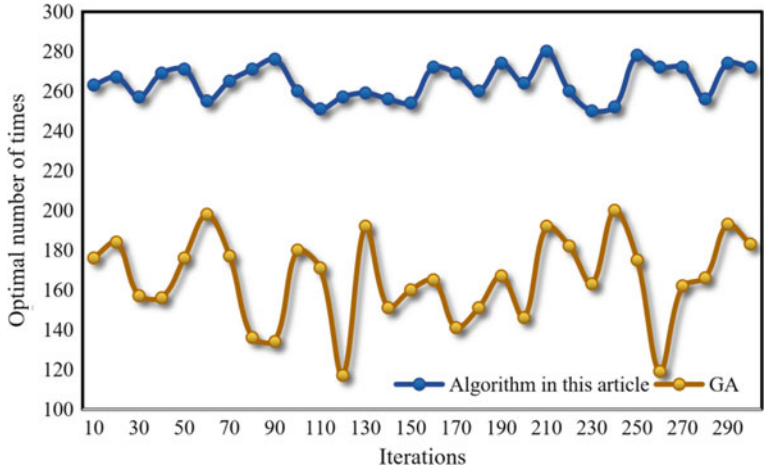


Fig. 8.3 Comparison of the number of optimal values

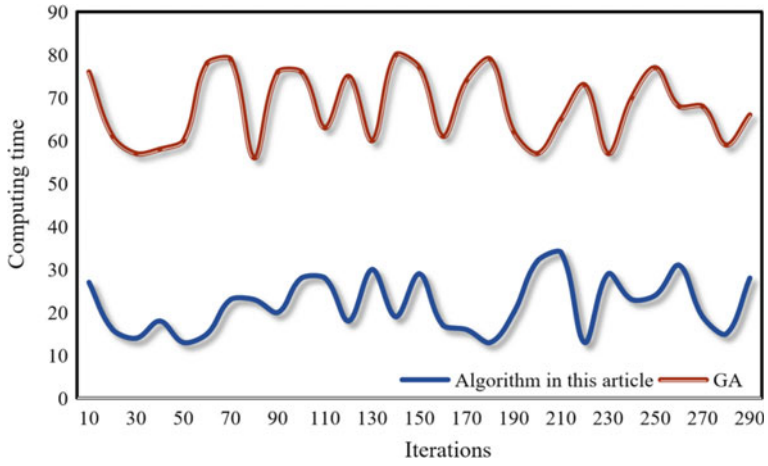


Fig. 8.4 Average calculation time

the algorithm in this paper is much lower than that of GA. The average calculation time required for this article is only 23 s, while GA takes 68 s. It can be seen that the algorithm proposed in this article can have a more significant effect on the scheduling problem of logistics delivery vehicles.

8.5 Conclusions

In this paper, the vehicle scheduling model of logistics distribution center is established based on AI algorithm, and an example is analyzed. Reasonable vehicle scheduling in logistics distribution center is based on the overall goal of logistics distribution center, using systematic principles and methods, making full use of all kinds of scheduling, selecting reasonable routes and means of transportation, and organizing goods transportation activities with the path, the least links, the fastest speed and the least labor consumption. The delivery vehicle scheduling problem is an NP complete problem, usually solved using some heuristic algorithms and intelligent optimization algorithms. The key to using AI algorithm to solve logistics vehicle scheduling problems is to solve particle encoding and decoding methods for scheduling problems. AI algorithm is different from other evolutionary algorithms. AI algorithm has an explicit computing model, which is easy to program and simple to implement. This paper will introduce AI-based algorithm to build and study the vehicle scheduling model of logistics distribution center. From the calculation results of two examples, the optimal times of the AI algorithm proposed in this paper can reach 264 times, while GA is only 165 times. In the test results of average calculation time, it can be clearly found that the average calculation required by this algorithm is much lower than that of GA. The average calculation time required in this paper is only 23 s, while GA takes 68 s. At the same time, the algorithm has the advantages of reverse application, such as changing the tasks and requirements of customers to those of suppliers, which becomes the optimal scheduling problem of goods-collecting vehicles. The AI algorithm can solve the problem of vehicle scheduling with integrated collection and delivery. The work done in this paper is a preliminary exploration in this direction, but only some relatively simple VSP, cargo loading and location positioning are studied, and there is still a lot of work to be done.

References

1. Yao, C.: Research on logistics distribution path analysis based on artificial intelligence algorithms. Int. J. Biom. **12**(1), 100–112 (2020)
2. Cui, W., Xu, X.L.: Model research of vehicle intelligent scheduling problem in distribution center based on improved differential evolution algorithm. J. Interconnect. Netw. **22**(02), 2143033 (2022)
3. Zhang, Y.: Logistics distribution scheduling model of supply chain based on genetic algorithm. J. Ind. Prod. Eng. **39**(2), 83–88 (2021)
4. Su, Z., Li, W.: The vehicle scheduling problem of third-party passenger finished vehicle logistics transportation: formulation, algorithms, and instances. IEEE Access **8**, 200597–200617 (2020)
5. Wang, C.L., Wang, Y., Zeng, Z.Y., et al.: Research on logistics distribution vehicle scheduling based on heuristic genetic algorithm. Complexity **2021**, Article ID 8275714 (2021)
6. Chang, Q.: Vehicle scheduling model of emergency logistics distribution based on internet of things. Int. J. Appl. Decis. Sci. **11**(1), 36–54 (2017)
7. Rigas, E.S., Ramchurn, S.D., Bassiliades, N.: Algorithms for electric vehicle scheduling in large-scale mobility-on-demand schemes. Artif. Intell. **262**, 248–278 (2018)

8. Wang, S.: Artificial intelligence applications in the new model of logistics development based on wireless communication technology. Sci. Program. **2021**, Article ID 5166993 (2021)
9. Zhao, J., Xiang, H., Li, J.B., Liu, J., Guo, L.Y.: Research on logistics distribution route based on multi-objective sorting genetic algorithm. Int. J. Artif. Intell. Tools Archit. Lang. Algorithms **29**(7–8), 2040020 (2020)
10. Zhang, X.: A comparative study on the influence of artificial intelligence robot distribution and traditional logistics distribution mode. J. Phys. Conf. Ser. **1966**, 012020 (1966)
11. Wang, Q.F.: A discrete artificial bee colony algorithm for solving the emergency logistics vehicle dispatching problem. J. Inner Mongolia Norm. Univ. (Nat. Sci. Ed.) **55**(19), 46–67 (2022)
12. Lan, X., Chen, H.: Simulation analysis of production scheduling algorithm for intelligent manufacturing cell based on artificial intelligence technology. Soft. Comput. **27**(9), 6007–6017 (2023)
13. Zhang, G., Xie, T., Min, L., et al.: Vehicle scheduling model based on data mining. Comput. Inf. Sci. **11**(1), 104–118 (2021)
14. Zhang, S., Lu, C., Jiang, S., Shan, L., Xiong, N.N.: An unmanned intelligent transportation scheduling system for open-pit mine vehicles based on 5G and big data. IEEE Access **8**, 135524–135539 (2020)
15. Hao, Z., Zhang, T.: Research on vehicle routing algorithm for supply chain logistics distribution. MATEC Web Conf. **227**, 02003 (2018)

Chapter 9
Construction of Intelligent Recognition System of Automobile State Based on Digital Image Processing

Danyi Zhang

Abstract Intelligent driving can collect environmental information through onboard sensors and actively control the driving operation of vehicles to replace drivers. In order to improve driving safety and promote the construction of Intelligent Transportation System (ITS), this article improves the traditional Deep Learning (DL) method and puts forward an intelligent recognition model of automobile state based on digital image processing to assist the path tracking control of intelligent vehicles. The experimental results show that this algorithm has a good effect on intelligent recognition of automobile state and has a certain resistance to target occlusion. Compared with the contrast algorithm, the accuracy of the proposed model for intelligent identification of automobile state is improved by 27.11%, and the error is reduced by 38.75%, so that the driving state of the vehicle in front of the automobile can be located more accurately. Therefore, the image processing algorithm designed in this article can accurately track the target in the visual image of smart car, and has good robustness and real-time performance.

9.1 Introduction

With the continuous increase of automobile production and sales, the problems faced by the automobile industry and the negative impact on the development of society and national economy are increasingly prominent, and the safety issue is the focus of attention. Intelligentization is an important direction of automobile technology development in the future, and environmental sensing is an important part of intelligent driving system [1]. In the twenty-first century, with the rapid development of science and technology and global economy, the problems of traffic safety and convenience brought by the increasing number of cars, high-speed cars, increasing

D. Zhang (✉)
Wuxi South Ocean College, Wuxi 214000, Jiangsu, China
e-mail: 790693032@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_9

number of unprofessional drivers and intensive traffic flow around the world need to be solved urgently [2]. In the actual driving process, most of the information obtained by drivers comes from vision. Intelligent vehicle, as the basic unit of intelligent vehicle system, can integrate many latest intelligent technologies, such as vision technology, tactile technology, autonomous control and decision-making technology, multi-agent technology, intelligent control technology, multi-sensor integration and fusion technology, so as to complete a lot of highly intelligent work [3]. Data-driven DL method provides an excellent solution for vehicle identification and state estimation [4]. However, this method needs a large number of labeled data sets for model training. The more complex the convolutional neural network (CNN) is, the more parameters it has, and the network may have higher fitting ability, but at the same time, it also needs more data for model training [5].

Intelligent driving collects environmental information through on-board sensors, makes driving decision-making planning through the processing and fusion of sensor information, and actively controls the vehicle to partially replace the driver's driving operation until it finally completely replaces the driver's operation and realizes unmanned driving [6]. In the stage of automobile intelligence, many perceptual sensors are used as the eyes and touch of smart cars. People hope to make cars have the same perception and judgment ability as drivers with the help of the wave of artificial intelligence, and help drivers make correct judgments when necessary [7]. Intelligent vehicle system is an integrated system with the ability of perceiving environment, planning and decision-making to realize automatic driving and multi-level assisted driving [8]. It can drive autonomously in different road environments through the stage of perceiving the environment, analyzing the behavior and controlling the actuators. At present, the construction of intelligent road system is still in its infancy, and the corresponding infrastructure construction period is long and the investment is large. Therefore, the development of intelligent vehicles and vehicle autonomous driving systems, through improving the vehicle's own intelligence, is the best choice to achieve safe and efficient autonomous driving at present [9]. In order to improve driving safety and promote the construction of ITS, this article improves the traditional DL method and puts forward an intelligent recognition model of automobile state based on digital image processing to assist the path tracking control of intelligent vehicles.

9.2 Methodology

Multi-modal information is effectively coded and globally reasoned, and the above individual features are fused with global features. Based on the attention mechanism, the information obtained by the inference module is integrated from the time domain perspective and the final recognition result is obtained. Through the collaborative processing of each module, the multi-modal effective information contained in the video image is extracted and effectively combined, so as to realize the function of group behavior recognition of smart car visual images. Any unprocessed original

image has a certain degree of noise interference [10]. Noise deteriorates the image quality, blurs the image, and even drowns the features to be detected, which brings difficulties to image analysis. Because of the correlation of image gray values, the energy of an image is mainly concentrated in the low frequency region [11]. The information of noise and false contour in the image is concentrated in the high frequency region, and the energy of the detailed part of the image is also in the high frequency region. The kinematics method does not consider the influence of vehicle load and tire cornering characteristics on the actual vehicle movement, but only pays attention to the position relationship between the actual vehicle trajectory and the expected trajectory.

The driver can judge whether the vehicle stays in the driving lane by perceiving the lane marking line on the expressway with his eyes. If the wheels cross the marking line on one side, it is considered that the vehicle has deviated from the channel; whether there is a curve ahead of the road, if there is a curve ahead, the driver turns the steering wheel to keep the lane. CNN performs downward unsupervised training and upward supervised error feedback, so it is need to establish a sample set composed of vector pairs of positive and negative samples. Before starting training, it is need to configure all the weight matrix values, and select differentiated small random numbers for this initialization. This method ensures that CNN's saturation is not caused by excessive weight. The training stage of automobile driving image recognition and tracking model is shown in Fig. 9.1.

The statistics of vehicle running data in smart car visual images belong to the multi-target tracking task in dynamic video scenes, and its difficulties mainly include occlusion between individuals, complex tracking and matching logic, and high real-time requirements of the algorithm [12]. Because some unstructured roads are generally located in remote areas and it is difficult to capture different weather changes of the same road type, it is very difficult and costly to collect diversified data on the spot. When the vehicle changes from an unobstructed state to an occluded state, it is need to accurately determine the occurrence of occlusion. In this chapter, the area of connected domain and the change of centroid coordinates are used to judge the occurrence of occlusion together. Usually, when the vehicle is blocked, the area and centroid coordinates of the vehicle connected domain will change greatly compared with the previous frame.

The traditional solution is to carefully design the mechanical sensing device of intelligent vehicles, or to specify and construct the environment in detail, or both. However, the problems brought about are that the cost is increased and the autonomy of intelligent vehicles is reduced, so it is difficult to be applied to any ambient intelligence vehicle. The control architecture is to study the relationship and function distribution among the parts of the vehicle system structure, and to study the reasonable hierarchy and logical structure of the system control logic in order to determine the intelligent structure and logical calculation structure of the intelligent vehicle system. For a specific intelligent vehicle, the architecture can be said to be the overall structure of the intelligent vehicle information processing and control system. Assume that $f''(x, y)$ represents the original automobile visual image, x'_{jui} represents the neighborhood of any pixel c'_{sg} in the automobile visual image, w'_{wer}

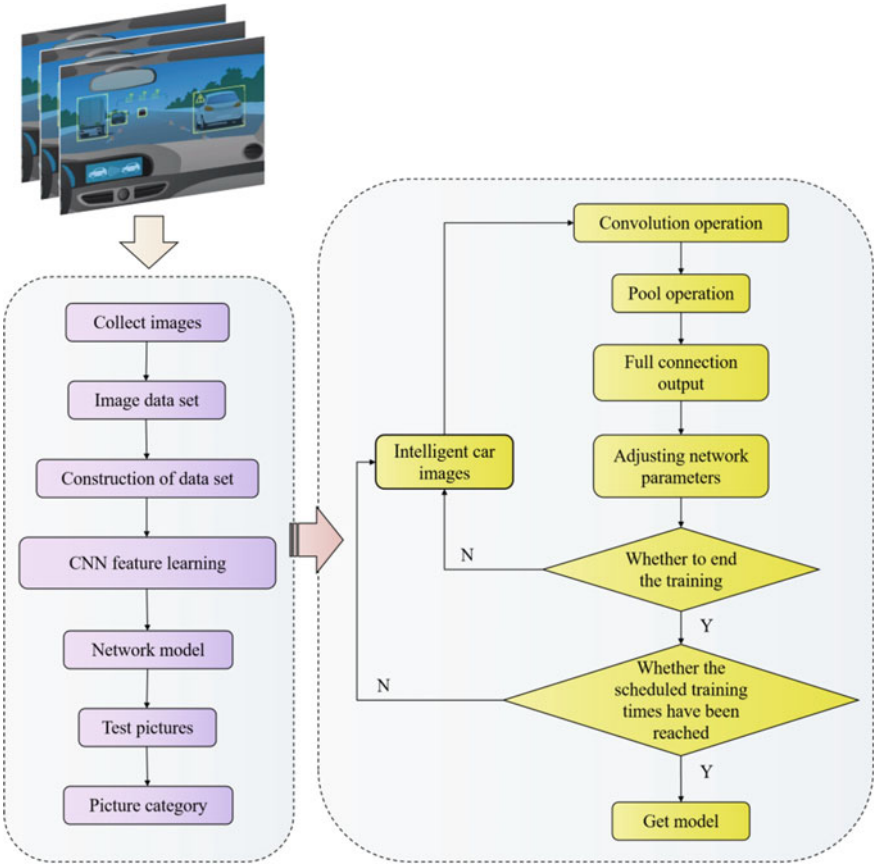


Fig. 9.1 Training stage of vehicle driving image recognition and tracking model

represents the number of values of each image feature, and b'_{wep} represents the total number of pixels in the image. The image is enhanced in frequency domain:

$$E'_{\text{qwu}} = \frac{b'_{\text{wep}} \times f''(x, y)}{x'_{\text{jui}} * c'_{\text{sg}}} \times \left\{ w'_{\text{wer}} \oplus e'_{\text{gtu}} \right\} \tag{9.1}$$

where e'_{gtu} represents the distribution state of image gray value. Suppose ξ'_{uio} represents the number of occurrences of each gray value in the image, O'_{hjk} represents the gray level of the image, (x, y) represents any point of the image, $(x + a, y + b)$ represents the disturbance point of the image, and $(x + a, y + b)^{kl}$ represents the corresponding point of (x, y) and $(x + a, y + b)$. Form a new image gray scale:

$$b''_{\text{poi}} = \frac{(x + a, y + b)^{kl}}{(x, y) \times (x + a, y + b)} \oplus \frac{\xi'_{\text{uio}}}{O'_{\text{hjk}}} \tag{9.2}$$

Because of the existence of preview distance, the intelligent vehicle can predict the road information ahead in the driving process, so as to control and adjust the vehicle in advance to obtain good control performance. In visual preview tracking, the selection of preview distance is also very important. In principle, preview distance is related to the driving speed of the vehicle. When the driving speed of the vehicle is high, preview distance should be increased accordingly. In automobile driving, the understanding of vehicle driving data has become an indispensable part of automobile driving state analysis. Let X_i^k denote the sum of the inputs of the neurons i in layer k and Y_i^k as the output. The weight of the neuron j in the $k-1$ layer to the neuron i in the k layer is W_{ij}, then there is the following functional relationship:

$$Y_i^k = f\left(X_i^k\right) \tag{9.3}$$

$$X_i^k = \sum_{j=1}^{n+1} W_{ij} Y_j^{k-1} \tag{9.4}$$

Usually f is taken as the asymmetric Sigmoid function:

$$f\left(x_i^k\right) = \frac{1}{1+\exp\left(-X_i^k\right)} \tag{9.5}$$

Assuming that the output layer is the mth layer, the actual output of the ith neuron in the output layer is Y_i^m. Let the corresponding vehicle signal be Y_i, and define the error function e as:

$$e = \frac{1}{2}\sum_i \left(Y_i^m - Y_i\right)^2 \tag{9.6}$$

CNN is essentially a mathematical mapping between input and output that can be learned independently and unsupervised. It does not need artificial feature extraction, and on the basis of clear sample classification, CNN can be trained to get ideal output [13]. Considering that the expansion and contraction of the lens will lead to the change of the area and the left and right movement will lead to the change of the centroid coordinates, two conditions are adopted to jointly judge the occurrence of occlusion. The result of target detection in smart car vision image is to form several marked connected areas in the field area of smart car vision image.

In the training stage of CNN, the training data is first input, and semi-supervised learning is adopted from top to bottom and then supervised reverse learning is carried out from bottom to top. The so-called supervised learning refers to solving the projection matrix according to the label guidance information of the training data set and various constraints. When occlusion is detected, the size of the search area needs to be changed, and the method of setting the search area without occlusion can no longer be adopted, because other adjacent vehicles other than the occluded vehicle will interfere with the occurrence of separation.

9.3 Result Analysis and Discussion

For the system described by state space, it is need to analyze controllability and observability in order to determine whether the system is observably controllable and then design the control system [14]. The simulation operating system is Windows 11, the processor is Core i5 12500f, the graphics card is RTX 3060, the memory is 16 GB, and the hard disk capacity is 1 TB. Considering that the speed of visual navigation intelligent vehicle is not high and the relative deviation is small in the stage of path tracking, the angle deviation can be expressed by radian, which can make the system meet the linear condition, thus simplifying the system into a linear time-invariant system for analysis. The problem of intelligent identification of automobile state is regarded as a classification problem. The comparison of the algorithm MAE is shown in Fig. 9.2.

It can be seen that compared with the traditional SVM method, this method has obvious advantages in the later stage of operation, and the error is reduced by 38.75%. The feature points that need to be extracted in the intelligent recognition and tracking of automobile state must ensure its stability, that is, the positions of these feature points will not change greatly because of the slight change of image quality. These feature points need to be invariant in scale and rotation, so that the same target object can be detected in different scenes.

Considering the driver's driving behavior, when driving, the driver always previews the information of the road ahead, and then determines the turning direction, corner size and speed of the vehicle according to the bending degree of the road ahead. For the path tracking of visual navigation intelligent vehicles, it should be similar to the driving behavior of car drivers. Therefore, the preview control of intelligent vehicle path tracking is proposed to adjust the driving speed and preview

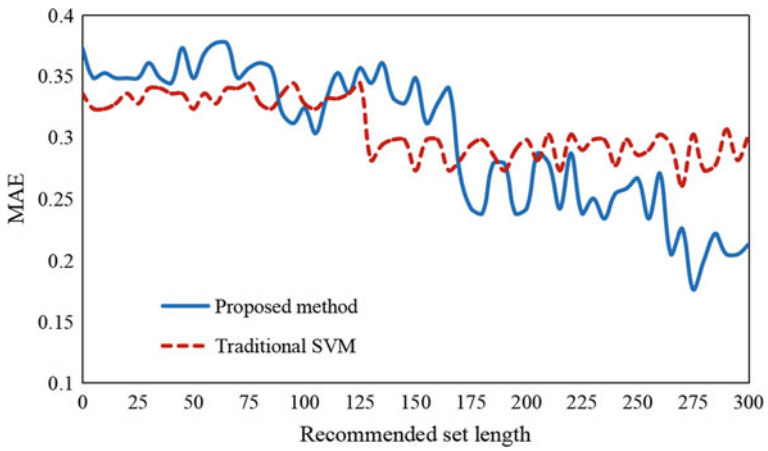


Fig. 9.2 Comparison of algorithm MAE

distance of intelligent vehicle according to the bending degree of the front path, thus improving the intelligence and robustness of intelligent vehicle path tracking.

The number of extracted feature points can be neither too much nor too little. Too few feature points will greatly reduce the matching logarithm of feature points, thus reducing the accuracy of target object detection. Too many feature points will consume a lot of calculation time, including the time of feature point location and feature point description sub-calculation. Taking the accuracy of intelligent recognition of automobile state as the test index, the traditional SVM method is selected as the comparison object, and the experimental results are shown in Tables 9.1 and 9.2.

According to the experimental data, when the number of test samples began to increase, the classification accuracy of automobile state images of all methods showed a downward trend. However, compared with the other two methods, the accuracy of intelligent identification and tracking of football car state based on this method is obviously higher. Intercept a number of image blocks with the same pixel size in the range of lanes from each image in the photo set as the image data part of the pavement material data set, and artificially filter out the image blocks that do not meet the use conditions. For vehicles, which are specific moving targets, it is superior to use the moving object segmentation method based on single image features and identified by sequence images. It is suitable for vehicle navigation, which requires high real-time performance, and has good robustness because of the outstanding characteristics of the target object. Before formal training, all neuron weights are randomly initialized. The initial value of weights plays an important role

Table 9.1 Accuracy of intelligent recognition of vehicle status using the method proposed in this article

Sample size	Accuracy of automobile state recognition (%)
15	98.88
30	98.21
45	97.76
60	97.10
75	96.94
90	96.88
105	96.57

Table 9.2 Accuracy rate of intelligent recognition of vehicle state based on traditional SVM algorithm

Sample size	Accuracy of automobile state recognition (%)
15	96.75
30	94.55
45	93.01
60	91.22
75	90.48
90	89.79
105	89.32

in the training stage of CNN. In order to get a better update in the training process, it is need to choose a suitable initial value. Too large or too small will make the weight update smaller, which will increase the number of iterations.

For different image contents, the most suitable scale coefficient is also different. Among them, the decisive factor is the texture richness of the target image. When the occlusion is judged, the occlusion relationship is stored and the search area is updated. According to the change of the number of vehicles in the new search area, the separation is judged. After separation, different matching strategies are adopted according to the driving state of the occluded vehicles. Compare the accuracy of the algorithm for intelligent identification of automobile state, as shown in Fig. 9.3.

It can be seen from the test results that this algorithm is more accurate in intelligent identification of automobile state, which is 27.11% higher than the comparison algorithm, and can accurately locate the driving state of the vehicle in front. The key technologies of visual intelligent vehicle based on path navigation are computer vision technology and intelligent control technology. After obtaining the road environment information in front of the intelligent vehicle by machine vision, it is need to make the intelligent vehicle walk along the navigation path by intelligent control technology. Working in a dynamic environment, intelligent vehicles may encounter various complex or unexpected situations, such as the sharp change of navigation paths. If the intelligent vehicle can only focus on the current path near it, it will be difficult to ensure that it can complete its task safely and effectively.

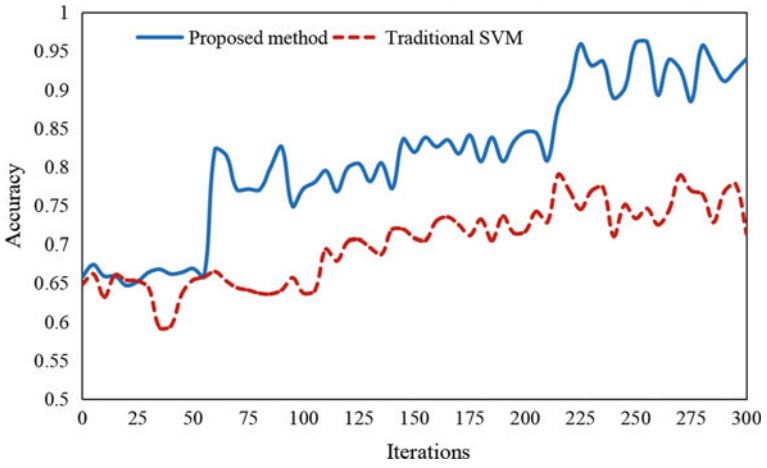


Fig. 9.3 Comparison of accuracy of intelligent recognition of automobile state

9.4 Conclusions

In the stage of automobile intelligence, many perceptual sensors are used as the eyes and touch of smart cars. People hope to make cars have the same perception and judgment ability as drivers with the help of the wave of artificial intelligence, and help drivers make correct judgments when necessary. Intelligent vehicle, as an integrated system with multiple functions, has been widely used in ITS, flexible manufacturing and assembly system, military field and special environment. Intelligentization is an important direction of automobile technology development in the future, and environmental sensing is an important part of intelligent driving system. In this article, the traditional DL method is improved, and an intelligent recognition model of automobile state based on digital image processing is proposed to assist the path tracking control of intelligent vehicles. The more complex CNN is, the more parameters there are, and the network may have higher fitting ability, but at the same time, more data is needed for model training. It can be seen from the test results that this algorithm is more accurate in intelligent identification of automobile state, which is 27.11% higher than the comparison algorithm, and can accurately locate the driving state of the vehicle in front. According to the continuity characteristics of the path, this method looks for the following path edge points near the identified path edge points, thus greatly reducing the number of image pixels to be processed and improving the real-time performance of path identification.

Acknowledgements This chapter was supported by the General Project of Natural Science Research in Higher Education Institutions in Jiangsu Province—A Research based on a Safe Search Vehicle Controlled by Raspberry Pi.

References

1. Endo, K., Ishiwata, T., Yamazaki, T.: Development of ultralow-cost machine vision system. Int. J. Autom. Technol. **11**(4), 629–637 (2017)
2. Xun, Y., Liu, J., Kato, N., et al.: Automobile driver fingerprinting: a new machine learning based authentication scheme. IEEE Trans. Ind. Inf. **16**(2), 1417–1426 (2020)
3. Shohei, N., Tomozawa, H., Mori, Y., Nakamura, H., Fujiwara, H.: Damage detection method for buildings with machine-learning techniques utilizing images of automobile running surveys aftermath of the 2016 Kumamoto earthquake. J. Disaster Res. **13**(5), 928–942 (2018)
4. Li, X.Y.: UTMD-based path planning optimization algorithm for automobile autonomous driving. J. Automob. Saf. Energy Conserv. **9**(4), 7 (2018)
5. Xu, Z.Z., Qu, D.Y., Hong, J.L., et al.: Research on decision-making method of intelligent networked automobile autonomous driving behavior. Complex Syst. Complex. Sci. **18**(3), 7 (2021)
6. He, C.H.: Research on auto-driving device system. Microcomput. Appl. **36**(9), 3 (2020)
7. Cabodi, G., Camurati, P., Garbo, A., et al.: A smart many-core implementation of a motion planning framework along a reference path for autonomous cars. Electronics **8**(2), 177 (2019)
8. Siddiqi, M.R., Milani, S., Jazar, R.N., et al.: Ergonomic path planning for autonomous vehicles—an investigation on the effect of transition curves on motion sickness. IEEE Trans. Intell. Transp. Syst. **99**, 1–12 (2021)

9. Su, S., Yang, T.T., Li, Y.J., et al.: Electric vehicle charging path planning considering real-time dynamic energy consumption. Power Syst. Autom. **43**(07), 136–143 (2019)
10. Zhu, B., Han, J.Y., Zhao, J., et al.: Intelligent vehicle path planning method based on improved RRT* algorithm of safety field. Automot. Eng. **42**(9), 7 (2020)
11. Hu, J., Zhu, Q., Chen, R.P., et al.: Global path planning of smart cars with necessary point constraints. Automot. Eng. **45**(3), 11 (2023)
12. Peng, X.Y., Hao, X., Huang, J.: Research on local path planning algorithm for driverless cars. Automot. Eng. **42**(1), 10 (2020)
13. Liu, Y.L., Chen, H.L.: Distribution path planning and charging strategy of pure electric vehicles under the influence of load. Comput. Appl. **40**(10), 7 (2020)
14. Wang, M.Q., Wang, Z.P., Zhang, L.: Research on local path planning method of intelligent vehicles based on collision risk assessment. J. Mech. Eng. **57**(10), 14 (2021)

Chapter 10
A Survey of Target Orientation Detection Algorithms Based on GPU Parallel Computing

Kewen Wang, Yu Wu, and Jiawei Tian

Abstract The ocean has rich natural resources and irreplaceable strategic role. The urgent demand for ocean exploration has promoted the rapid development of underwater acoustic engineering. Beamforming has the capability of spatial filtering, is also an important means of target bearing estimation. Beamforming is the key technology of underwater acoustic system now, and it is also one of the most computationally intensive parts in underwater acoustic signal processing. In recent years, the signal type of underwater acoustic detection has transited from narrow-band signal to wide-band signal which can carry more information, which further increases the calculation amount of beamforming, and the system has higher requirements for the speed of signal processing. Therefore, the application of parallel processing technology based on graphics processor platform to the acceleration of underwater acoustic array signals will have significant practical value.

10.1 Introduction

Underwater acoustic beamforming technology can not only act on the transmitting end of the underwater acoustic signal to make the signal emit a directional waveform, but also can be applied to the receiving end for spatial filtering and azimuth estimation, which has always been a hot spot in research. The ocean is noisy and heavily reverberated, and beamforming allows the base array to spatially reduce the effects of spatial noise and ocean reverberation. Through beamforming processing, the signal-to-noise ratio of the base array signal can be significantly improved, and the full-angle scanning using beamforming can achieve wave arrival direction estimation

K. Wang (✉) · Y. Wu
Dalian Scientific Test and Control Technology Institute, Binhai Street 16, DaLian 116013, China
e-mail: heaven996@hotmail.com

J. Tian
The Third Military Representative Office of Shenyang Maritime Decoration Bureau in Dalian, Dalian, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_10

(Direction Of Arrival, DOA). Therefore, beamforming technology is widely used in underwater detection systems, guidance systems and communication systems, so that these underwater acoustic systems have good technical performance. Therefore, beamforming technology is widely used in underwater detection systems, guidance systems and communication systems, making these underwater acoustic systems have good technical performance.

As a coprocessor, the advantage of GPU compared with FPGA is that the peak computing performance of GPU is higher. The single precision computing performance of NVIDIA Tesla P100 can reach 10.6Tflpos, while FPGA is usually less than 1Tflops. In the process of ray tracing, the number of rays to be traced may reach tens of millions, which requires a lot of floating-point operations, and floating-point operations occupy a lot of logical resources in FPGA, which may limit the performance of FPGA. In addition, GPU usually uses GDDR, Tesla P100 uses HBM2 with better performance, and FPGA uses traditional DDR, and the access bandwidth of GPU is obviously higher than FPGA. Moreover, the cost of hardware board based on embedded processors such as DSP and FPGA is expensive, lack of mature and efficient software development models and environments, and development costs and risks are not easy to control.

10.2 Development Status of Narrowband Beamforming and Realization Technology

In many underwater acoustic systems such as underwater target detection and positioning, underwater acoustic communication and ocean survey, various hydrophone arrays are equipped, which makes array signal processing a basic part of underwater acoustic signal processing and plays a pivotal role in the field of underwater acoustic detection [1]. Array signal processing mainly uses the known information such as array shape and array element properties to extract the needed useful information through signal fusion of each channel. Array signal processing technology began in the 1990s, and its theoretical research can be divided into three development stages [2]. The first phase focuses on linear conventional beamforming, the second phase focuses on adaptive beam control and adaptive zero suppression, and the third phase focuses on spatial spectrum estimation. Nowadays, array signal processing technology is gradually deepening in the research process, and the continuous innovation of hardware technology and software tools has made it not only stay in theoretical research but gradually apply to various technical fields.

Among the methods to reduce spatial background noise and ocean reverberation, beamforming is one of the most simple and effective ways, and also an important tool for target bearing estimation. Beamforming can be simply understood as a delay weighted summation of the output of signals passing through each array element in

space. The two important parameters affecting the spatial resolution of beamforming are the size characteristics of the base array and the frequency of the transmitted signal, the larger the array size, the higher the frequency of the transmitted signal, and the more accurate the spatial resolution of beamforming. Beamforming mainly studies the directivity of the array receiving system in space, and filters the specific direction through beam control to enhance the strength of the desired signal and reduce the impact of noise and interference in the environment. At the same time, the signal can be converted from time domain to space domain by using full-angle beam scanning, and the direction of the signal reaching the receiving end can be estimated by using the spatial spectrum distribution, which provides information for target positioning. After beamforming, target detection can be better realized [3].

With the increasing investment in ocean exploration, hydroacoustics has been developed unprecedentedly, and the beamforming technology of narrow-band signals has become increasingly mature. Among various beamformers, the most cost-effective is the Delay And Sum (DAS) beamformer, which has been widely used in underwater acoustic engineering because of its low cost and good effect. At the same time, when the array model has some errors, the DAS beamformer can still maintain relatively stable performance. Due to the limitation of array aperture size, its spatial resolution is not high, and the effect of restraining interference is insufficient [4]. It can be inferred from Rayleigh criterion that for two point sources located in the far field of the array, DAS method can accurately distinguish two targets only when their included angle is greater than the reciprocal of the aperture of the array. Therefore, the number of array elements and the distance between array elements will determine the spatial resolution accuracy of the array. Capon proposed the Minimum Variance Distortionless Response (MVDR) beamformer in 1969 [5]. He improved the conventional beamforming method, so that the ideal array can obtain the highest gain in the desired direction, and can effectively distinguish multiple targets. However, in order to fully and effectively realize the performance of MVDR beamformer, it is necessary to establish an accurate array signal model. Scholars began to focus on both of the performance and effect of beamforming algorithm. At the 11th IBCAST Conference in 2014, Hamid provided a performance comparison of various classical time-domain and frequency-domain beamforming technologies by taking into account the execution time of general-purpose processors, hardware and processing requirements [6]. In beamforming, in order to suppress the interference of the signal in the unexpected direction, the sidelobe level of the beam is required to be low. Therefore, the optimization design of the beam pattern is an important issue. Based on Chebyshev's polynomial, Dolph designed a weight method for linear arrays with the same element spacing that can reduce the sidelobe level of the beam and distribute it evenly, but this method requires that the array element of the receiving array can receive the signal in all directions [7]. Cui et al. designed a method that can minimize the sidelobe level of non-uniform array through Olen and Compton adaptive beamforming methods [8]. In order to ensure the stability of the algorithm in the

complex sonar system, Wu et al. improved the iterative step factor in Feng and Wu's method [9], and used the adaptive ability of the interference spectrum structure to iterate the step factor. Experiments show that this robust beamforming method has lower sidelobe level and higher spatial resolution than the conventional method [10].

10.3 Development Status of Broadband Beamforming and Realization Technology

At present, the processing method of broadband beamforming mainly refers to narrowband beamforming. The broadband beamforming is decomposed into multiple sub-bands through frequency domain division, and each sub-band is processed according to the narrow-band beamforming method. Finally, the broadband beamforming result is obtained by adding the energy of all sub-bands [11]. When a broadband signal passes through an array with certain parameters, the response of different frequency components in the signal to the array is different within the half-power beam width, which will cause the output of the array to be distorted when the signal is incident on the array within the main lobe width angle of the beam, which will affect the accuracy of the system in subsequent processing. The design purpose of constant beam width beamforming is to ensure that the beam shape formed by the signal after passing through the array has no fluctuation in all frequencies of broadband signal. The earlier method is to use sub-arrays with different apertures to receive signal components of different frequencies, so that the compensation filter changes with frequency. In practice, only a finite length of subarray can be used, which makes it difficult to ensure that the beam width is constant in the whole design frequency band, and the implementation of this method is very complex and expensive. For wideband array, Lee sets the manifold vector of narrowband array with the central frequency of each sub-band through the subspace transformation of coherent signals, and uses the MUSIC algorithm in the beam domain to achieve the target DOA estimation method through a constant beam width technology [12]. Later, Ward et al. proposed a new adaptive broadband beamforming algorithm based on adaptive LMS algorithm, which solved the minimum variance beamforming problem and had the beam constraint of constant structural frequency. This constraint reduces the dimension of the adaptive problem and only uses some parameters of the adaptive array, reducing the computational complexity and increasing the convergence speed [13]. Liu et al. proposed a beamforming method for three-dimensional arrays with uniform distribution, which does not use FIR/IIR filter. By using two alternative conditions, the dependence of the system on time and frequency is eliminated. The simulation results show that the method can well meet the requirements of constant beam width [14]. As time goes by, broadband beamforming technology is also being optimized. Zhou et al. divided all bands and receiving arrays of broadband signals into a group

of sub-bands to obtain a fast convergent adaptive beamforming weight, which effectively reduces the misalignment effect of spatial correlation matrix while reducing the complexity of the algorithm [15]. Manamperi et al. introduced second-order cone planning technology into beam forming design in order to optimize the array beam. In the next year [16], Wang et al. proposed a design method of broadband beamformer with constant main lobe response based on the second-order cone planning, and used this method to design the beam to transform the signal into the beam domain for target bearing estimation. This method can not only ensure that the main lobe response of the beam is constant, but also make the side lobe form a certain width and depth depression at the desired angle, so as to suppress interference signals in space [17].

Broadband beam processing technology is divided into time domain and frequency domain processing methods. In frequency domain processing, array data is decomposed into several sub-bands by DFT transform of signals, and then narrowband processing is carried out in each sub-band, and then IDFT transform is carried out on the sub-band output to obtain broadband output time domain waveform. Time domain processing is to let the time domain signal pass through a set of corresponding transverse FIR filters.

The linear conditional adaptive beam processing method proposed by Kondratyuk can be considered as the broadband application of MVDR beamformer [18], but this algorithm requires high-precision mechanical rotation to achieve the purpose of accurate delay compensation, which is very expensive [19]. Gershman found that when the size of the input data block of DFT beamforming is the same as the length of the time-domain FIR beamforming filter, the frequency-domain weighting vector value of each array element of DFT beamforming and the impulse response function of the time-domain beamforming FIR filter are just Fourier transform pairs [20]. Godara proposed a time-domain FIR broadband beamforming implementation method based on FFT, using the weighted vector value of each frequency subband [21] to calculate the impulse response of the FIR filter of each array, However, because the FIR filter is simply obtained by the method of inverse Fourier transform according to the expected response of the beam, it is the simplest window function design method, and the design and implementation are not ideal. Moreover, because the length of the FIR filter must be equal to the length of DFT data, in order to improve the design accuracy, the subbands need to be divided densely, and the filter order needs to be increased.

In time domain processing, the impulse response of FIR filter affects the space and frequency characteristics of beamformer. The performance of conventional beamformer depends on the accuracy of FIR filter. At present, there are many design and implementation methods of fractional delay filter. The window function truncation method is realized by adding a window to the impulse response function of the filter and directly truncating the sinc function to approximate the ideal system response. According to Paswal's theorem, the order of the filter directly affects the design

error [22]. The least squares method uses the norm error criterion to construct the objective function according to the design and expected filter frequency response error, and calculates the gradient extremum of the two norm error function to solve the filter impulse response. Maximally Flat makes the error function reach approximately optimal at a specific frequency point, the impulse response function of the filter is solved as the solution of the Lagrange nonlinear constraint function, the performance is better than the window function method in the case of a certain filter length, the low-frequency response error is small, but the high-frequency error is large, and the overall is not as stable as the window function method, and the passband cannot be accurately controlled. The second-order cone programming design method, based on the idea of approximating the expected response function, uses the norm constraint criterion to optimally solve the error function, which can flexibly control the constraints according to needs, and the design performance is good and the error is small. However, for some cases where there is no optimal solution, it needs to increase the order, which will increase the calculation amount [23].

10.4 Overseas and Domestic Research Status of GPU Parallel Computing

In 2007, NVIDIA officially released CUDA. Compared with the CPU focusing on logical judgment, the GPU is designed to solve intensive data computing problems. A GPU can have thousands of CUDA cores, and the controller on the core can schedule 32 threads to process the same instruction at the same time. Benefit from the super floating-point computing power of GPU, as shown in Fig. 10.1. Tianhe 1A supercomputer won the TOP500 championship in 2010 and Titan supercomputer won the TOP500 championship in 2012. In the same year, in the ILSVRC image classification competition, the AlexNet multi-layer neural network trained by two NVIDIA GTX580 GPUs won the championship with scores more than ten percentage points higher than the traditional method. The popularity of deep learning also promoted the continuous improvement of GPU computing ability. NVIDIA released Maxwell, Pascal, Volta and Turing architectures in 2015, 2016, 2017 and 2018, respectively. The latest Turing architecture implements ray tracing through hybrid rendering. The new RTCore processes rays 25 times faster than Pascal. The processing performance of computing core is not getting faster and faster, but developing towards parallelism. With the advantages of multi-core, GPU has become a high-performance parallel processing platform for beamforming algorithm [24].

GPU has proved its feasibility and processing efficiency in the application of ultrasonic imaging beam focusing similar to underwater acoustic beam forming [25]. Yoshida et al. used NVIDIA Geforce8800 GPU and Intel Xeon CPU to perform conventional time-domain and frequency-domain beamforming calculations and

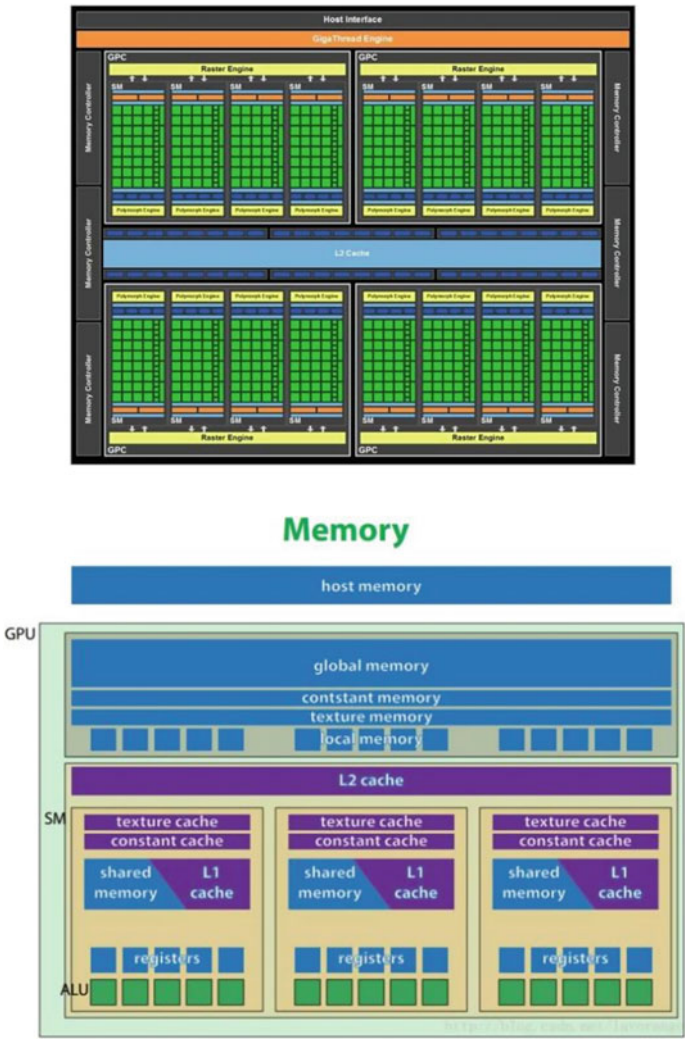


Fig. 10.1 GPU architecture

measure their performance. The results show that using GPU is more efficient [26]. Liu proposed a parallel algorithm for the frequency domain broadband beamformer and conducted experimental tests based on CUDA architecture, which proved that the CUDA technology can achieve real-time multi-beam forming of array signal data blocks, and has the performance advantage of high efficiency and high speed in high-density computing with large amount of data [27]. Furthermore, Sun applied GPU to underwater 3D imaging and simulated 3D imaging of complex underwater scenes using broadband signals, which reflected the correctness and efficiency of the underwater 3D imaging method based on GPU [28]. Khan adopted GPU based

on CUDA parallel programming model, combined with Intel MKL library to solve SVD, and achieved the common beam forming processing method with high real-time requirements in the active detection system, which shows the feasibility of CUDA technology in real-time underwater acoustic detection field [29]. Yiu, B.Y.S., et al. proposed a method of delay superimposed beamforming using GPU groups, and used two GTX-470s and a single GTX-275 for simulation, highlighting the processing speed advantage of GPU groups [30]. Inspired by Smirnov's scientific research achievements, Dhillon proposed the idea of implementing the CUDA Real FIR algorithm on the GeForce8800 platform, which has been verified to be nearly 100 times faster than the CPU [31]. Differences between CPU and GPU architectures as shown in Fig. 10.2, and comparison of CPU and GPU processing speeds as shown in Fig. 10.3. Vranjes-Wessely et al. have designed a parallel computing scheme based on GPU platform that uses multiple channels for FIR filtering [32]. This method deduces the basic principle of matrix multiplication, analyzes its parallelizable method, creates a parallel method model based on GPU platform, and successfully achieves the goal of balancing the load of GPU processing data, which not only effectively relieves the bandwidth limitation problem of I/O, but also successfully makes it possible to realize fast filtering by using multiple channels. Yang et al. studied a high-speed parallel FFT computing method based on GPU [33], and proposed an optimization scheme that can realize thread parallel mapping, the core idea of the scheme is to combine computing characteristics of FFT and advantages of GPU storage architecture, so that the parallelism of FFT can be effectively improved, so as to effectively improve the parallelism of FFT and performance optimization of FFT algorithm. Experiments show that the computation speed of the parallelized FFT algorithm is 2 to 6 times faster than the CUFFT function provided by the system.

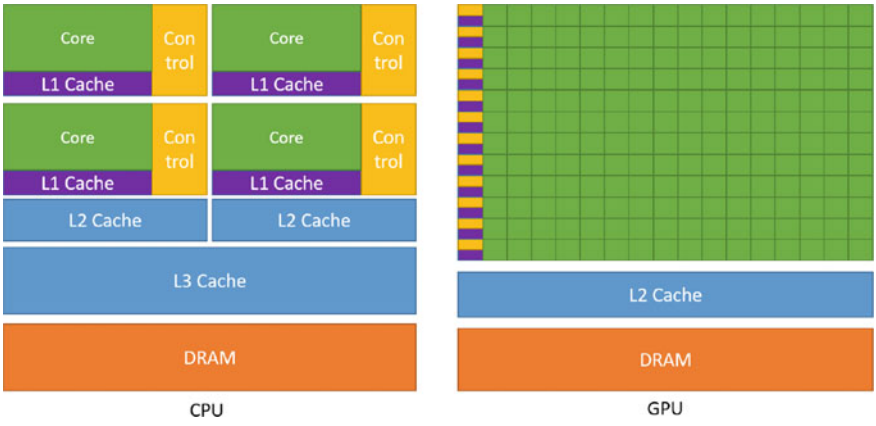


Fig. 10.2 Differences between CPU and GPU architectures

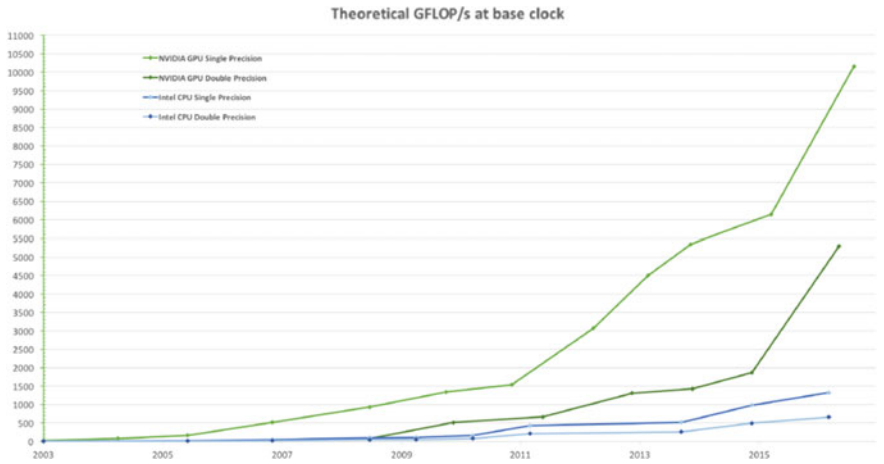


Fig. 10.3 Comparison of CPU and GPU processing speeds

10.5 Conclusion

GPU has been commercialized on a large scale, with mature standard hardware modules and shelf products with standard interfaces. The software development of CPU and GPU is unified through CUDA, which makes CPU focus on system management and GPU focus on computing. Compared with the development cost and risk of embedded processors such as DSP and FPGA, parallel computing technology based on CUDA platform can significantly reduce development cost and risk, and has high software development efficiency and strong software portability, which has broad development prospects in the field of signal processing.

References

1. Wang, J., Li. Y. A.: Research on wideband underwater acoustic signal processing based on Mellin transform. Binggong Xuebao/acta Armamentarii **28**(1), 87–90 (2007)
2. Barua, S., Rong, Y., Nordholm, S., Chen, P.: A labview-based implementation of real-time adaptive modulation for underwater acoustic OFDM communication. In: Global Oceans 2020: Singapore–US Gulf Coast, pp. 1–6. IEEE (2020)
3. Huang, L., Li, M., Shen, B., Liu, Z.: Subband method for wideband adaptive beamforming with conjugate quadrature mirror filter bank. In: 2007 IET Conference on Wireless, Mobile and Sensor Networks (CCWMSN07), pp. 701–704. IET (2007)
4. Trees, H.: Detection, Estimation, and Modulation Theory, Optimum Array Processing. Publishing House of Electronics Industry (2013)
5. Capon, J.: High-resolution frequency-wavenumber spectrum analysis. Proc. IEEE **57**(8), 1408–1418 (1969)
6. Hamid, U., Qamar, R. A., Waqas, K.: Performance comparison of time-domain and frequency-domain beamforming techniques for sensor array processing. In: Proceedings of 2014 11th

International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 14th–18th January 2014, pp. 379–385. IEEE (2014)
7. Deshmukh, S.B., Deshmukh, A.A.: Series fed designs of planar circular and hexagonal microstrip antenna arrays for high gain and reduced first side lobe level radiation. J. Microwaves Optoelectronics Electromagnetic Appl. **21**, 114–130 (2022)
8. Cui, C., Li, W.T., Ye, X.T., Rocca, P., Hei, Y.Q., Shi, X.W.: An effective artificial neural network-based method for linear array beampattern synthesis. IEEE Trans. Antennas Propag. **69**(10), 6431–6443 (2021)
9. Feng, W., Hu, D.: A modified whale optimization algorithm for pattern synthesis of linear antenna array. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **104**(5), 818–822 (2021)
10. Xiong, Z., Wang, D.: OFDM-MIMO radar main-lobe transmitting interference suppression method based on robust beamforming. In: 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), pp. 53–60. IEEE (2021)
11. Li, X. M., Huang, H. N., Li, Y., et al.: A broadband high resolution direction of arrival estimation algorithm based on conditional wavenumber spectral density. Chin. J. Acoust. **39**(4), 482–497 (2020)
12. Sanderovich, A., Kasher, A. Y., Eitan, A. P.: Efficient Beamforming Technique, US11228348B2 (2022)
13. Guo, W.: An adaptive beamforming algorithm for sound source localisation via hybrid compressive sensing reconstruction. J. Vibroengineering **24**(3), 13 (2022)
14. Wei, L., Mclernon, D.C., Ghogho, M.: Design of frequency invariant beamformer without temporal filtering. IEEE Trans. Signal Process. **57**(2), 798–802 (2009)
15. Quan, Z., Zhang, Y., Liu, J., Wang, Y.: Maximum element dichotomous coordinate descent based minimum variance distortionless response DoA estimator. Electronics **10**(23), 2966 (2021)
16. Manamperi, D.I., Ekanayake, J.B.: Performance comparison of optimum power flow based on the sequential second-order cone programming in unbalanced low voltage distribution networks with distributed generators. Int. Trans. Electrical Energy Syst. **31**(12), e13224 (2021)
17. Wang, P., Li, X., Du, T., et al.: A high-resolution minimum variance algorithm based on optimal frequency-domain segmentation. Biomed. Signal Process. Control **67**, 102540 (2021)
18. Kondratyuk, N., Nikolskiy, V., Pavlov, D., et al.: GPU-accelerated molecular dynamics: state-of-art software performance and porting from Nvidia CUDA to AMD HIP. Int. J. High Perform. Comput. Appl. **35**(4), 312–324 (2021)
19. Li, S.: Research on the development status and tendency of general hospital buildings under the background of smart city in the future. In: International Symposium on Advancement of Construction Management and Real Estate, pp. 243–254. Springer Singapore, Singapore (2021)
20. Gershman, A. B.: Robust adaptive beamforming: an overview of recent trends and advances in the field. In: 4th International Conference on Antenna Theory and Techniques (Cat. No. 03EX699), vol. 1, pp. 30–35. IEEE (2003)
21. Zou, X., Zhang, Q., Zhang, W., Hu, J.: Robust adaptive beamforming based on the direct biconvex optimization modeling. In: 2021 IEEE Radar Conference (RadarConf21), pp. 1–5. IEEE (2021)
22. Sun, J., Wang, Y., Shen, Y., Lu, S.: Design of variable fractional delay FIR filter using the high-order derivative sampling method. Digital Signal Process. **123**, 103394 (2022)
23. Liu, R., Cui, G., Lu, Q., Yu, X., Feng, L., Zhu, J.: Constant beamwidth receiving beamforming based on template matching. In: 2021 IEEE Radar Conference (RadarConf21), pp. 1–5. IEEE (2021)
24. Donepudi, P.K.: Leveraging cloud computing and high performance computing (HPC) advances for next generation projects and technologies. Am. J. Trade Policy **7**(3), 73–78 (2020)
25. Kwon, J., Song, J. H., Bae, S., Song, T. K., Yoo, Y.: An effective beamforming algorithm for a GPU-based ultrasound imaging system. In: 2012 IEEE International Ultrasonics Symposium, pp. 619–622. IEEE (2012)

26. Yoshida, S., Kameda, S., Suematsu, N., et al.: Re-evaluation of a dual-feed linear polarized 2-by-2 circular patch array antenna for 60-GHz-band digital beamforming applications. IEEJ Trans. Electr. Electron. Eng. **16**(12), 1653–1655 (2021)
27. Liu, X., Yu, C., Xin, F.: Gradually perforated porous materials backed with Helmholtz resonant cavity for broadband low-frequency sound absorption. Compos. Struct. **263**(1), 113647 (2021)
28. Sun, Y., Zhang, Q., Zhang, A.:. Research and simulation of motion planning for underwater hexapod robot based on improved artificial potential method. In: 2021 6th International Conference on Control and Robotics Engineering (ICCRE), pp. 7–12. IEEE (2021)
29. Khan, C., Dei, K., Schlunk, S., et al.: A real-time, GPU-based implementation of aperture domain model image reconstruction. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **68**(6), 2101–2116 (2021)
30. Yiu, B. Y., Tsang, I. K., Alfred, C. H.: Real-time GPU-based software beamformer designed for advanced imaging methods research. In: 2010 IEEE International Ultrasonics Symposium, pp. 1920–1923. IEEE (2011)
31. Dhillon, V., Nair, S., Pabarekar, A., Kumbhare, M., Thakur, K., Krishnan, R.: Implementation of FIR digital filter on FPGA. In: 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), pp. 1–5. IEEE (2021)
32. Vranjes-Wessely, S., Misch, D., Kiener, D., Cordill, M. J., Frese, N., Beyer, A., et al.: High-speed nanoindentation mapping of organic matter-rich rocks: a critical evaluation by correlative imaging and machine learning data analysis. Int. J. Coal Geol. **247**, 103847 (2021)
33. Yang, S.C., Wang, Y.L.: A hybrid MPI-CUDA approach for nonequispaced discrete Fourier transformation. Comput. Phys. Commun. **258**, 107513 (2021)

Part II
Intelligent Image and Data Processing

Chapter 11
Optimal Design of Hydrodynamic Journal Bearing Based on BP Neural Network Optimized by Improved Particle Swarm Algorithm

Xinliang Hu, Jun Wang, Shifan Zhu, and Wangyan Dong

Abstract In this paper, an algorithmic model of hydrodynamic journal bearing with the goal of optimizing particle algorithm is set up, and particle algorithm is optimized by combining artificial neural network. The effect of particle algorithm coupled with neural network on the optimized design results of hydrodynamic journal bearings is studied, and the computational flow of the coupled two algorithms is proposed. The calculation example shows that the selection of the parameters of neural network based on the improved algorithm majorization in hydrodynamic journal bearings proposed in the paper is reasonable and can be extended to other fields of optimization design based on particle swarm algorithm.

11.1 Introduction

Hydrodynamic journal bearing is important support component part in machine, and they are quite widely used in various mechanical equipments; thus, the study of optimal design of hydrodynamic journal bearing is of great importance. With the continuous efforts of many scholars, the research on the optimal design of hydrodynamic journal bearings has achieved rapid development, and the research has continued to deepen [1].

The particle swarm intelligent optimization algorithm (PSO algorithm) recommended by Russell has features of simple computational process, no gradient information, less parameter setting, and high search density of the algorithm, which can efficiently deal with discrete computational problems and has good applicability in the optimization design of bearings [2]. However, it is found in the research of the particle swarm algorithm which is unstable when particle swarm algorithm with fixed

X. Hu (✉) · J. Wang · S. Zhu · W. Dong
Department of Mechanical EngineeringDepartment of Arms Engineering, PLA Army Academy of Artillery and Air Defense, Hefei Anhui 230000, China
e-mail: Huxinliang99@yeah.net

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_11

parameters is used to optimize under the constraints, and sometimes the real optimal solution cannot even be obtained.

Neural networks are not only widespread used in problem of predicting the tribological properties of materials, freeing people from complex intrinsic mechanism analysis, but also making reasonable use of the powerful ability of neural networks to mine the intrinsic laws of data, and the corresponding target parameters can be obtained directly from the input parameters [3, 4].

To address the above problem, particle swarm algorithm is used to optimize neural network by encoding the initial value into particles, updating the position and velocity information of particle swarm algorithm, and obtaining the optimal particles after evolutionary iteration, i.e., the optimized initial weights and thresholds [5, 6]. The BP neural network for particle swarm optimization fully uses both local optimization-seeking ability of gradient descent algorithm and global heuristic search property of particle swarm algorithm, and the two algorithms complement each other to achieve stable improvement of the bearing optimization calculation process [7].

11.2 Hydrodynamic Journal Bearing Lubrication Analysis

11.2.1 Reynolds Equation

The Reynolds equation is the most fundamental formula in hydrodynamics theory, and general form of Reynolds equation is [8]:

$$\frac{\partial}{\partial\theta}\left(h^3\frac{\partial p}{\partial\theta}\right)+R_b^2\frac{\partial}{\partial y}\left(h^3\frac{\partial p}{\partial y}\right)=6\eta R_b\left(u\frac{\partial h}{\partial\theta}+2R_b\frac{\partial h}{\partial t}\right) \tag{11.1}$$

In Eq. (11.1), p is oil film pressure, h is oil film thickness, η is lubricant viscosity, $u = u_j + u_b$, $uj = R_j\omega_j$ is speed of bearing surface, R_j is radius of bearing, ω_j is journal rotational angular velocity, $u_b = R_b\omega_b$ is velocity of the bearing surface, R_b is bearing radius, ω_b is bearing rotational angular velocity. The *Reynolds* boundary conditions are applied. The *Reynolds* formula is solved by the finite difference method. The nodes inside the solution domain are in central difference format, and the nodes at the boundary of the solution domain (located at the front and rear end faces) are in front or back difference format along the axis of bearing.

11.2.2 Oil Film Thickness Equation

The oil film thickness equation is an important equation for calculating oil film pressure, which is used solve minimum thickness of film and determine capacity of film.

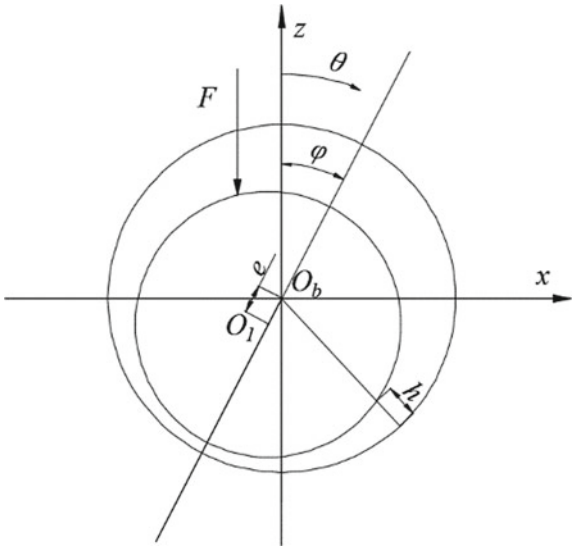


Fig. 11.1 Schematic diagram of journal bearing in the Cartesian coordinate system

In Fig. 11.1, φ is eccentricity angle and e is eccentricity distance. θ is the circumferential coordinate of bearing expansion, and $\theta = 0°$ is the vertical direction. Since the bearing clearance value c is much smaller than the bearing diameter, after a series of simplifications, the expression of oil film thickness is:

$$h \approx c + e\cos(\theta - \phi) \tag{11.2}$$

11.2.3 Bearing Oil Film Counterforce

The bearing oil film reaction force is force exerted by hydrodynamic pressure generated by wedge oil film on bearing bush during the high-speed operation of the sliding bearing. The oil film reaction force is an important parameter to verify whether the sliding bearing can operate normally.

$$F_x = -\int_0^L \int_{\theta_1}^{\theta_2} pR\sin\theta \mathrm{d}\theta \mathrm{d}y \tag{11.3}$$

$$F_z = -\int_0^L \int_{\theta_1}^{\theta_2} pR\cos\theta \mathrm{d}\theta \mathrm{d}y \tag{11.4}$$

$$F = \sqrt{F_x^2 + F_z^2} \tag{11.5}$$

For a finite width bearing, let the oil film reaction force in the horizontal be F_x and film reaction force in the vertical direction be F_z,θ is the circumferential coordinate of bearing expansion, R is bearing radius. The component of the oil film reaction is obtained by integrating film pressure.

11.2.4 Discharge Flow

Discharge flow rate Q refers to the volume of lubricating oil leaked from both ends of a sliding bearing during operation. This physical quantity represents the direct loss of lubricating oil during the bearing's use. The more lubricating oil loss, the worse the bearing's lubrication performance.

The flow rate of lubricant from the front face and rear face of the bearing is:

$$Q_1 = -\int_0^{2\pi} \frac{h^3}{12\eta} \frac{\partial p}{\partial y}\Big|_{y=0} \mathrm{d}\theta \tag{11.6}$$

$$Q_2 = -\int_0^{2\pi} \frac{h^3}{12\eta} \frac{\partial p}{\partial y}\Big|_{y=L} \mathrm{d}\theta \tag{11.7}$$

$$Q = Q_1 + Q_2 \tag{11.8}$$

The variables involved in the equation are mentioned earlier.

11.2.5 Friction Power Loss

Friction power consumption N_F represents the energy loss during the operation of sliding bearings, and the magnitude of friction power consumption directly affects the lubrication effect of bearings.

$$F_j = \int_0^L \int_0^{2\pi} \left(\frac{h}{2} \frac{\partial p}{R\partial \theta} + \frac{u\eta}{h} \right) R \mathrm{d}\theta \mathrm{d}y \tag{11.9}$$

$$N_F = F_j \cdot u \tag{11.10}$$

F_j is bearing surface friction force, and u is linear velocity of bearing surface.

11.2.6 Solution Method

The *Reynolds* equation is solved by the finite difference method [9]. The nodes within the solution domain are in the central difference format, and the nodes at the boundary of the solution domain (located at the front and rear end faces of the bearing) are in the forward or backward difference format along the bearing axis. The solution domain is taken as the full length of bearing in axial and 360° in the circumferential direction, and is divided into equally spaced grids along the circumferential and axial directions. Grid division of bearing surface along the plane is shown in Fig. 11.2. The circumferential direction $\theta = 0°$ to 360° is divided into m nodes, i.e., $i = 1$ to m, with $i = 1$ corresponding to $\theta = 0°$ and $\theta = 360°$; the axial direction $y = 1$ to B is divided into n nodes, i.e., $j = 1$ to n (Fig. 11.3).

Calculation of bearing load capacity, end leakage flow, and friction power consumption in the formula for the integration of the application of Simpson's formula for numerical integration, the partial derivative of the application of the four-point difference formula.

11.3 Mathematical Model for Optimal Design of Hydrodynamic Journal Bearing

The objective function is an optimization model of frictional power consumption with design variables: eccentricity ε, gap ratio ψ, width–diameter ratio λ, misalignment coefficient m_1, and preload coefficient m [10].

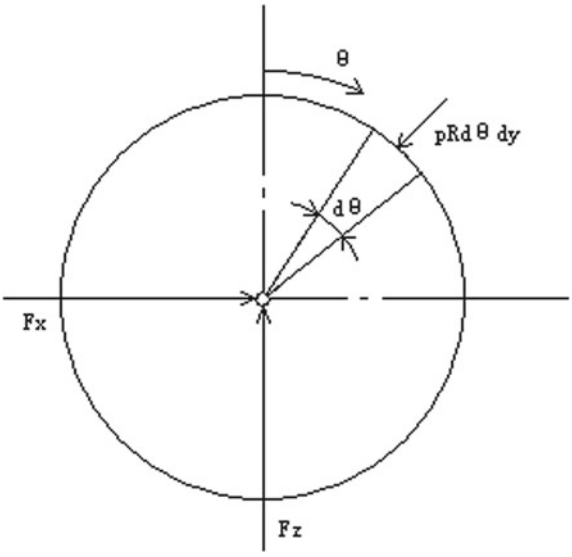


Fig. 11.2 Cross-section diagram of bearing

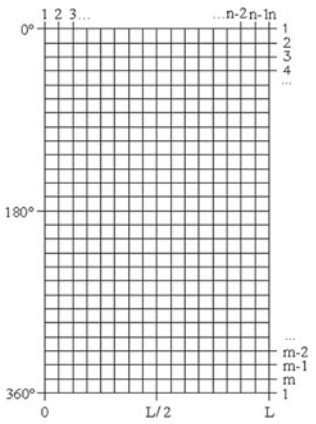


Fig. 11.3 Meshing

$$X = [\varepsilon, \psi, \lambda, m_1, m]^{\mathrm{T}} = [x_1, x_2, x_3, x_4, x_5]^{\mathrm{T}} \tag{11.11}$$

The constraint conditions are as follows.

11.3.1 Bearing Minimum Oil Film Thickness $\mathbf{h}_{min}$ Constraint

In Eq. (11.11), $[h_{\min}]$ is the permissible minimum oil film thickness of the bearing, and is mainly considered sum of arithmetic mean deviation of the journal and bearing working surface roughness.

$$g_1(X) = h_{\min} - [h_{\min}] \geq 0 \tag{11.12}$$

11.3.2 Bearing Width-to-Diameter Ratio Λ

In Eq. (11.12), A_{L} is the minimum value of width-to-diameter ratio λ; A_{U} is the maximum value of width-to-diameter ratio λ. The values of A_{L} and A_{U} are determined according to the specific conditions of the designed bearing.

$$g_2(X) = A_{\mathrm{U}} - \lambda \geq 0 \tag{11.13}$$

$$g_3(X) = \lambda - A_{\mathrm{L}} \geq 0 \tag{11.14}$$

11.3.3 Average Specific Pressure of the Bearing $\overline{p}$

In Eq. (11.14), $[\overline{p}]$ is the permissible specific pressure of the bearing, and its value is determined according to the specific requirements of the design bearing.

$$g_4(X) = [\overline{p}] - \overline{p} \geq 0 \tag{11.15}$$

11.3.4 Bearing Clearance Ratio Ψ Constraint

In Eq. (11.15), ψ_{U} is the maximum value of bearing clearance ratio and ψ_{L} is the minimum value of bearing clearance ratio. ψ_{U} and ψ_{L} take values according to the specific requirements of the designed bearing to determine.

$$g_5(X) = \psi_{\mathrm{U}} - \psi \geq 0 \tag{11.16}$$

$$g_6(X) = \psi - \psi_{\mathrm{L}} \geq 0 \tag{11.17}$$

11.3.5 Bearing Lubricant Flow Q Constraint

In Eq. (11.17), $[Q]$ is the permissible (required) bearing lubricant flow rate, the value of which is determined according to the specific requirements of the designed bearing.

$$g_7(X) = [Q] - Q \geq 0 \tag{11.18}$$

11.3.6 Bearing Lubricant Temperature Rise Δt

In Eq. (11.18), $[\Delta t]$ is the permissible (required) bearing lubricant temperature rise, the value of which is determined according to the specific requirements of the designed bearing.

$$g_8(X) = [\Delta t] - \Delta t \geq 0 \tag{11.19}$$

11.3.7 *Bearing Load Capacity* F

In Eq. (11.19), $[F]$ is the required bearing load capacity, the value of which is determined according to the specific requirements of the designed bearing.

$$g_9(X) = F - [F] \geq 0 \tag{11.20}$$

Mathematical model:

$$\begin{aligned} &X = [x_1, x_2, x_3, x_4, x_5]^{\mathrm{T}} \\ &\text{s.t. } g_i(X) \geq 0 (i = 1, 2, \ldots, 9) \end{aligned} \tag{11.21}$$

11.4 Hydrodynamic Journal Bearing Algorithm Coupling

Bearing lubrication analysis calculations are very complex and time consuming, and the optimization requires solving Reynolds [11] equation for different design variables to obtain the corresponding lubrication performance data (friction power consumption, minimum oil film thickness, average specific pressure, load-carrying capacity, lubricant flow rate, lubricant temperature rise, dynamic characteristic coefficient, etc.) [12], which will account for a long calculation time and thousands of times in one optimization process.

The method used in the paper is that a particle swarm algorithm is used to optimize neural network, and the initial values are encoded into particles, and the particle swarm algorithm is used to update position and velocity information of particles, and optimal particles are obtained after evolutionary iteration, i.e., the optimized initial weights and thresholds.

11.4.1 *Particle Swarm Algorithm*

Basic thought of the particle swarm algorithm is to randomly initialize particles with no mass, treating each particle as solution to the optimization problem, with goodness of the particle decided by a predetermined fitness function. Each particle will move in the space of feasible solutions, and its direction and distance will be determined by a velocity variable. Particle will follow the current optimal particle and finally get solution after a generation-by-generation search. In each iteration, particle will track two results: one is solution found so far by the particle itself, called individual optimum (p_{best}), and the other solution found so far by the whole, called global optimum (g_{best}). Suppose a population of M particles in the D-dimensional search space at presets speed, and the state properties of particle i at moment t as follows.

location: $x_i^t = \left(x_{i1}^t, x_{i2}^t, \ldots x_{id}^t\right)^T$.
$x_{id}^t \in$ [Ld, Ud], Ld, Ud are the lower and upper limits of the search space.
Velocity: $v_i^t = \left(v_{i1}^t, v_{i2}^t, \cdots v_{id}^t\right)^T$.
$v_{id}^t \in [v_{\min}, d, v_{\max}, d]$, $v_{\min}$, $v_{\max}$ are the minimum and maximum velocity.
Individual optimal position: $p_i^t = \left(p_{i1}^t, p_{i2}^t, \cdots p_{id}^t\right)^T$.
Global optimal position: $p_g^t = \left(p_{g1}^t, p_{g2}^t, \cdots p_{gd}^t\right)^T$, $1 \le d \le D$, $1 \le i \le \mathrm{M}$.

The position of the particle at the moment $t + 1$ is obtained by updating the following equation.

$$v_{id}^{t+1} = wv_{id}^t + c_1 r_1\left(p_{id}^t - x_{id}^t\right) + c_2 r_2\left(p_{gd}^t - x_{id}^t\right) \tag{11.22}$$

$$x_{id}^{t+1} = x_{id}^t + av_{id}^{t+1} \tag{11.23}$$

where a is the step size, r_1, r_2 are random numbers uniformly distributed in the (0, 1) interval; c_1, c_2 are called learning factors, usually take $c_1 = c_2 = 2$.

The w is inertia parameter, and its size determines how much the particle inherits the current velocity, and choosing a suitable w helps the particle population to equalize its exploration ability. Larger inertia weights facilitate the unfolding of the global optimum, while smaller inertia weights facilitate the local search for the optimum. Therefore, a particle swarm algorithm with inertia weights can better guide the particles to jump out of optimal solution and improve convergence of the algorithm. Inertia weights based on the form of exponential curve. Its expression is as follows

$$w = ww = w_{\text{end}}(w_{\text{start}}/w_{\text{end}})^{1/(1+10t/T)} \tag{11.24}$$

where w_{start} is the maximum inertia parameter, w_{end} is the minimum inertia parameter at the end of the iteration, t is the current number of iterations, and T is total number of iterations.

Equation (11.12) consists of three main contents: the first part is the inheritance of the particle's previous velocity, which represents the particle's trust in the current state of its own motion and inertial motion based on its own velocity; the second part is the cognitive part, which represents the particle's own thinking, i.e., the comprehensive consideration of its own previous experience to achieve the next behavioral decision, which is cognitive and reflects an increasing learning process. Third part is the social sector, which represents the information sharing and mutual cooperation among particles. During the search, particles remember results while considering the experiences of their peers. When a single particle perceives that the experience of its peers is better, it will make adaptive adjustments and seek a consistent cognitive process.

11.4.2 Neural Network Modeling Framework

The BP neural network includes three parts: the input layer, the hidden layer, and the output layer, and the structure is shown in Fig. 11.1.

Training and learning process consists of two stages: forward propagation and backward propagation of errors. In the forward propagation process of BP neural network, the input vector of the input layer is $x = (x_1, x_2,..., x_n)$, and the weights and thresholds between the input and hidden layers are w_{ij} and b_{1i}, then the input net_i of the i neuron in the hidden layer is

$$\text{net}_i = \sum_{j=1}^{M} w_{ij} x_j + b_{1i} \tag{11.25}$$

F(-) is the activation function, and the output O_i of the ith neuron in the hidden layer is.

$$O_i = f(net_i) = f\left(\sum_{j=1}^{M} w_{ij} x_j + b_{1i}\right) + b_{2k} \tag{11.26}$$

The input *netk* of the k-th neuron in the output layer is

$$\text{net}_k = \sum_{i=1}^{M} w_{ki} f\left(\sum_{j=1}^{M} w_{ij} x_j + b_{1i}\right) + b_{2k} \tag{11.27}$$

The output *yk* of the k-th neuron of the output layer is

$$\text{y}_k = \text{f}(\text{net}_k) = f\left(\sum_{i=1}^{M} w_{ki} f\left(\sum_{j=1}^{M} w_{ij} x_j + b_{1i}\right) + b_{2k}\right) \tag{11.28}$$

In the back-propagation process of the error, the neural network undergoes forward propagation to obtain the output y_k of the output layer, and the output error E is obtained by calculating with the expected output T_k.

$$E = \frac{1}{2} \sum_{k=1}^{N} (T_k - y_k)^2 \tag{11.29}$$

The BP neural network in error adjusts parameters and thresholds in negative gradient direction according to descent algorithm to gradually reduce the error until condition is satisfied. The adjustment of the output layer weights and thresholds are Δw_{ki} and Δb_{2j}, respectively, and similarly, the adjustment of the hidden layer

weights and thresholds are Δw_{ij} and Δb_{1i}, respectively.

$$\Delta w_{ki} = -\eta \frac{\partial E}{\partial w_{ki}} \ \Delta b_{2k} = -\eta \frac{\partial E}{\partial b_{2k}} \ \Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \ \Delta b_{1j} = -\eta \frac{\partial E}{\partial b_{1j}} \quad (11.30)$$

η is the learning rate of the back-propagation process of the BP neural network. The topology of the BP neural network includes the number of layers of hidden layers and the number of neurons per layer. The number of neurons in the input and output layers has been determined according to the input and output parameters, and the number of neurons in each hidden layer of is determined by the following empirical formula.

$$N = \sqrt{I + O} + A \quad (11.31)$$

In Fig. 11.4, I is the number of neurons in the input layer; O is number of neurons in the output layer; A is a value between 1 and 10. To predict the friction coefficient by BP neural network, it is important to first determine the topology of the neural network with the same input parameters as the minimum film thickness prediction model, so the number of input neurons is 8 and the output parameter is the friction coefficient, i.e., the number of output neurons is 1.

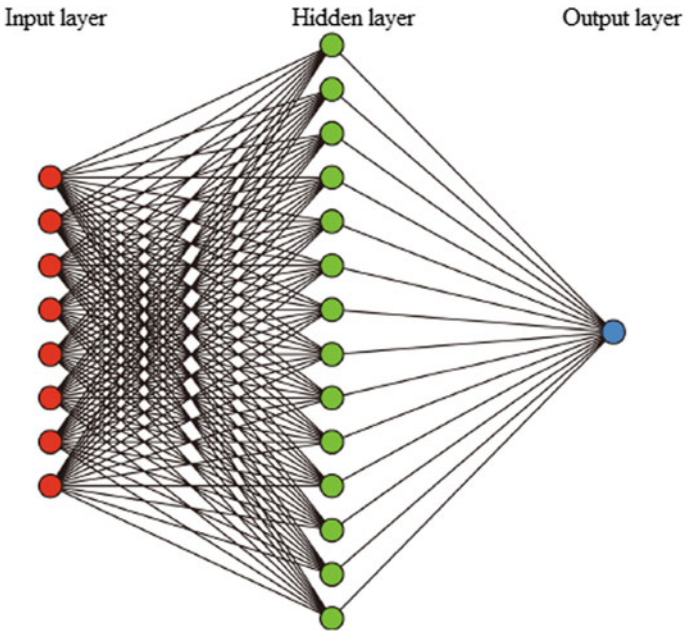


Fig. 11.4 Principle of neural network computation

11.5 Calculation Examples

11.5.1 Verify the Accuracy of the Calculation

In order to verify the effectiveness of the proposed variable-parameter particle swarm optimization algorithm applied to the hydrodynamic journal bearing optimization design problem, a round bearing is used as an example for computational analysis and comparison. Parameters of the circular bearing: diameter 0.05 m, eccentricity 0.2/0.5, bearing speed 2000/3000/4000 r/min, and the objective function is the lubricant flow. Each constraint: bearing load P is 300 N, permissible minimum oil film thickness $h_{\min}$ is 0.02 mm, minimum value of width–diameter ratio is 0.6, maximum value of width–diameter ratio is 1, permissible specific pressure is 0.21 MPa, minimum value of clearance ratio is 0.001, maximum value of clearance ratio is 0.002, permissible flow rate is 0.04 L/min, permissible frictional power consumption is 0.04 kW, and permissible lubricant temperature rise is 30 °C.

Comparing Tables 11.1, 11.2, 11.3 and 11.4, 11.5, 11.1.6, it can be seen that under the condition of changing the speed with constant eccentricity, the traditional algorithm and the coupled algorithm calculate the results under various conditions, respectively, among which, comparing the results of lubrication flow, maximum oil film pressure, and surface friction, it is found that the relative error of the traditional algorithm and the coupled algorithm is less than 1%, while the relative error of is less than 2%.

Similarly, comparing Tables 11.1, 11.2, 11.3, 11.4, 11.2, 11.3, 11.4, 11.5, 11.6, and 11.3, 11.4, 11.5, 11.6, it can be obtained that changing the eccentricity at the same speed, for lubrication flow rate, maximum oil film pressure, and surface friction,

Table 11.1 Eccentricity 0.2 rotation speed 2000 r/min calculation results

Algorithm type	Q (m^3/s)	p_{max} (MPa)	Fj (N)	Δt (°C)
PSA	1.71E-06	1.3	33.54	10.29
PSANN	1.71E-06	1.29	33.52	10.33

Table 11.2 Eccentricity 0.2 rotation speed 3000r/min calculation results

Algorithm type	Q (m^3/s)	p_{max} (MPa)	Fj (N)	Δt (°C)
PSA	2.32E-06	1.5	42.62	16.85
PSANN	2.32E-06	1.48	41.99	16.97

Table 11.3 Eccentricity 0.2 rotation speed 4000r/min calculation results

Algorithm type	Q (m^3/s)	p_{max} (MPa)	Fj (N)	Δt (°C)
PSA	2.95E-06	1.75	52.22	21.00
PSANN	2.95E-06	1.74	52.07	20.98

Table 11.4 Eccentricity 0.5 rotation speed 2000r/min calculation results

Algorithm type	Q (m^3/s)	p_{max} (MPa)	Fj (N)	Δt (°C)
PSA	4.30E-06	5.14	37.50	12.42
PSANN	4.30E-06	5.15	37.54	12.64

Table 11.5 Eccentricity 0.5 rotation speed 3000r/min calculation results

Algorithm type	Q (m^3/s)	p_{max} (MPa)	Fj (N)	Δt (°C)
PSA	5.85E-06	6.22	48.58	18.43
PSANN	5.85E-06	6.24	48.63	18.47

Table 11.6 Eccentricity 0.5 rotation speed 4000r/min calculation results

Algorithm type	Q (m^3/s)	p_{max} (MPa)	Fj (N)	Δt (°C)
PSA	7.50E-06	6.93	55.66	26.35
PSANN	7.50E-06	6.91	55.59	26.73

the relative errors of the conventional algorithm and the coupled algorithm are less than 1%, and the relative errors of oil film temperature rise are less than 3%. For judging the lubrication status of the bearing, the maximum oil film pressure is the key judgment factor, and the relative error between the calculation results of the coupled algorithm and the traditional algorithm and the actual measured value is less than 10% under the six working conditions. The relative error between the minimum film thickness calculated by both methods and the measured value is also less than 10%, which meets the actual engineering needs and proves that the coupled algorithm can replace the traditional algorithm in practical applications.

In the process of calculating lubrication flow, maximum oil film pressure, and other parameters, it needs suitable solution parameters, such as relaxation factors, in order to obtain convergence results with high accuracy, if the relaxation factor is not selected appropriately, on the one hand, it may lead to pressure, temperature, and other parameters in the iterative process of dispersion, cannot obtain the calculation results; on the other hand, even if the calculation error reaches the convergence accuracy, but the convergence process. On the other hand, even if the calculation error reaches the convergence accuracy, the convergence process is long, there is a certain fluctuation phase, and the calculation efficiency is not high (Fig. 11.5).

As the number of bearing meshes increases, the calculation accuracy of the bearing increases exponentially, resulting in an increase in the calculation volume of the traditional and coupled algorithms. Therefore, this paper compares the computational efficiency of the traditional algorithm and the coupled algorithm considering the increase of the number of surface meshing of the bearing.

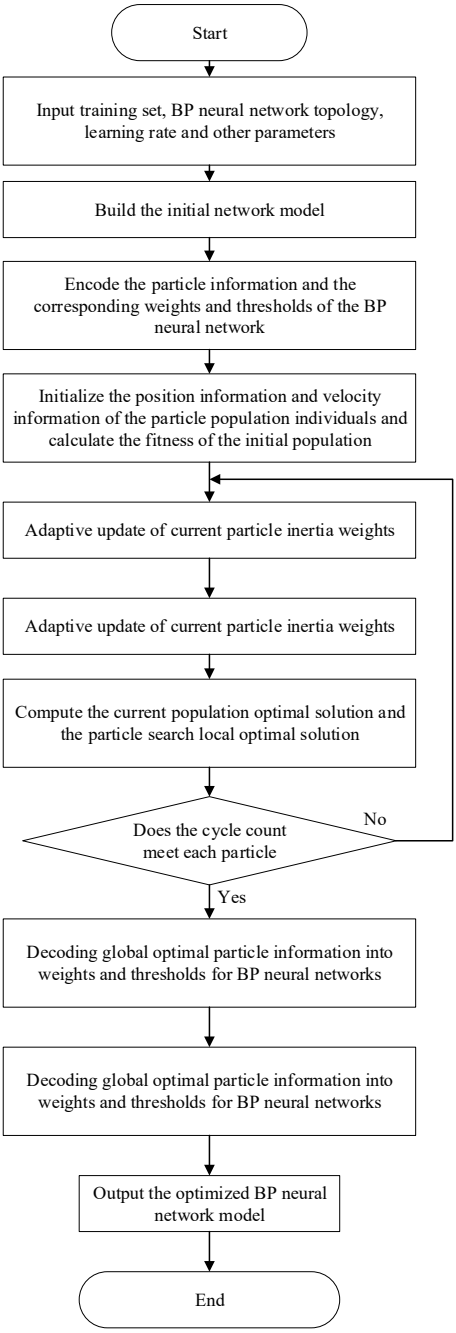


Fig. 11.5 Calculation process

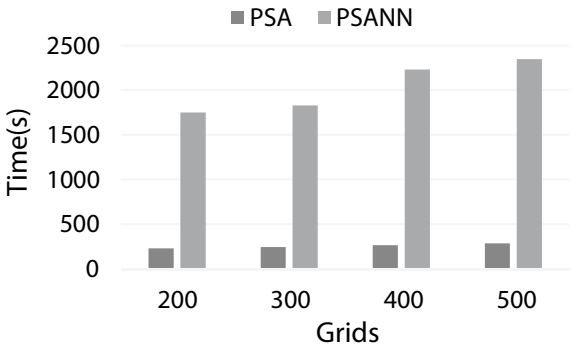


Fig. 11.6 Comparison of computational efficiency with different grid numbers

11.5.2 Validate Computational Efficiency

As shown in Fig. 11.3, in the process of increasing the number of grids from 200 to 300, the computation time of the traditional algorithm increases by 590 s; the computation time of the coupled algorithm increases by 92 s, and the time change of the coupled algorithm is more gentle. Under the test conditions of this paper, the number of grids is 500, the coupling algorithm can save 2050s compared with the traditional algorithm. Based on this test result, it can be inferred that the numerical calculation of lubrication analysis may have singularity problem and convergence difficulty under the case of high precision of bearing calculation, and the advantage of high efficiency of coupling algorithm calculation will be more and more prominent. It can be seen that improving the computational efficiency of the basic lubrication analysis is of great significance to the dynamic performance calculation of the cascading digital rotor system (Fig. 11.6).

11.6 Conclusion

(1) Initial weights and thresholds of neural network were optimized by the particle algorithm, and the optimization model for the optimization of the particle algorithm was determined. The neural network models for calculating lubrication flow, maximum oil film pressure, oil film temperature rise, and surface friction were established, respectively, and the computational performance of the neural network models was evaluated. Judging from the results of multiple indicators, it was proved that it is feasible to use the neural network method to couple the particle swarm algorithm to calculate the relevant parameters.

(2) The accuracy of the coupling algorithm is analyzed by changing the working parameters of the bearing, and the error of the calculation results of the performance parameters is less than 2% regardless of whether the real eccentricity is changed with constant speed or the speed is changed with constant eccentricity.

(3) In the case of increasing the number of bearing mesh division and improving the calculation accuracy of the algorithm, the calculation time of the traditional particle algorithm has increased, while calculation time of neural network coupled particle swarm algorithm still remains at a low level, and the calculation efficiency of the coupled algorithm is significantly higher than that of the traditional particle swarm algorithm.
(4) Overall, the neural network coupled particle swarm algorithm will have a wide range of applications in the future for its comprehensive computational performance.

References

1. Saeed, N.A., Kandil, A.: Two different control strategies for 16-pole rotor active magnetic bearings system with constant stiffness coefficients. Appl. Math. Model. **92**, 1–22 (2021)
2. Iseli, E., Schiffmann, J.: Experimental and numerical investigation of the unbalance behavior of rigid rotors supported by spiral-grooved gas journal bearings. Mech. Syst. Signal Process. **174**, 109080 (2022)
3. Argatov, I.I., Chai, Y.S.: An artificial neural network supported regression model for wear rate. Tribol. Int. **138**, 211–214 (2019)
4. Zakaulla, M., Parveen, F., Amreen, Harish, Ahmad, N.: Artificial neural network based prediction on tribological properties of polycarbonate composites reinforced with graphene and boron carbide particle. Mater. Today: Proc. **26**(Part 2), 296–304 (2020)
5. Zhao, C., Sun, J.L., Lin, S.L., Peng, Y.: Rolling mill bearings fault diagnosis based on improved multivariate variational mode decomposition and multivariate composite multiscale weighted permutation entropy. Measurement **195**, 111190 (2022)
6. Tomar, A.K., Sharma, S.C.: Non-Newtonian lubrication of hybrid multi-recess spherical journal bearings with different geometric shapes of recess. Tribol. Int. **171**, 107579 (2022)
7. Hu, T., Guo, Y.M., Gu, L.D., Zhou, Y.F., Zhang, Z.S., Zhou, Z.T.: Remaining useful life estimation of bearings under different working conditions via Wasserstein distance-based weighted domain adaptation. Reliab. Eng. Syst. Saf. **224**, 108526 (2022)
8. Wen, S.Z.: Tribological Principle. Tsinghua University Publishing House Co., Ltd (2002)
9. Yu, C., Meng, X., Xie, Y.: Numerical simulation of the effects of coating on thermal elastohydrodynamic lubrication in cam/tappet contact. J. Eng. Tribol. **231**(2), 221–239 (2017)
10. Xiang, G., Yang, T.Y., Guo, J., Wang, J.X., Liu, B., Chen, S.A.: Optimization transient wear and contact performances of water-lubricated bearings under fluid-solid-thermal coupling condition using profile modification. Wear **502–503**, 204379 (2022)
11. Zhang, W., Deng, S., Chen, G., Cui, Y.S.: Impact of lubricant traction coefficient on cage's dynamic characteristics in high-speed angular contact ball bearing. Chin. J. Aeronaut. **30**(2), 827–835 (2017)
12. Questa, H., Mohammadpour, M., Theodossiades, S., Garner, C.P., Bewsher, S.R., Offner, G.: Tribo-dynamic analysis of high-speed roller bearings for electrified vehicle powertrains. Tribol. Int. **154**, 106675 (2021)

Chapter 12
The Intelligent Human–Computer Interaction Method for Application Software of Electrical Energy Metering Based on Deep Learning Algorithm

Weijie Zeng, Liman Shen, Wei Zou, Yeqin Ma, Songya Jiang, Mouhai Liu, and Libing Zheng

Abstract With the vigorous development and wide application of mobile communication device technology, various mobile communication devices have emerged. Aiming at the application of mobile communication equipment application software in the field of electrical energy metering, the intelligent human–computer interaction deep learning method of mobile communication equipment of application software for electrical energy metering based on embedded Linux is studied in this paper deeply. First, the architecture and working principle of application software for electrical energy metering are analyzed briefly. Secondly, the basic principle of neural network based on deep learning algorithm are analyzed, and then, the principle process of intelligent human–computer interaction of electric energy metering application software obtained through neural network depth learning is studied in detail. Finally, a simulation example is given. The results of theoretical and simulation studies show that the method studied in this paper has the advantages of simple human–computer interaction, real-time, accurate, and intelligent.

W. Zeng (✉) · M. Liu
Hunan Province Key Laboratory of Intelligent Electrical Measurement and Application Technology, Changsha, China
e-mail: 529454053@qq.com

W. Zeng · L. Shen · W. Zou · Y. Ma · S. Jiang · M. Liu
State Grid Hunan Electric Power Limited Company Power Supply Service Center (Metrology Center), Changsha, China

L. Zheng
Beijing Smartchip Microelectronics Technology Company Limited, Beijing, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_12

12.1 Introduction

With the rapid development of power system and science and technology, the electric energy metering management technology is constantly innovating, and the remote intelligent management of electric energy metering using mobile communication equipment comes into being.

Literature [1] points out the current problems of traditional managed energy metering. First of all, since most of the traditional energy metering is collected by the employees of power supply enterprises on site, this will lead to inaccurate energy measurement data, and it is inevitable that some users will seize this loophole to steal electricity, resulting in losses of power supply enterprises; Secondly, with the development of science and technology, the current equipment for energy measurement and verification has lagged behind significantly, resulting in errors in energy measurement data. The advantage of mobile communication equipment management energy measurement is that the energy measurement data can be viewed through mobile communication equipment, and the electrical appliances can be controlled through human–computer interaction on mobile communication equipment, and it is faster and more convenient to check the power level. Literature [2] refers to the advantages and disadvantages of human–computer interaction of mobile communication devices at present. As an instant messaging mobile device, the wireless communication function of mobile communication equipment is very portable, and in addition, mobile communication equipment has a variety of human–computer interaction methods, such as fingerprints, faces, voice, touch screens, and buttons. Literature [3] points out the advantages and disadvantages of common electric energy measurement applications (APPs) in the market at present. Nowadays, although most of the energy metering apps such as the "Online State Grid" APP [4] have been able to allow users to apply online for most of the energy metering services provided by power supply enterprises through the network, there are still problems such as opaque business data, slow response, insufficient intelligence level, and inability to realize human–computer interaction, which seriously affects the customer experience.

This paper proposes a deep learning method for human–computer interaction of energy metering APP based on embedded Linux. Firstly, it introduces the system architecture of embedded Linux [5] and the composition and functions of intelligent energy metering APP, and then analyzes the basic principles of deep learning algorithm based on neural networks. The principle process of human–computer interaction obtained by neural network deep learning is studied in detail. On this basis, the further study of the deep learning based on neural network of mobile communication equipment of electric energy metering device application software implementation process and the principle of intelligent human–computer interaction, the fingerprint identification intelligent human–machine interactive simulation, proved in this paper, based on the deep study of electric energy metering APP intelligent human–computer interaction method is feasible.

12.2 Composition and Function of Energy Metering APP

A complete energy metering App includes two parts: the server and the client, as shown in Fig. 12.1. The role of the server is to be responsible for reading the energy data of each power user from the energy meter and storing it in the database, and then moving it out, handling some logical problems, and providing it to the client. Therefore, databases and human–computer interfaces are an essential part of it.

The goal of the client is to facilitate users to access the server, view the user's electricity consumption information and perform operations such as payment, start and stop of electrical equipment, etc. Therefore, a good display interface and human–computer interaction interface are essential. Taking residential users as an example, the APP on this mobile communication device has functional modules such as electricity analysis, electricity billing, payment, and data download. Among them, the electricity bill introduces the monthly electricity and electricity usage of the residential user, including the ladder electricity usage, meter code, historical electricity consumption, payment records, etc. The electricity consumption analysis is a multi-dimensional electricity consumption analysis in the form of a histogram and a line chart, which clearly shows the use of the user's ladder electricity and the trend of electricity consumption in the past two years. In addition, there are recharge records, manual services, cloud systems, self-service payment, rich data categories, high-precision smart metering, and anti-theft. The highlight of this APP is that the power information can be visually analyzed through fingerprint recognition, which allows residents to clearly and transparently see the use of electricity, especially when the user's remaining power is insufficient, the APP will issue an alarm to remind the user to pay in time. For example, when a user wants to check his electricity bill balance and says, "Check my electricity balance", it will respond quickly, and the real-time balance can be displayed after completing fingerprint recognition verification. In the future, the smart assistant will continue to focus on user needs, and continue to launch

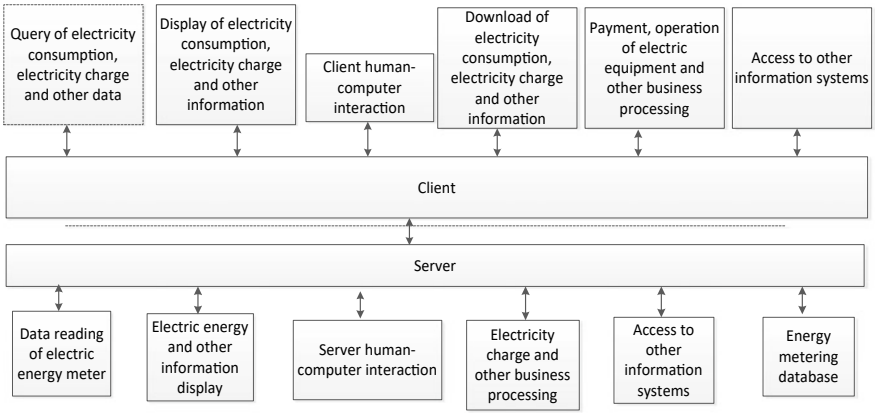


Fig. 12.1 Composition and function of intelligent electric energy metering APP

the forecast of electricity charges in the next few months according to users' recent electricity expenses and weather changes, thereby reminding users to pre-deposit electricity bills.

Therefore, human–computer interaction is an important part of smart energy metering APP, and the following studies several main human–computer interaction methods of smart energy metering APP.

12.3 Main Methods of Man–Machine Interaction of Energy Metering APP

Several main human–computer interaction modes of mobile communication devices are introduced, and their advantages and disadvantages are analyzed [6].

12.3.1 Classical Human–Computer Interaction Method

Classical human–computer interaction methods mainly include keypad, mechanical keyboard, software keyboard, and touch screen. Nowadays, the keys still play an irreplaceable role.

12.3.2 Intelligent Human–Computer Interaction Method

Keys and Software Keyboard

The human–computer interaction of keys is mainly achieved by gently pressing the keys on mobile communication devices with fingers. The software keyboard is a virtual keyboard realized by software based on touch screen. Nowadays, although the keys are more traditional and are gradually being replaced by software keyboards, they still play an irreplaceable role. Although these days face recognition and fingerprint recognition technologies are very mature, mobile communication devices still retain physical switches and volume control keys, which can be seen from the irreplaceable nature of keys.

Touch

Touch mainly refers to the feeling of contact between fingers and mobile communication devices. Fingerprint recognition is widely used, followed by handwritten character recognition. Through the fingerprint identification of user input, it can unlock mobile communication devices or applications, achieve the purpose of human–computer interaction, and has good confidentiality. However, haptic interaction has

its unique advantages, such as short reaction time, small space occupation, and convenience, so it has become the mainstream and accepted by the public.

Visual

Visual interaction mainly refers to the interaction between the eyes and the camera of the mobile communication device. Face recognition is widely used. After entering the features of the face, the camera of the mobile communication device tracks the subtle changes around the eyes, unlocks them, or issues other instructions, so as to achieve the purpose of human–computer interaction. It has good confidentiality.

Voice

Speech recognition interaction mainly refers to the process of interaction with mobile phones by giving instructions [6] through human voice. The advantage of language recognition lies in the simple and direct communication with the machine, and the efficiency is very high. The page has good confidentiality, so the public acceptance is high.

Expression and Awareness

Consciousness is at the stage of imagination and has not yet been put into practice.

The composition of the human–computer interaction system of the smart energy metering app is shown in Fig. 12.2. In view of the relatively simple principle of human–computer interaction between keys and software keyboards, the recognition principles of fingerprints, handwriting, faces, voices, and other methods of fingerprints, handwriting, faces, voices, and other methods are relatively complex and need to be in-depth researched. Taking fingerprint recognition as an example, this paper focuses on the human–computer interaction method of deep learning-based APP suitable for fingerprint, handwriting, face, speech, and other recognition.

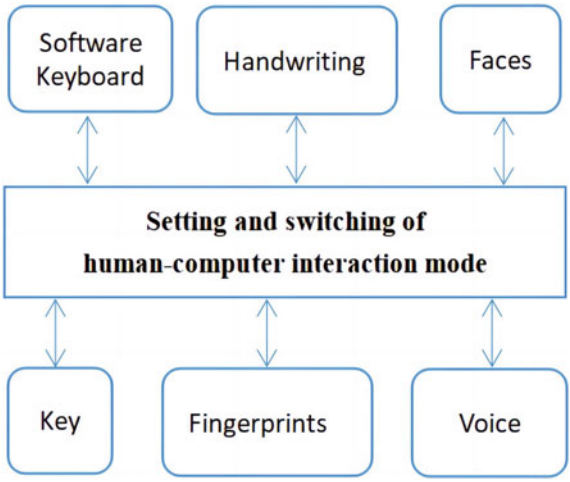


Fig. 12.2 Composition of human–computer interaction system for intelligent electric energy metering APP

12.4 Intelligent Human–Computer Interaction Principle of APP Based on Deep Learning

12.4.1 Neural Network Structure Types of Deep Learning

Deep learning is a special form of neural network, and the classical neural network consists of three layers: input layer, intermediate layer, and output layer [7]. When dealing with complex problems, multiple middle tiers may be required. This neural network structure with multiple intermediate layers is called a deep learning network.

In the deep learning network, there is no need to manually extract the feature structure by yourself, the network system will automatically extract the features and feature combinations of different pictures, and find effective features from them.

The neural network based on deep learning studied in this paper is composed of input layer, multiple intermediate layers, and output layer. The deep learning network will automatically extract the features and feature combinations of different pictures, and find effective features from them. The specific deep learning model realizes a variety of difficult to realize complicated functions through layers of intermediate layers. In this paper, the neural network with four intermediate layers is adopted:

Layer 1: Extract the most basic underlying features of the object
Layer 2: Arrange and combine the features of layer 1, and find useful combination information
Layer 3: Arrange and combine the features of layer 2, and find useful combination information
Layer 4: Arrange and combine the features of layer 3, and find useful combination information.

12.4.2 Deep Learning Principle Based on Convolutional Neural Network

The neural network used in this paper is a kind of deep learning neural network based on convolutional neural network, which is a feed-forward neural network in deep learning [7]. The feed-forward input is the semaphore, which is called feed-forward because the signal is conducted forward. Structurally, a convolutional neural network is a mapping network that can get the relationship between output and input without the need for complex and precise mathematical formulas. The design structure of the convolutional neural network can be shown in Fig. 12.3.

It can be seen from Fig. 12.3 that the convolutional neural network is a multi-layer network, each layer has multiple two-dimensional planes, and there are many neurons in each plane. In this paper, a shallow convolution neural network with four 2D convolutions is established.

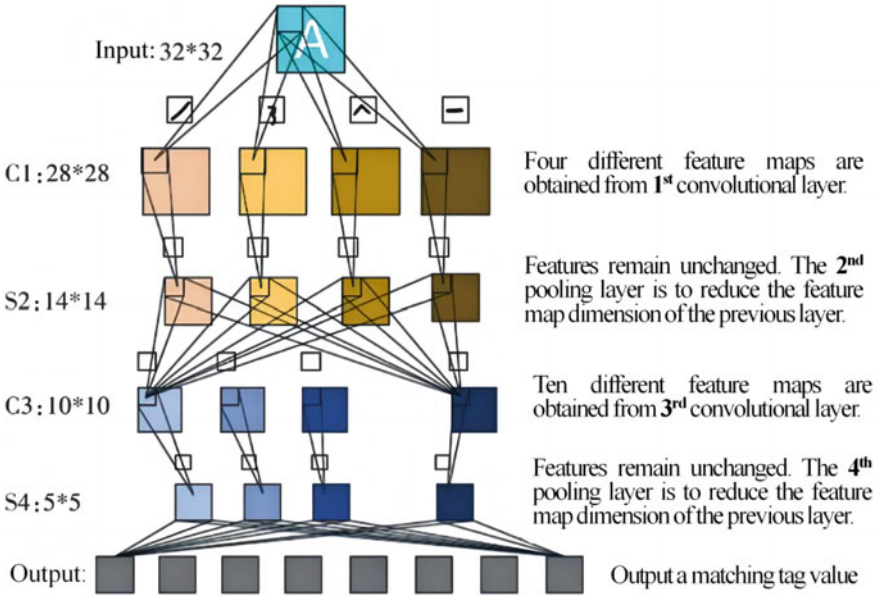


Fig. 12.3 Convolutional neural network structure diagram

The principle process of deep learning based on convolutional [8] neural network is as follows:

Convolution

The core idea of convolutional neural network is the existence of convolutional layer. The basic idea of convolution operation is to conduct inner product according to different window data and filter matrix [8]. Through the convolution operation, the convolution layer can obtain the feature image of the input data, highlight some features in these features, and filter some impurity information.

The pixel points in each image are mathematically numbered and represented in the form of a series of numbers, with $X_{i,j}$ referring to the data in the i-th row and the j-th column. Each weight in the filter is then data numbered. $W_{m,n}$ represents the data in the n-th column of the m-th row, and the bias term of the filter is expressed in W_b. Each data of the feature map is numbered, and $a_{i,j}$ represent the data of the i-th row and the j-th column of the feature map.

$$a_{i,j} = f\left(\sum_{m=0}^{2}\sum_{n=0}^{2} w_{m,n}x_{i+m,j+n} + w_b\right) \tag{12.1}$$

The convolution step is 1 in (12.1), which is the convolution under one-way communication channel. Since the convolution step has a certain impact on the final feature map, there is a specific calculation to determine the formula of the impact of

the convolution step on the feature map, as shown below.

$$W_1 = \frac{W_2 - F + 2P}{S} + 1 \tag{12.2}$$

$$H_2 = \frac{H_1 - F + 2P}{S} + 1 \tag{12.3}$$

W_1 the width of the graph before convolution;
W_2 the width of the graph after convolution;
F wide filter;
P the number of turns that make up 0 around the original data image;
S step size;
H_1 the height of the graph before convolution;
H_2 the height of the graph after convolution.

However, most of them use multi-channel convolution operations in general.

$$a_{d,i,j} = f\left(\sum_{m=0}^{F-1}\sum_{n=0}^{F-1}\sum_{d=0}^{D-1} w_{d,m,n} x_{d,i+m,j+n} + w_b\right) \tag{12.4}$$

D depth;
F height or width of the filter;
w weight;
a pixels.

Pooling

After the convolution operation of the convolution layer, the basic feature map of the input image is obtained. However, some interference parameters will be generated after convolution, so the dimension of the feature map obtained is high. The role of pooling layer is to further reduce the dimension of the obtained feature map after the convolution operation, so that when the new feature map has as many features as possible, it can reduce the dimension of the feature map.

Full Connection

Full connection is the operation process of spatial transformation of feature matrix, integrating the feature map data obtained after the operation of convolution layer and pooling layer, and reducing the dimension of high-dimensional feature map data.

12.4.3 Implementation Process of Deep Learning Algorithm Based on Convolutional Neural Network

Assuming a series of image inputs I, a system T is designed with n layers. By improving different parameters of each layer in T, the output is still input I, so that various features in I can be obtained. In terms of deep learning neural networks, the fundamental idea is to stack multiple intermediate layers between the input layer and the output layer, and the output of a certain layer in the middle serves as the input signal of the next layer [9]. In this way, it is a good way to realize the hierarchical expression of the input information.

In addition, it was previously directly assumed that the output of the previous layer is equal to the input of the next layer, but considering the data loss and deviation of the transmission process, the requirement that the output is strictly equal to the output can be slightly reduced, as long as the difference between input and output is as small as possible. The above is the basic idea of deep learning.

Figure 12.4 shows the flowchart of the deep learning algorithm based on convolutional neural network. The signal of the electric energy metering device is collected and stored, and the fingerprint stored in the electric energy metering device is extracted through convolutional neural network and stored in the fingerprint database as a storage signal. After the same fingerprint extraction, the fingerprint is compared in the same device, and the fingerprint is verified one-to-one with the existing fingerprint recorded in the fingerprint database for identity verification.

12.5 Principle of APP Fingerprint Recognition Human–Computer Interaction Based on Deep Learning

12.5.1 Deep Learning of Fingerprint Recognition Principle

This section proposes a multi-tasking full-depth convolutional neural network that extracts details from non-contact fingerprints, which can jointly learn the detection of detail positions and the calculation of detail directions [10]. Utilize attention mechanisms to make the network focus on detail points for directional regression.

Detail feature is a special pattern of ridge-valley staggered flow in fingerprint recognition, which is an important feature widely used in fingerprint recognition. There are several types of detail points, such as ridge ends and bifurcations. In this work, regardless of the type of minutiae, all minutiae are considered points of interest. Detail point extraction consists of two tasks: detail point position detection and detail point orientation calculation. Thus, a detail can be represented as a triplet (x_i, y_i, θ_i), where (x_i, y_i) represents its position and $\theta_i \in [0, 2\pi]$ represents its direction. There are dozens of points of detail in the fingerprint image.

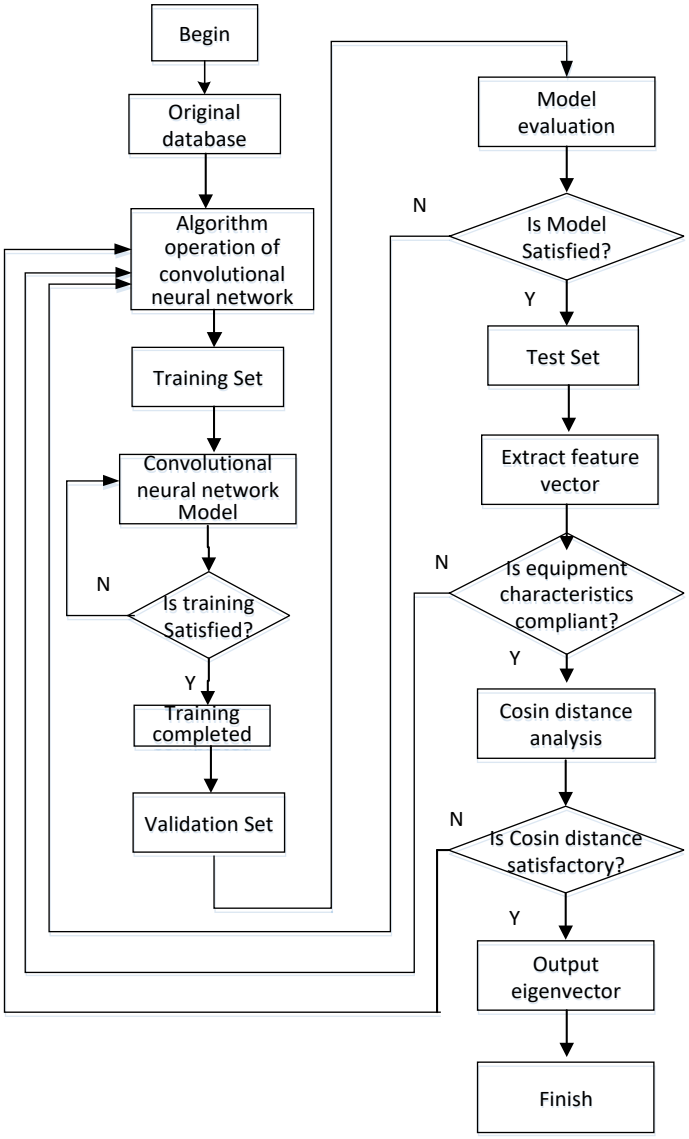


Fig. 12.4 Algorithm design flowchart

Detail extraction is often thought of as a type of object detection, where every detail is a target to be inspected. In the past few years, deep CNNs have been widely used for object detection. Therefore, detail extraction based on depth CNNs is usually divided into two steps: detecting detail points, and then calculating detail point direction [11]. The two-step method of training two deep networks separately is not only

time-consuming, but also does not make full use of the characteristics of position detection and direction calculation tasks.

To solve the above problems, a non-contact fingerprint detail extraction method based on multi-task full-depth convolutional neural network is proposed. Since detail position detection and orientation calculation are two related tasks, a multi-task deep network is proposed to learn these two tasks together and share their calculations and representations. Unlike the two-stage algorithm based on image chunks, this algorithm is based on a full-size fingerprint image, locates pixel-level detail points, and simultaneously calculates their corresponding orientation. Among them, it includes offline training and online testing phases.

In the offline stage [10–12], the training data is used to train a multi-task full-depth convolutional neural network, and the training data is composed of a full-size non-contact fingerprint and its corresponding detailed truth, which are used as the input and output of the network, respectively. Since there is currently no publicly available contactless fingerprint database, with marked granular coordinates and directions, marking details from scratch is time-consuming and labor-sensitive, so commercial fingerprint recognition software is used to extract candidate details and manually check them to generate basic facts about details. First, the COTS Verifinger SDK is used for detailed position and orientation extraction. A GUI tool was then developed to correct detail points and orientations in three ways: (a) adding new details that Verifinger missed; (b) removing false details; and (c) modifying coordinates and orientations that were marked incorrectly. Finally, the position and direction of human detection are used as the ground truth values of the training image.

In the online phase [13, 14], simply input a raw, full-size fingerprint into the trained network to generate two heat maps: one to detect detailed locations and the other to calculate their orientation. Local peaks on the location map are used as detail points, and the pattern map values on detection points are used as detail point directions.

12.6 Results of Fingerprint Recognition Simulation Based on Deep Learning

12.6.1 Fingerprint Identification Simulation

The template is designed, so that author affiliations is not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (e.g., do not differentiate among departments of the same organization). This template was designed for two affiliations.

In the same energy metering device, the fingerprint of the user in the input device is extracted, and then the fingerprint of the same person is extracted outside the device. The fingerprint in the device is compared with the fingerprint extracted again. The original figure and comparison figure are shown in Fig. 12.5.

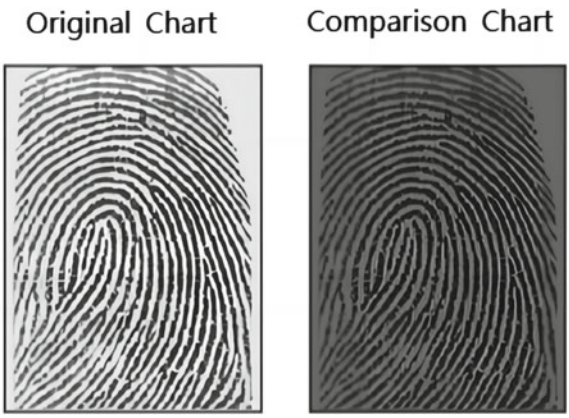


Fig. 12.5 Original fingerprint image and comparison image

Through the convolutional neural network algorithm, we can extract the feature map of the fingerprint, as shown in Fig. 12.6.

The feature map can be compared with the original map to improve the fingerprint recognition rate of power metering equipment. Figure 12.7 shows the recognition rate of fingerprint after using convolutional neural network algorithm.

In order to make the convolutional neural network algorithm used in this paper more comparable, the convolutional neural network algorithm will not be used in this paper to conduct the simulation again, and the simulation results are shown in Fig. 12.8.

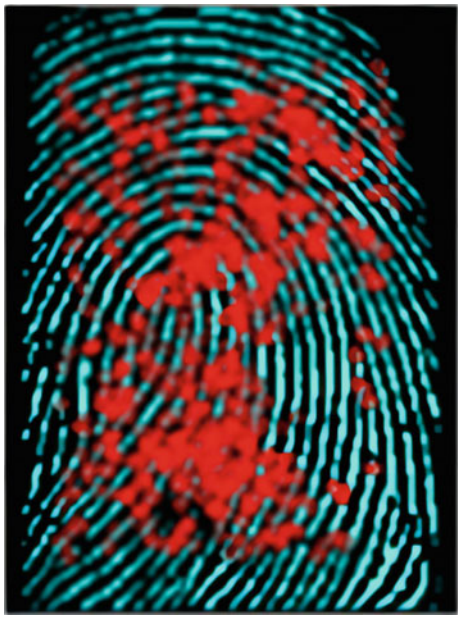


Fig. 12.6 Feature map of fingerprint

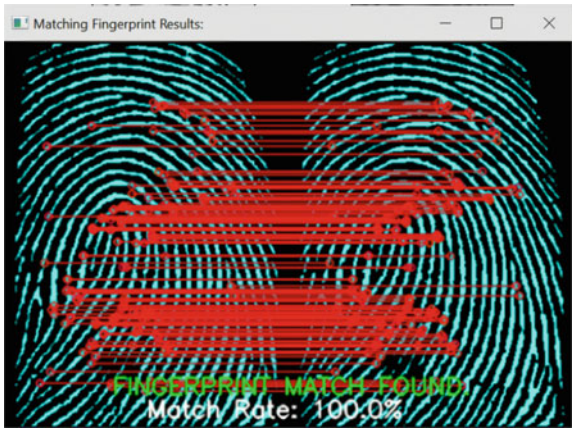


Fig. 12.7 Results of convolutional neural network algorithm

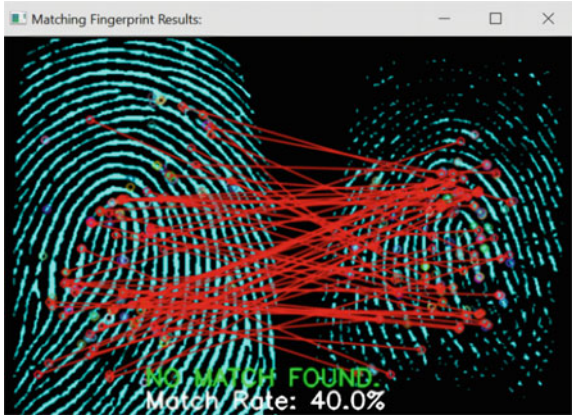


Fig. 12.8 Results of unused convolutional neural network algorithm

From the above simulation results, it can be seen that the fingerprint recognition algorithm based on deep learning convolutional neural network algorithm has higher accuracy in one-to-one fingerprint recognition, which is beyond the traditional point pattern matching algorithm. When classifying different fingerprints, the accuracy of the proposed algorithm is greatly improved.

12.7 Summary

In today era of big data, in order to achieve rapid recognition between human–computer interactions, a new human–computer interaction algorithm is proposed, which is a human–computer interaction recognition system based on deep learning

neural network [13]. This algorithm improves the recognition accuracy of human–computer interaction in the realization of human–computer interaction. Taking fingerprint recognition as an example, the experiment proves that the fault tolerance rate of traditional fingerprint recognition methods is low. After combining with the deep learning neural network, when using the convolutional neural network model, the fingerprint recognition rate of human–computer interaction has been better improved, with an error rate of about 4%.

Through simulation comparison, the recognition rate of traditional point pattern algorithm based on human–computer interaction is lower than that of convolutional neural network algorithm. But the training of convolutional neural network is very difficult to achieve, and a lot of experimental data are needed to provide data support for training. But the experimental results show that the human–computer interaction method based on deep learning neural network is promising in the future, and the deep learning algorithm will be more widely used in the field of human–computer interaction, but the experimental accuracy needs to be strengthened to further improve the accuracy.

Acknowledgements Supported by the State Grid Technology Project (5700-202055484A-0-0-00).

References

1. Shi, S., Peng, L., Cheng, L.: Current situation and prospect of electric energy measurement technology in China. Sci. Technol. Innov. (05), 44–45 (2019)
2. Xi, J.: Discussion on power measurement management problem and optimization strategy. Sci. Technol. Wind (18), 176 (2018)
3. Ding, F., Jiang, Z.; Research on human-computer interaction design in mobile devices. Packaging Eng. (16), 35 (2014)
4. Meng, F.: Improve the function of the "Online State Grid" app improve the service level of power measurement. Agric. Power Manage. **04**, 37–38 (2021)
5. Wu, X.: Design and implementation of embedded linux file system. Comput. Eng. Appl. (09), 111–112 (2005)
6. Tao, J., Wu, Y., Yu, C., Weng, D., Li, G., Han, T., Wang, Y., Liu, B.: Overview of multi-mode human-computer interaction. Chin. J. Image Graphics **27**(06), 1956–1987 (2022)
7. Li, B., Liu, K., Gu, J., Jiang, W.: A review of convolutional neural network research. Comput. Age (04), 8–12 (2021)
8. Nan, Z., Xinyu, O.: Development of convolutional neural network. J. Liaoning Univ. Sci. Technol. **44**(05), 349–356 (2021)
9. Li, B., Liu, K., Gu, J., Jiang, W.: A review of convolutional neural network research. Comput. Age (04), 8–12 (2021)
10. Zhang, Y.: Design and implementation of radio frequency fingerprint identification system based on deep learning. University of Chinese Academy of Sciences (National Space Science Center of Chinese Academy of Sciences), 000024 (2021)
11. Zeng, Y., Chen, X., Lin, Y., Hao, X., Xu, X., Wang, L.: Research status and trend of rf fingerprint identification. J. Radio Wave Sci. **35**(03), 305–315 (2020)
12. Chen, H., Li, X., Zheng, Y., Yuan, Shao, N., Yang, L., Liu, L.: Research on fingerprint recognition method based on deep learning. Intell. Comput. Appl. (03), 64–69 (2018)

13. Zhang, L.: Design of big data fingerprint identification system based on ARM and deep learning. J. Hunan Univ. Sci. Technol. (Nat. Sci. Edn.) **34**(01), 77–84 (2019)
14. Lv, Z., Huang, Z.: Fingerprint identification method and its application. Comput. Syst. Appl. (01), 34–36 (2014)

Chapter 13
Short Text Classification of Invoices Based on BERT-TextCNN

Jiuwei Zhang, Li Li, and Bo Yu

Abstract Traditional invoice text classification methods are labor-intensive and inefficient. In order to effectively identify the types of invoices, a Chinese text classification model based on deep learning BERT-TextCNN is designed, and a short text classification dataset of invoices is obtained from a municipal tax bureau to train and test the model, and to compare and analyze the performance of BERT-TextCNN model, BERT model, and TextCNN model. As a result, compared to traditional neural network models, the BERT + TextCNN model can accurately classify Chinese text, effectively prevent excessive fitting, and have good generalization ability. The performance of text classification is improved compared to both BERT model and TextCNN model alone. Draw a conclusion through experiments which show that the BERT-TextCNN model has good classification effect and good stability.

13.1 Introduction

The rapid development of computers, especially driven by online social networking, text data has gradually become a mainstream form of text. Due to the large amount of data and complex text semantics, text classification has become a challenge. Facing such a large and complex text data, it is especially important to classify them accurately and effectively. The length of text can be long or short, so we can classify these text data into short text data and long text data. Short text has the characteristics of short text content, easy to read, and easy to disseminate, and it exists widely in the Internet as a carrier of information dissemination and interaction, such as news headlines, social media information, invoice names, and other texts. Therefore, how to enable computers to classify large amounts of text data is becoming a topic of

J. Zhang (✉) · L. Li
College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang, China
e-mail: z1264641108@163.com

B. Yu
Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_13

interest to researchers. In general, text classification tasks have only few classes. When the classification task has a large number of classes, traditional Recurrent Neural Network (RNN) [1] (e.g., LSTM and GRU) algorithms perform poorly in terms of accuracy.

Therefore, in this article, we designed a BERT-based model and integrated its output into CNN to deal with classification matters. This method adopts the BERT Chinese model released by Google, which is pre-processed and then gets the word vector characteristics [2]. The feature of the word vector in the obtained sentence is the convolution kernel size from CNN. We combine the above two and use softmax to get the results. The reliability of this model in category task is demonstrated by comparing it with various text classification models.

13.2 Related Research

In the past, text categories were distinguished by plain Bayesian, KNN, decision tree, etc. With the rapid progress of deep learning, natural language processing technology has made rapid development. Deep neural networks are becoming a common method for text classification due to their powerful expressive power. Despite their attractiveness, neural text recognition models lack training data in many applications. In recent years, several Chinese classification methods emerge in an endless stream. Convolutional neural networks and recursive neural networks are applied to image processing and speech processing, and have made achievements. Later, they are applied to text processing technology. The first is to find a way to express words that can be recognized by computers that the computer will understand, making it possible for the computer to perform subsequent computation and analysis. The above is referred to as text representation. Word embedding is a kind of text representation that is often used. Words are put into the space, and these are expressed as vectors. One-hot, bag-of-words model, TF-IDF, etc., are the common text representation ways. But, the above method will result in problems, e.g., higher dimensionality and sparsity. They cannot explain two words with similar meanings very well. This is the reason why Word2vec 1 model emerged later. Word2vec has different models. The sequential grouping model is used to analyze the value at this time through the previous and subsequent articles. The sequential skip-gram model (Skip-gram) uses the value at this time to judge the meaning of the previous and subsequent articles. This way is associated with the previous and subsequent articles. The problem of too much computation and waste of resources is solved. Word2vec is not good at handling polysemy words. Word2vec is a static method, so it cannot be adjusted dynamically to enhance the specified things. Bidirectional Encoder Representations from Transformers (BERT) [3] is a pre-training model, which solves the problem of multiple meanings well by considering contextual information. BERT pays more attention to the early training of words, so it only needs to adjust the model according to different scenarios [4].

Liu et al. [5] proposed a multi-layer model construction, which can obtain the contents of previous and subsequent articles from the articles in a sequential manner. And LSTM is used to extract the contextual and sequential features of documents. The architecture of multi-layer models is more complex. It includes recursive neural networks such as LSTM, which require more computing resources and training time. The advantage of FastText text classification model is fast and efficient, but its direct use for distinguishing small text categories is of low accuracy rate. Feng Yong et al. proposed a method that fuses Term Frequency-Inverse Document Frequency (TF-IDF) and Implicit Dirichlet Distribution (IDD). Latent Dirichlet Allocation (LDA) for distinguish different texts [6]. The method performs TF-IDF filtering on the lexicon processed by the n-grammar model in the input stage of the FastText text classification model, performs corpus topic analysis using the LDA model, and complements the feature lexicon based on the obtained results, thus biasing the input word sequence vector mean in favor of highly discriminative entries and making it is more suitable for the environment where short and small texts are distinguished. Comparing the experimental results, it can be seen that the way has a higher accuracy rate in the classification of Chinese short texts. The application of TF-IDF and LDA is based on specific tasks and corpora, and may require adjustments and optimizations to each task and dataset. The generalization ability of this method may be relatively low, making it difficult to adapt to the needs of different tasks and domains.

Qiaohong Chen et al. proposed a novel text representation method to extract high-quality features from the entire training set by applying Gini impurity, information gain, and chi-square test from phrase features [7]. The phrase features extracted from each document must be linearly represented by these high-quality features, and then, after Word2vec word vector representation, advanced features are drawn out using convolutional neural network convolutional and pooling layers, and finally classified using Softmax classifier. This method depends on the feature selection of the whole training set in the feature extraction stage. This may lead to inaccurate feature selection in situations where the dataset is insufficient or unevenly distributed, affecting the final classification performance.

The attention mechanism is added to the text data coding, and the hierarchical structure of text classification is figured out. Attention mechanism is added to sentences and words, which is superior to long-term short-term memory (LSTM), CNN, and other models. Later, the transformer [8] model appeared, which abandoned the previous CNN and RNN, and the attention mechanism formed the entire network. It is a process of encoding and decoding, so this paper uses BERT. The BERT is used as an embedding [9] layer to access to other mainstream models and is trained and validated on the same invoice dataset. And it becomes one of the current mainstream models with good performance. Based on this, a network structure based on BERT-TextCNN is proposed in this paper, and a comparison experiment with BERT model and TextCNN model in the invoice text dataset is conducted.

13.3 Research Methodology

In Chinese text analysis, the model we propose is the BERT-TextCNN model. The model structure is shown in Fig. 13.1. In this pattern, words and phrases are embedded into a model, taking into account the pre prepared BERT model. Extract attachment vector from BERT and use it as input of CNN.

13.3.1 BERT

Contrasted with the ELMO model released by Google in 2018, BERT [3] changes the language pattern between BiLSTM and transformer [10], which truly implements the concept of bidirectional coding. The BERT model consists of a stack of encoders of transformer. This is a two-way encoding pattern. The input is the sum total token flushbonading, segment flushbonading, and location embedding. In addition, the output is a T-vector with feature message. The model composition of BERT is shown in Fig. 13.2.

In Fig. 13.2, E represents the embedding layer, which is the input layer of the BERT model. At this layer, each word or marker in the text sequence is converted into a

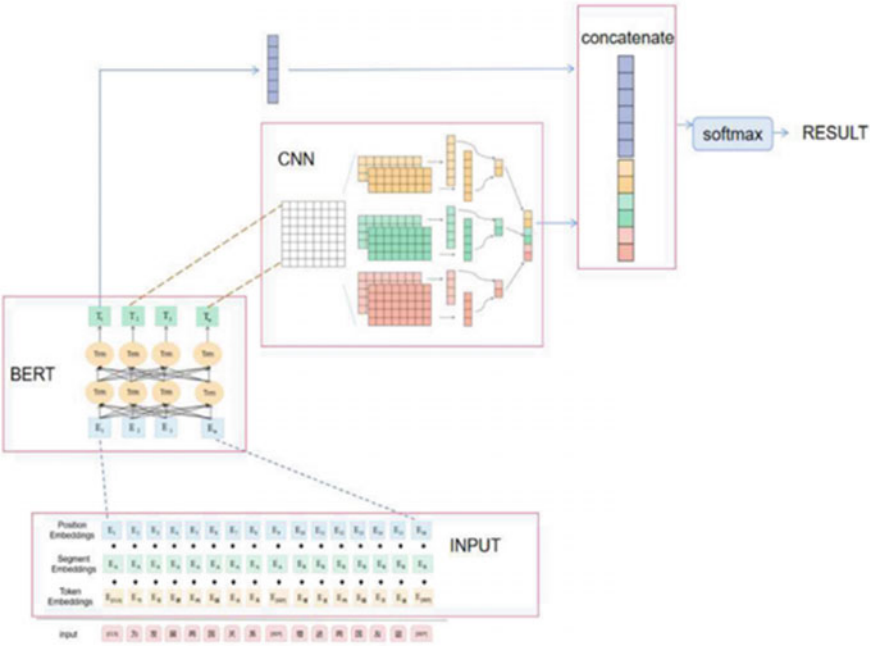


Fig. 13.1 BERT-TextCNN structure

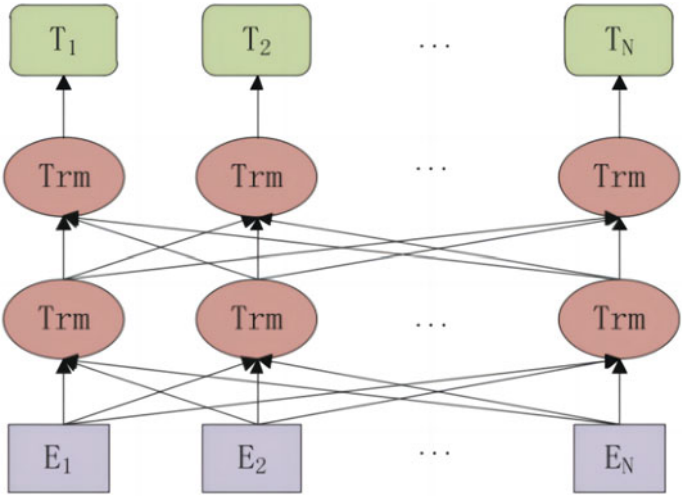


Fig. 13.2 BERT composition

corresponding vector representation, called an embedding vector. These embedding vectors will capture the semantic relationships and contextual information between words as input for subsequent processing. Trm represents a converter, which is a core component of the BERT model. BERT uses a multi-layer converter structure. Through the self-attention mechanism and the feedforward neural network layer, the input embedded vectors are encoded and feature extracted for multiple rounds. The transformer can capture contextual dependencies in text sequences and learn rich semantic feature representations.

The main purpose of the model is to generate a language model, so only multi-layer encoder construction is used. The encoder is mainly composed of feedback network layer and self-attention layer. If we want the computer to focus on some information, the attention mechanism can be implemented. Self-attention mechanism is to add the preceding and following words to the current word. This article is to add the features in the front and back of words to the features of words with different permissions, so that the computer can determine whether words in the sentence are more compactly connected with other words in the sentence.

In Fig. 13.3, multi-head attention is a self-attention mechanism used to capture the correlation between different positions in the input sequence. It maps the input sequence into multiple queries, keys, and values, and then aggregates the values by calculating attention weights. Multi-head attention allows the model to focus on different representation subspaces in the input sequence, thereby improving the model's expressive ability. Dropout is a regularization technique used to reduce model overfitting. During training, dropout randomly discards a portion of the output of neurons, making the model independent of specific features of individual neurons. This helps to improve the generalization ability and robustness of the model. Add represents adding the input to the output of the sub-layer in the residual connection.

In the encoder structure, after the self-attention and feedforward neural network sublayer, the residual connection will add the output of the sublayer to the input. This facilitates the flow of information and facilitates gradient propagation, promoting model training and convergence. Layer normalization is a normalization technique used to adjust the mean and variance of inputs at a hierarchical level. In encoder structures, layer normalization is usually followed closely by addition operations. It helps to alleviate the internal covariate offset problem and improve the stability and rate of convergence of the model. Feedforward is a sub-layer of the encoder structure, which processes the input by applying two linear transformations and nonlinear activation function. The feedforward neural network operates on the representation after position coding to extract higher level feature representation. It usually includes a hidden layer and an activation function, such as ReLU.

As we all know, attention mechanism can make the computer pay attention to the information that we want it to pay attention to. The self-attention mechanism is to integrate the context into the encoding of the current vocabulary. In this paper, the features of a word in a context are added to the features of the word with different weights, so that the computer can judge which words in a sentence are more closely related to the other words in the sentence.

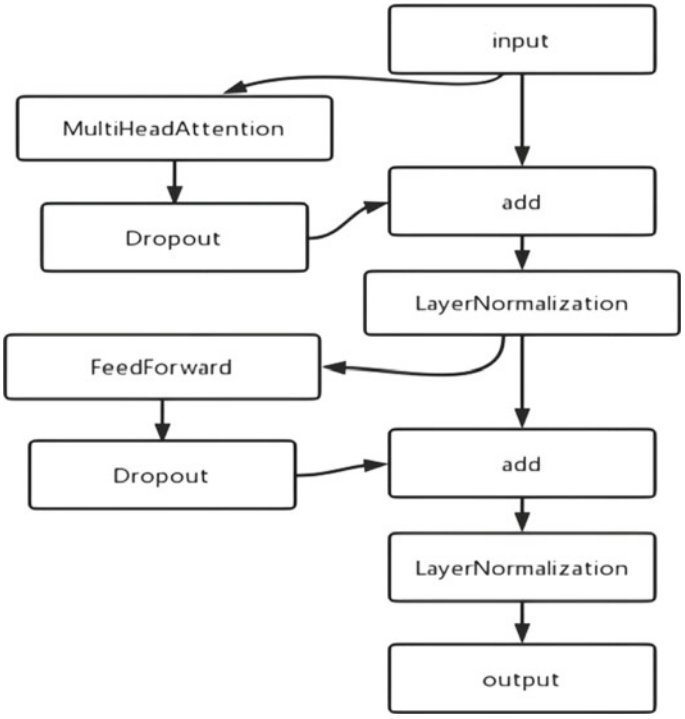


Fig. 13.3 Encoder structure diagram

The calculation of self-attention can be summarized as follows: first, prepare the input vector, and create a query vector, key vector, and value vector for each word. These vectors are obtained by multiplying the word embedding and three transformation matrices (W_Q, W_K, W_V), which are learned in training. Note that the dimensions of these new vectors are smaller than those of the input word vectors ($512 \rightarrow 64$), which is not necessary. This structure is intended to make the computation of the multi-headed attention more stable. Then, calculate the score, and calculate the self-attention of "Thinking" in "Thinking Matches". We need to calculate the score of "Thinking" for each word in the sentence, which determines the degree of attention paid to other parts of the sentence when encoding "Thinking". This score is obtained by calculating the dot product of the query vector of "Thing" and the key vector of other words. Second, divide the score by 8, so that the gradient will be more stable. Then, normalize the score by softmax to make the sum equal to 1. The softmax score determines how much each word pays attention to this position. Multiply the softmax score by the vector corresponding to value (to prepare for subsequent sum ups). The purpose of this is to retain the value of the concerned words and weaken the value of the irrelevant words (e.g., multiply by a small value of 0.001). Accumulate all weighted value vectors to produce the output result of self-attention at the location. Calculate the matrix of query, key and value, combine all input word vectors into matrix X, and multiply them by the trained weight matrix (WQ, WK, WV). The matrix is calculated as follows:

$$x \times w^q = q \tag{13.1}$$

$$x \times w^k = k \tag{13.2}$$

$$x \times w^v = v \tag{13.3}$$

$$Z = \text{soft}\max\left(\frac{q \times k^t}{\sqrt{d_k}}\right) \times v \tag{13.4}$$

In (13.4), the calculation result of matrix Q, K inner product shows the matching degree of the two vectors. After softmax function, the influence degree (weighted result) of the current word to the coding position can be obtained. Dividing by the root sign d_k is to prevent the score from expanding with the increase of dimensions. Without this step, softmax will get a smooth and indistinguishable result. Then, multiply the value matrix to get the self attention score of the current word. Finally, calculate all the words according to the above steps.

A set of Q, K, and V matrices can get a current word's eigenvalue through calculation. The multi-attention mechanism is like a filter in convolutional neural network, which can help us to extract a variety of features. As shown in the figure below, multiple feature expressions are obtained through different heads, all features are

spliced together, and finally, dimension reduction is carried out through the full connection layer. This algorithm uses 8 heads for feature stitching.

13.3.2 TextCNN

Convolutional neural network CNN [11] is used for graphics processing. As its variant model, text convolutional neural network (TextCNN) extracts local features of different sizes in text sequences by setting filters of different sizes. The convolution layer is more important in TextCNN model. It requires less parameters than other deep learning models. Different features of input information can be extracted by convolution. The convolution layer is composed of several convolution kernel modules. The fully connected layer is shown in Fig. 13.4.

In the traditional neural network, each neuron is connected to each neuron in the next layer, which is called full connection. In CNN, the input layer is convoluted to get the output, which is not all connected but becomes local connection, that is, the local area of the input is connected to a neuron, and each layer uses different convolution kernels, and then combines them. The pooling layer is an important structure in convolutional neural networks. It is applied after the convolution layer. The pooling layer downsamples its input. The most common method is to retain the maximum information, which is generally the maximum pooling through windowing.

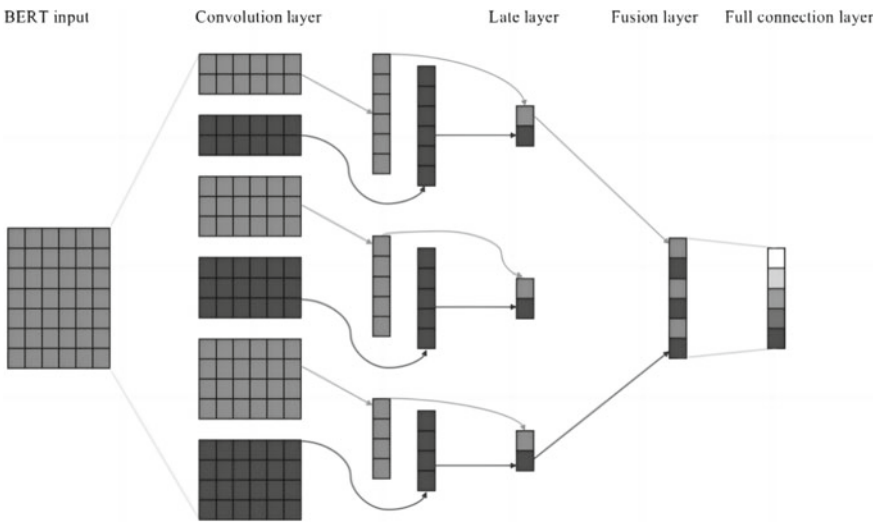


Fig. 13.4 TextCNN model

13.4 Experiment

13.4.1 Experimental Environment and Data Set

The experiment in this paper is implemented under the deep learning TensorFlow. The Python version is 3.7. The operating system is Windows 10 (64 bit). As for experimental hardware, the CPU is i3-9100f.

In supervised learning, the performance of the model is largely dependent on the dataset. The learning of neural networks also depends on datasets, and if the number of datasets is small, the learning will be insufficient. In order to provide suitable datasets for model training and model result evaluation, this paper selects real data from tax offices. Among them, 200,000 invoices are selected, and there are ten categories: tea, pet supplies, textile supplies, clothing, handicrafts, goods, furniture, wine, toys, and jewelry. Each category has 20,000 items with an average text character length of 15–30. 180,000 of them are used as the training set, 10,000 are used as the validation set, and the remaining 10,000 are used as the test set.

13.4.2 Experimental Setup

In training TextCNN, BERT, and BERT-TextCNN, we use cross-entropy as the loss function. TextCNN uses ADAM as the optimizer with a learning rate of 0.001. Meanwhile, in the model, BERT acts as the encoder of comment text and uses the embedding function of BERT language model to encode each comment into a sentence formed by stacking word vectors. As a new feature, it is used as the input of the CNN layer. In order to prevent overfitting, a dropout layer with a discarding rate of 0.5 is added in front of the full connection layer. The hyperparameter settings in this paper are given in Table 13.1.

Table 13.1 Hyperparameter

Parameter	Value
Embedding	64
Learning rate	1e−3
Train epoch	100
Dropout	0.3
Batch size	128
Epoch	20

13.4.3 Evaluation Indicators

The commonly used evaluation indicators for classification tasks include precision, recall, and F1 score. The calculation formula is as follows.

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{TF}} \tag{13.5}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{13.6}$$

$$\mathrm{F1} = \frac{2 \cdot P \cdot R}{P + R} \tag{13.7}$$

13.4.4 Analysis of Experimental Results

In the experiment, we compared the effects of different convolution kernel sizes on the model.

As given in Table 13.2, the best effect is obtained when using convolution kernels of (3, 4, 5) sizes. Therefore, we chose convolution kernel sizes of 3, 4, and 5. Model comparison was performed with the same dataset.

In this paper, we perform comparison experiments on invoice text dataset classification using different models of BERT-TextCNN, BERT, TextCNN, and CNN + Attention. The experiments measure the average accuracy (P), average recall (R), and average F1 value for ten labels. BERT model [14]: word vectors are trained by BERT model, and CLS flag bit feature vectors are used directly for downstream classification task. TextCNN [12] is implemented with Word2vec. CNN + Attention [13] obtains important local information from CNN and then calculates the score through attention.

As given in Table 13.3, the accuracy of BERT-TextCNN is 3.16% and 6.15% higher than that of BERT and TextCNN, respectively. It shows that this model is better than other models in invoice classification. BERT-TextCNN model has a high accuracy rate in invoice classification, which shows that the model has a good effect in invoice text classification. Its fine-tuning based on pre-training can effectively

Table 13.2 Comparison of convolutional kernel size

Size	Acc	Pre	Rec	F1
(2, 3, 4)	93.28	93.36	93.28	93.25
(3, 4, 5)	93.97	94.04	93.97	93.93
(4, 5, 6)	93.32	93.46	93.32	93.29

Table 13.3 Model performance comparison

Model	Acc	Pre	Recall	F1
TextCNN	87.57	85.52	83.02	82.45
CNN + Attention	89.96	90.73	90.17	90.21
BERT	90.56	92.89	92.50	90.78
BERT + TextCNN	93.72	94.83	92.96	92.73

Table 13.4 Comparison between test set and validation set

Data	Acc	Pre	Rcc	F1
Validation	95.96	96.03	95.96	95.92
Test set	96.34	96.40	96.44	96.40

solve the problem of polysemy of traditional word vectors, which is the key to obtain high accuracy of the model (Table 13.4).

There is hardly any difference between the test set and the verification set, so the model has good generalization.

13.5 Conclusion

In this paper, an improved BERT-TextCNN classification model based on deep learning algorithm is proposed for invoice short text data. The model uses BERT pre-training to generate word vectors and embeds words into convolutional neural networks. The test results show that the model performs well in all aspects, with high efficiency and accuracy. However, the data used in this paper is not enough, and a large number of data are needed to better train, so it may perform better with the increase of samples.

Although BERT-TextCNN has a significant improvement over TextCNN and BERT in classification, there are still some problems that need to be improved. The number of model parameters is large, and it takes a lot of time for training and loading, so it is an important research work to study the compression of BERT model and reduce the complexity of the model without suffering a large loss of model accuracy.

References

1. Zaremba, W., Sutskever, I., Vinyals, O.: Recurrent Neural Network Regularization. arXiv preprint arXiv:1409.2329 (2014)

2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 (2018)
3. Chen, Q., Zhuo, Z., Wang, W.: Bert for Joint Intent Classification and Slot Filling. arXiv preprint arXiv:1902.10909 (2019)
4. Sun, C., Qiu, X., Xu, Y., Huang, X.: How to fine-tune bert for text classification? In: Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20 (2019)
5. Liu, J., Xia, C., Yan, H., Xie, Z., Sun, J.: Hierarchical comprehensive context modeling for Chinese text classification. IEEE Access **7**, 154546–154559 (2019)
6. Chawla, S., Kaur, R., Aggarwal, P.: Text classification framework for short text based on TFIDF-FastText. In: Multimedia Tools and Applications, pp. 1–14 (2023)
7. Wang, J., Wang, Z., Zhang, D., Yan, J.: Combining knowledge with deep convolutional neural networks for short text classification. IJCAI **350**, 3172077–3172295 (2017)
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
9. Goldberg, Y., Levy, O.: word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. arXiv preprint arXiv:1402.3722 (2014)
10. Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., Okruszek, L.: Detecting formal thought disorder by deep contextualized word representations. Psychiatry Res. **304**, 114135 (2021)
11. Kim, Y.: Convolutional Neural Networks for Sentence Classification. arXiv preprint arXiv (2014)
12. Song, P., Geng, C., Li, Z.: Research on text classification based on convolutional neural network. In: 2019 International Conference on Computer Network, Electronic and Automation (ICCNEA), pp. 229–232. IEEE (2019)
13. Chen, Z., Tang, Y., Zhang, Z., Zhang, C., Wang, L.: Sentiment-aware short text classification based on convolutional neural network and attention. In: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence, pp. 1172–1179 (2019)
14. Jing, W., Bailong, Y.: News text classification and recommendation technology based on wide & deep-bert model. In: 2021 IEEE International Conference on Information Communication and Software Engineering, pp.209–216 (2021)

Chapter 14
Design of Fully Automatic Calibration Scheme for Load Box Based on Visual Recognition

Jinjin Li, Fanghua Mo, Keying Huang, and Junli Huang

Abstract At present, the calibration of the transformer load box adopts a "semi-automatic" device, which automatically measures the load value and manually switches gears. This article designs an automatic shift device for load box gear based on visual recognition. The device can recognize the load box gear information through the visual subsystem, upload it to the upper computer to form a calibration plan, and then automatically switch the mechanical gear of the load box to the corresponding load point on the plan through a motion control adapter for calibration, achieving full automation of load box calibration.

14.1 Introduction

The transformer load box is an auxiliary equipment in transformer calibration and requires regular calibration. At present, the automatic calibration device of the load box is usually used to calibrate the load box. The process is that the automatic calibration device outputs the working voltage U_0/working current I_0, and calculates the load value of the load box by sampling the current I/voltage U at both ends of the load box. The entire measurement process can be switched to different gears based on the different values of the tested object, and the current voltage ampere number, power factor, and error of the load box can be automatically calculated based on the measurement results. Due to the fact that a load box has multiple load value ranges, switching between different load values requires manual operation by humans, with each measurement being adjusted one by one, resulting in a cumbersome and redundant process and low efficiency [1].

J. Li · K. Huang · J. Huang
Guangxi Power Grid Corporation Metering Center, Nanning 530023, Guangxi, China

F. Mo (✉)
Guangxi Power Grid Corporation, Nanning Power Supply Bureau, Nanning 530023, Guangxi, China
e-mail: 357030634@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_14

At present, there is relatively little research on fully automatic measurement of load boxes in China. Reference [2] introduces an intelligent calibration system for load boxes, which can achieve automation of load value measurement. Manual operation is required for gear switching of load boxes; Reference [3] introduces a simple mechanical rotation scheme for switching the gear knob of the load box. This article designs a visual recognition based automatic adjustment output system for the load value of the load box, which can recognize the load gear information of the load box through the visual subsystem, upload it to the upper computer to form a calibration plan, and then automatically switch the mechanical gear of the load box to the corresponding load point on the plan through the motion control adapter for calibration, achieving full automation of the load box calibration process without manual intervention throughout the entire process.

14.2 Technical Proposal

14.2.1 Overall Design Plan

The automatic detection system device for the payment process of fee controlled electricity meters is a highly representative detection device formed by combining three major systems: mechanical and electrical automation, software-integrated control, and calibration system. This device completes the transmission, transportation, positioning, monitoring, etc., of various links through electrical control. Realize data exchange between the energy meter data collection and calibration device through software, as well as overall equipment operation scheduling and screen monitoring, as shown in Fig. 14.1.

Manually place the meter into the standard meter fixture, which provides a power supply slot for the meter and can automatically control the 220 V power supply of

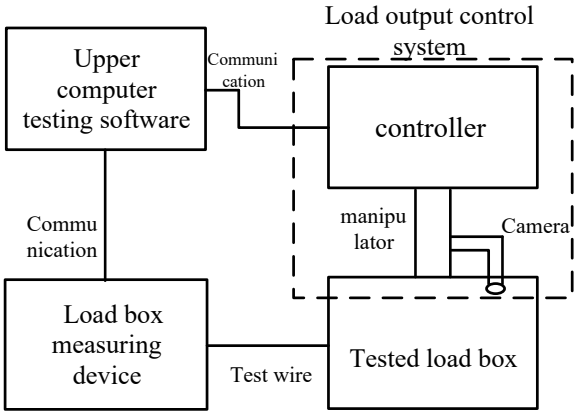


Fig. 14.1 Overall frame diagram

the meter. The robotic arm can automatically insert and remove the payment card of the meter and press the display button on the meter. Two sets of visual inspection systems, with cameras placed directly above the electricity meter, take photographs to determine the presence or absence of two models of electricity meters (one photograph position), precise positioning of the sockets of two electricity meters (two photograph positions), and display detection of two electricity meters (two photograph positions). There are a total of five photograph positions, and the mechanical arm presses the electricity meter. The designated 10 page display content of the camera's photograph display is identified [4].

14.2.2 Load Output Control System

The load output control system consists of a combination of visual subsystem, motion control subsystem, and motion mechanical subsystem to achieve the above functions. The detailed system diagram is shown in Fig. 14.2.

The vision subsystem is used for camera recognition and feedback control of the gear output of the load box. It is composed of an image acquisition card, an industrial camera, and a light source generator. The industrial camera captures the gear information and pointer position of the tested load box and transmits them to the image acquisition card. The image acquisition card is preprocessed and uploaded to

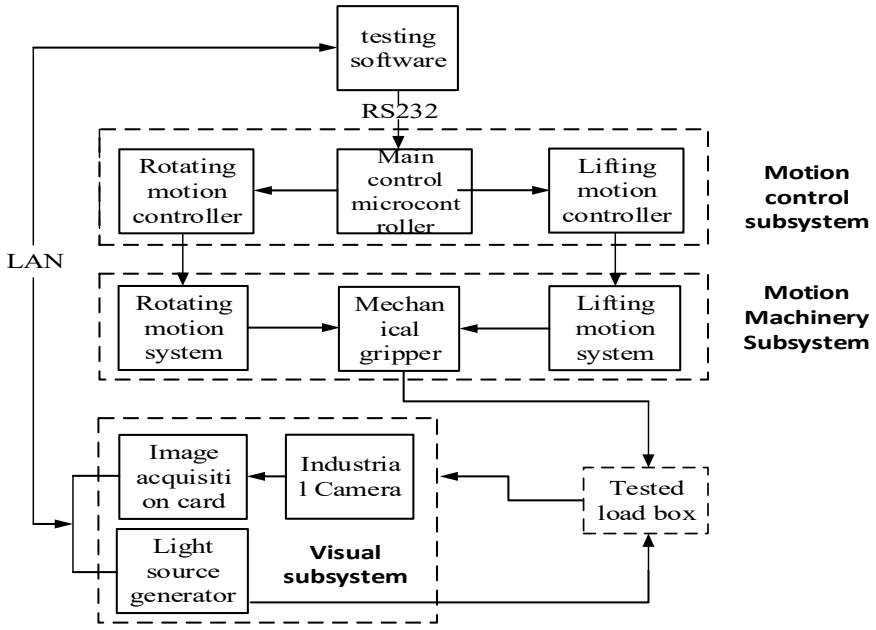


Fig. 14.2 Load box output regulation control system diagram

the image processing module of the upper computer testing software, which recognizes the image, Match the current pointer position with the required adjusted gear to provide control information for the motion control subsystem [5].

The motion control subsystem consists of a main microcontroller, a knob sleeve lifting controller, and a knob sleeve rotation controller. The master MCU communicates with the upper computer test software, and the test software sends the knob sleeve control information to the master MCU according to the image information uploaded by the vision subsystem, and the master MCU distributes the information to the lifting motion controller and the rotating motion controller [6].

The operating machinery subsystem is composed of a rotating motion system, a lifting motion system, and a knob clamp. The rotating motion system is composed of a rotating motor and a rotating screw [7]. The rotating motor receives instructions from the rotating motion control system for movement, while the lifting operation system is composed of a lifting motor and a lifting screw. The lifting motor receives instructions from the lifting motion control system for lifting movement. The lifting and rotating motion system controls the lifting and rotation of the sleeve clamp, and positions and rotates the load box gear knob [8].

14.3 Hardware Composition

14.3.1 Pin Automatic Insertion Power Supply Device

The device adopts a cylinder to push and crimp the pins, and the device has one set of single-phase and three-phase electricity meters each. This plug-in device is a professional plug-in device for the power grid, as shown in Fig. 14.3.

The upper computer uses LAN and RS232 ports to communicate with the vision subsystem and operation control subsystem, responsible for controlling the entire system work, reading industrial camera images, and analyzing the content contained

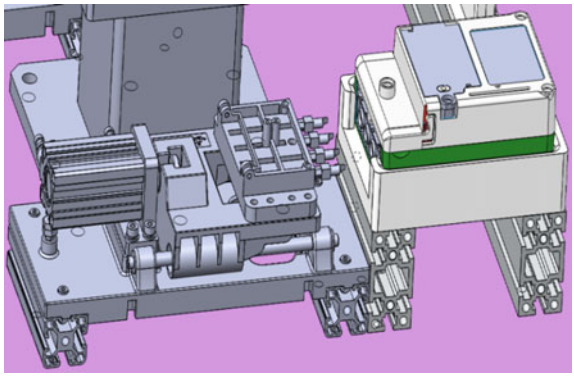


Fig. 14.3 Schematic diagram of single-phase plug-in device

in the images; send instructions to the microcontroller controller to control the rotating motor and lifting motor for related actions.

The microcontroller chip (SMT32F4 series) of the mainstream Italian semiconductor used for industrial control is used as the main control chip of the lower computer [9]. The chip adopts an ARM-32 bit core, including 32KRAM, RS232, ADC, and 32 IO control ports.

14.3.2 Visual Inspection System

There are two sets of visual inspection, one set for each camera and one set for each light source. The visual camera adopts an industrial grade 10 million pixel camera, which meets the equipment usage requirements. The light source adopts a strip visual light source. The form is shown in the following figure, as shown in Fig. 14.4.

The industrial camera captures the text, color, area, and other features corresponding to the knob of the tested object (load box), and uploads the captured content to the upper computer software through the LAN port. Industrial cameras need to be equipped with a gigabit LAN network interface. Select a 25 mm lens and set the working distance from the lens to the knob to approximately 170 mm. Design field of view is 100 mm × 80 mm with a resolution of 1292 × 964, with a design accuracy of 0.0772 mm/pix. The lens adopts a parallel light path design, placing the aperture diaphragm of the visual system on the focal plane of the imaging side to eliminate perspective distortion and adapt to the visual image acquisition of the system.

The light source adopts a circular diffuser design, with a distance of about 100 mm from the measured object knob. It is equipped with an NG-1000 light source controller to control the brightness and lighting status (on/off).

A PCI image acquisition card with a 32bit/66 MHz LAN interface is used for the collection, recording, and processing of characteristic image data such as text, color, and area of the load box, completing functions such as testing, measurement, and signal processing analysis.

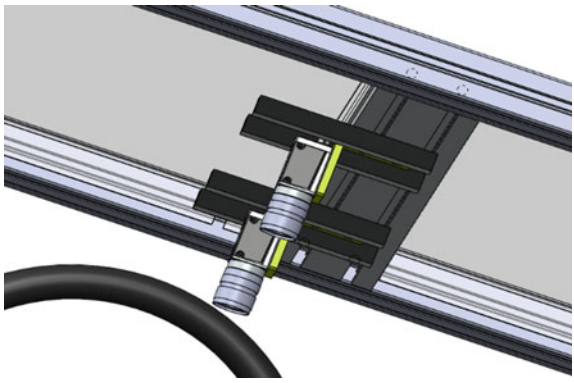


Fig. 14.4 Visual inspection system

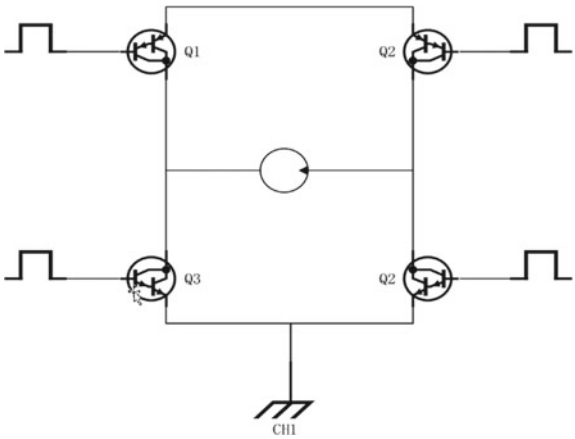


Fig. 14.5 Schematic diagram of motor controller

14.3.3 Motor Control System

The motor drive circuit consists of a bridge circuit composed of four high-power transistors, which can control the motor to perform forward, reverse, and brake actions, and read the corresponding protection signals to provide power outage protection for the motor [10]. The schematic diagram is shown in Fig. 14.5.

The rotating motor adopts a combination motor with a built-in reduction gear, which has simple control, high torque, and relatively low speed, making it easy to control.

The power supply adopts a 100 W AC–DC power conversion module, which can convert 220 V AC power into 24 V, 12 V, and 5 V DC.

14.4 Module Functions

14.4.1 CCD Detection

Liquid crystal image recognition is based on the image processing software of Convision in the United States, which is developed and deepened by our company. It uses industrial cameras to automatically recognize image acquisition, calibration devices load voltage, automatically analyze manufacturer information, cooperate with (image template), query and identify the image template of the manufacturer in the comparison system database, and automatically store the recognition results in the system database. The collected images are automatically stored in the corresponding barcode database, and process and upload.

Camera: The advantage of a smart camera is that it separates from the capture card and directly uses a network interface to capture images. The smart camera has

a DSP image processing system inside, which can integrate some image algorithms and more efficiently complete the tasks specified by the system. This system adopts imported high-end intelligent cameras, which have high stability and have been widely used in many industrial sites, with good practicality.

14.4.2 Identification Scheme

LCD image recognition is divided into two parts

Step 1: Template production, select a standard image as the template image, and create an image comparison template. The operation process is as follows:

(1) Template production

Click on the green small box on the interface, adjust its size and position to cover the desired positioning area. Click on the Add Locator button on the interface to confirm the positioning template.

Adjust the green small box, select the character you want to create, enter the name of the character in the character name field of the interface, set the score for the character template, and then click the Add Character button. The system will automatically create a character template for the selected area and save it to the database with the set name. Create all the characters to be recognized in sequence and complete the template production.

(2) Image recognition

After creating the template, exit the template creation interface and return to the main interface. Select the type of recognition you want to perform on the interface, and then load the image for the recognition operation into the interface by capturing or importing images with a camera.

Click the start recognition button on the interface, and the system will automatically load the corresponding template and start the template recognition operation. During the recognition process, there will be prompts on the interface for qualified, unqualified, and recognition progress.

By comparing with LCD, the unqualified images will be saved in the database for easy viewing by operators; at the same time, in order to compensate for misjudgment in image recognition (such as differences in the LCD screen itself and misjudgment in image recognition itself), the system provides a function of manually reviewing unqualified images.

The system provides an operation interface to view unqualified images based on information such as image recognition time. At the same time, meters that have not uploaded packing conclusions can be manually reviewed and modified with image conclusions. The system will update the final verification conclusion according to the situation to avoid meter verification failure due to misjudgment.

14.4.3 Real-Time Alarm

Real-time collection of PLC alarms, control system abnormal alarms, and calibration system abnormal alarms display the alarm information in the alarm information column. Once an alarm occurs, the operator can immediately detect it and take corresponding measures based on the alarm information. At the same time, the system will save the alarm information in the database for future reference. The alarm is divided into different levels, and the system will pop up an alarm information box for alarms that require human intervention based on the situation, and provide suggestions for handling and solving them. The operator can handle it according to the suggested solutions.

Output important links in the system operation process to the information list, such as starting the robot to grasp the meter, and starting the calibration line to enter the board. Record the event and its occurrence time. This way, operators can better grasp the verification process and understand the operation of the system.

This will display the current calibration process of each device in real time. If the calibration line is currently in the loading state, unloading state, or calibration state; if it is in the verification state, the specific verification phase and the estimated time required for the current phase will be displayed. The processing batch represents the total number of batches verified by the device. The last column of the table displays the connection status between the control system and the calibration system. If the calibration system program connection is closed or the connection port is caused by network reasons, this column will be displayed.

14.5 Algorithm Design

14.5.1 Image Preprocessing

Before image recognition or analysis, it is generally necessary to preprocess the image first, eliminate irrelevant information in the image, recover useful real information, and enhance the detection of information, so as to improve the reliability of feature extraction, image segmentation, matching and recognition, and make the recognition effect of image recognition algorithm more accurate. The main steps of preprocessing include grayscale, geometric transformation, and image enhancement.

(1) Graying

In this paper, the component method is used to grayscale the image. Component method refers to the brightness of the three components in the color image as the gray value of the three gray images, that is

$$f_1(i, j) = R(i, j)\ f_2(i, j) = G(i, j)\ f_3(i, j) = B(i, j)$$

where f_k (i, j) $(k = 1, 2, 3)$ is the gray value of the converted gray image at (i, j).

(2) Geometric Transformation

In order to correct the systematic error of the image acquisition system and the random error of the instrument position, the geometric transformation of the image acquisition is carried out by the translation method. In addition, a grayscale interpolation algorithm is required, as the pixels of the output image may be mapped to non-integer coordinates of the input image according to the calculation of this transformation relationship.

(3) Image Intensification

In order to enhance the useful information in the image, the image can be distorted to improve its visual effect. For a given image application, it purposefully emphasizes the overall or local features of the image, makes an otherwise unclear image clearer or emphasizes some interesting features, enlarges the difference between the features of different objects in the image, and suppresses the features of interest. It can improve image quality, enrich information, enhance image interpretation and recognition ability, and meet some special analysis needs.

14.5.2 Image Preprocessing

(1) Feature extraction

Image feature extraction refers to the processing and analysis of the information in the image that is not easily disturbed by random factors, and the extraction of iconic information as image feature information. Feature extraction algorithms include: SIFT, SURF, BRISK, FREAK, and MSER.

SURF algorithm is adopted in this paper. When generating feature vectors, integral graphs and fast Hessian detectors are used to determine whether the key points extracted in scale space are extremal points. The main direction of each extreme point is determined, the window area is constructed along the main direction, and the feature vector is extracted in the window to describe the key points. Compared with SIFT algorithm, SURF algorithm maintains scale-invariant and rotation-invariant characteristics, with high speed and good robustness.

(2) Image segmentation

Image segmentation is the process of dividing an image into regions with similar features. The most basic feature of image segmentation is the brightness amplitude of monochrome image and the color component of color photograph.

This paper adopts the method of region merging. The first step is the initial region segmentation of the image. In extreme cases, each pixel can be seen as a small region. The second step is to determine the similarity standard, according to the adjacent areas of gray, color, texture, and other parameters to compare. If the gray levels of

adjacent regions are evenly distributed, the average gray levels between regions can be compared. If the average gray difference is less than a certain threshold, it is considered that the similarity of the two regions can be combined. Step 3: Determine whether the adjacent areas in the image meet the similarity criteria, and merge if they are similar. Repeat this step until there are no more regions to merge.

14.6 Software Design

The intelligent measurement system consists of four parts: automatic detection, data recording, system management, and assistance. The software composition is shown in Fig. 14.6.

Automatic detection: including visual detection and motion control, where visual detection achieves functions such as control, acquisition, and analysis of industrial cameras, while outputting control information (including control signals and position information) to the motion control subsystem; Motion control achieves motor control and mechanical action control [11].

System management: including visual calibration and system parameter setting, where visual calibration realizes the calibration functions of the visual system, initial position, and accuracy; the system parameter settings include system settings such as industrial camera IP address and local IP [12].

Data management: achieved the function of recording process data for visual image detection (DATALOG).

The detection process is shown in Fig. 14.7.

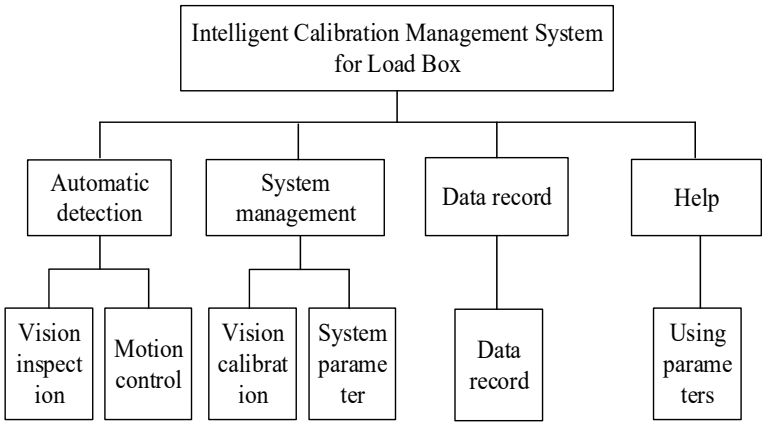


Fig. 14.6 Software composition

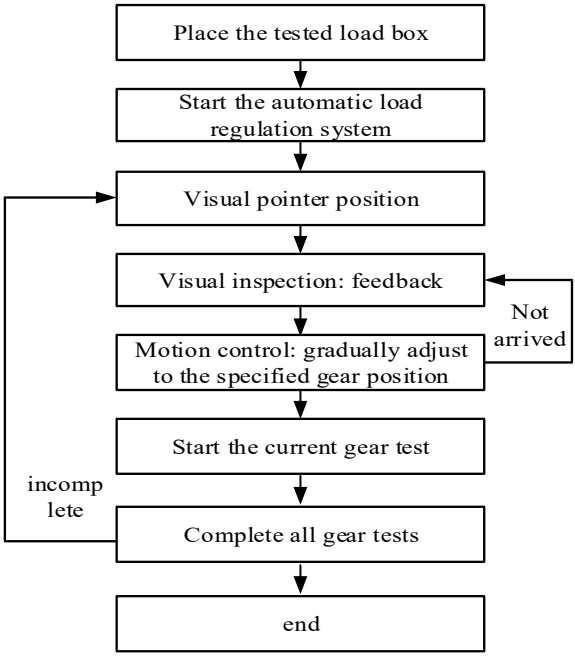


Fig. 14.7 Calibration software flowchart

14.7 Conclusion

The traditional load box calibration method of manually switching gears is cumbersome and inefficient. The difficulty of fully automatic calibration of load boxes lies in the accurate adaptive adjustment output of non-programmable load box values. This article designs a circular mechanical snap ring that can grasp knobs of different sizes. Visual recognition is achieved through industrial cameras and graphics acquisition cards, tracking and feedback the pointer position of the knobs, ensuring that the load value of the load box is consistent with the current value of the test plan, and achieving the entire process automation of load box calibration.

References

1. Verification regulation of JJF1264-2010 “Measurement Transformer Load Box”
2. Lu, J., Wang, H., Li, W., Liu, Y., Wang, T., Liu, Y., Jia, J.: Transformer load box intelligent verification system design. Electrical Meas. Instrumentation **51**(11), 6–8 (2014)
3. Fu, L., Liao, Y.: A design scheme of full automatic measuring device for load boxes. Electrical Meas. Instrumentation **47**(539), 55–57 (2014)
4. Wu, Y.: Image target recognition. Nanjing University of Science and Technology School of Electronic Engineering and Optoelectronics Technology (2016)
5. Chen, H.: A computer vision recognition and matching method and its application. Comput. Telecommun. **08**, 86–88 (2007)

6. Chen, W., Wang, J., Li, X., Liang, Y., Li, L.: Discussion on calibration method of transformer load box tester. Autom. Instrumentation **02**, 137–138 (2013)
7. Chen, G.: Transformer load box method of calibration errors. Metrol. Meas. Technique 6 (2009)
8. Zhang, J., Huang, H., Fang, J.: Current transformer load box detection and error analysis. Metering Technol. 12 (2010)
9. Ren, M.: Overview of image visual recognition algorithm. Intell. Comput. Appl. **9**(3), 294–297 (2019)
10. Ren, Y.: Research on visual recognition technology of power equipment with improved maximum category variance method. China New Technol. New Products **4**, 20–22 (2021)
11. Xu, Z., Wu, W., Luo, X.-H., Zhu, D., Chen, X.-W.: Research on power equipment recognition technology based on machine vision application. Experimental Res. 23–25 (2022)
12. Wu, X., Huang, G., Sun, L.: Fast visual identification and location algorithm for industrial sorting robots based on deep learning. ROBOT **38**(6), 713–719 (2016)

Chapter 15
DeepScan: Revolutionizing Garbage Detection and Classification with Deep Learning

Guohao Li, Anas Bilal, Muhammad Ibrar, Shahid Karim, and Haixia Long

Abstract With the rapid increase in human household waste due to urbanization, garbage classification has become a critical issue in environmental protection. The traditional manual classification method is no longer sufficient to meet the growing demand, highlighting the need for an efficient and accurate garbage classification and identification system. This paper proposes a garbage classification and recognition system that uses a convolution neural network for feature extraction and classification. The system was trained and tested using datasets containing five types of garbage. The experimental results show that the proposed system achieves an accuracy of 90.3%, outperforming traditional machine learning methods. Furthermore, confusion matrix analysis and feature visualization provide additional evidence of the system's effectiveness and interpretability. This deep learning-based garbage classification and recognition system has significant potential for practical use in urban environmental management and garbage classification and treatment enterprises.

G. Li · A. Bilal (✉) · H. Long
College of Information Science Technology, Hainan Normal University, Haikou City 571158, Hainan Province, China
e-mail: a.bilal19@yaoo.com

M. Ibrar
Software College, Shenyang Normal University, Shenyang, China
e-mail: ibrar@synu.edu.cn

S. Karim
Research and Development Institute of Northwestern, Polytechnical University in Shenzhen, Shenzhen 518057, China
e-mail: shahid@nwpu.edu.cn

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_15

15.1 Introduction

In terms of environmental issues, the increase in global population and the acceleration of urbanization, as well as the living environment in rural areas under the dual carbon strategy, have led to a sharp increase in the amount of human household waste [1]. Garbage disposal has become a global environmental problem that should be faced together. In terms of resource utilization, garbage classification can recycle and reuse potentially valuable and reusable resources such as paper, plastic, and metal, thereby improving resource utilization efficiency. In terms of laws and regulations, garbage classification regulations are important measures that governments, international organizations, and even global citizens should respond to [2].

In terms of social ideology, garbage classification is a way of life based on the increasing environmental awareness of residents. Establishing a correct concept and awareness of garbage classification is the responsibility and obligation of all citizens. The traditional garbage classification method is limited to human and material resources, resulting in low classification efficiency and accuracy influenced by subjective human factors. In addition, traditional garbage classification methods can only handle a small amount of garbage, and the processing cost is relatively high. Therefore, a garbage classification and recognition system can be developed based on the strong adaptability of deep learning algorithms and the ability to extract a large number of learning features and adapt to different data and scenarios.

The research and practice of garbage classification can promote the process of environmental protection and sustainable development, and contribute to the sustainable development of human society; it can effectively separate recyclable and harmful substances, reduce the amount of landfill and incineration, reduce environmental and social costs, and promote the development of the circular economy; the research on garbage classification involves computer science, chemistry, environmental science, sensor technology, and even sociology and psychology. The cross-integration of these disciplines can not only promote innovation in garbage classification-related technologies, but also promote the application, promotion, and industrialization of technologies [3, 4].

Current research status of garbage classification:

1. Classification standards and methods for garbage classification: Domestic research mainly focuses on this aspect, studying how to establish standards and methods suitable for each city based on the characteristics and attributes of different garbage.
2. The influencing factors and effectiveness evaluation of garbage classification: Evaluate the causes of garbage classification and its effectiveness in environmental protection, resource recovery, and other aspects.
3. Technical means and facilities for garbage classification: Research on technical means and facilities for garbage classification, including classification and recycling equipment, treatment equipment, etc.

4. Social behavior and management of garbage classification: Research how to promote and improve the implementation effect of garbage classification through social behavior and management methods.

However, there are also shortcomings in the research:

1. Lack of long-term tracking research: Currently, most of the research on garbage classification is horizontal and short-term, lacking long-term tracking research, making it difficult to comprehensively evaluate the effectiveness and impact of garbage classification.
2. Lack of international research: Garbage classification is a global issue, but currently research is mainly focused on certain countries and regions, lacking international research and exchange. Different countries and regions have vastly different garbage classification standards, methods, and technical means [5].
3. Lack of technological innovation research: The research on waste classification technology and facilities mainly focuses on traditional classification and recycling equipment and treatment equipment, lacking technological innovation research, making it difficult to meet the actual needs of waste classification.
4. Lack of social psychology and behavioral research: Garbage sorting behavior is affected by cognitive attitude, social norms, and self-efficacy, and social psychology and behavioral factors need to be considered, but there is a lack of relevant research at present.

15.2 Theory of Convolutional Neural Networks

15.2.1 Traditional Neural Networks

The traditional neural network refers to a kind of feedforward feedback neural network based on the artificial neuron model, which is composed of multiple layers of neurons. Each layer of neurons receives the output of the upper layer of neurons as the input, and through the role of linear transformation and nonlinear activation function, produces the output of the next layer of neurons, and finally obtains the output of the network [6]. Traditional neural network training uses the hair propagation algorithm to update the parameters of the network by minimizing the loss function. However, in training, traditional neural networks are prone to problems such as vanishing and exploding gradients, which affect network performance. A series of improvement methods have been proposed to solve this problem, such as adding regularization terms, using different activation functions and optimization algorithms, and improving the network structure [7].

15.2.2 Convolutional Neural Network

In terms of environmental issues, the increase in global population, convolutional neural network is a forward feedback neural network that has been widely used in the fields of image and speech processing. The main components of convolutional neural networks are convolutional layers, pooling layers, and fully connected layers. The convolutional layer uses convolutional verification to perform convolutional operations on input data to extract the features of the input data [8]. The pooling layer is usually followed by the convolution layer to reduce the space size and quantity of input data to reduce computational complexity and control overfitting. The fully connected layer connects the outputs of the convolutional and pooling layers and passes them on to the final output layer.

In terms of environmental issues, the increase in global population and the acceleration of urbanization, as well as the living environment in rural areas under the dual carbon strategy, have led to a sharp increase in the amount of human household waste. Garbage disposal has become a global environmental problem that should be faced together. In terms of resource utilization, garbage classification can recycle and reuse potentially valuable and reusable resources such as paper, plastic, and metal, thereby improving resource utilization efficiency. In terms of laws and regulations, garbage classification regulations are important measures that governments, international organizations, and even global citizens should respond to.

1. Convolutional layers use convolutional kernels to extract features from input data. In the convolutional layer, the convolutional kernel performs convolution operations on the input data through a sliding window, multiplies its corresponding weights with the corresponding parts of the input data, and sums all elements to obtain the output of the convolutional layer [9]. The convolution operation is shown in Fig. 15.1.

Another important feature of convolutional layers is parameter sharing, which explains that the weights of convolutional kernels are the same when applied in different positions. The parameter-sharing mechanism can reduce the number of unnecessary parameters, reduce the complexity and computational cost of the model, and also make the model more robust and generalization capable.

In addition to standard convolutional layers, scholars have also proposed various improved convolutional layer structures, such as deep separable convolutional layers, hollow convolutional layers, etc., to further enhance the performance of convolutional neural networks. These improved convolutional layer structures can better adapt to different application scenarios, such as the need for faster running speed, larger Receptive field, etc.

2. Common pooling methods include maximum pooling and average pooling, and the pooling layer operations are shown in Fig. 15.2. Through the down-sampling operation of the pooling layer, the size of input data can be reduced, and the robustness of the model can be enhanced, the sensitivity of the model to small

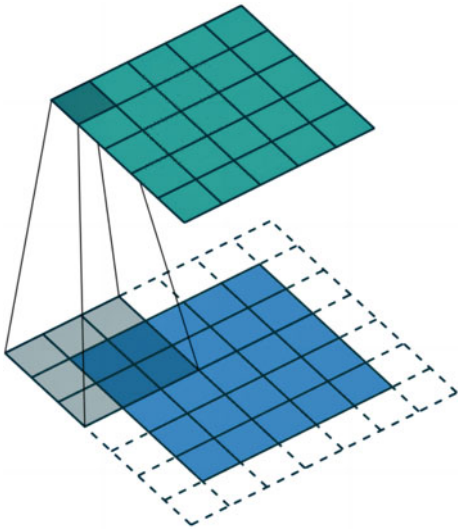


Fig. 15.1 Convolutional operation

changes in the input data can be reduced, at the same time, the number of parameters to be trained in the model can be reduced, and the risk of overfitting can be reduced [10].

3. The main function of the fully connected layer is to map the output features of the previous network layer into classification or regression results and connect them with neurons. It is usually used for tasks such as classification, recognition, and prediction of features extracted from convolutional and pooling layers [11, 12].

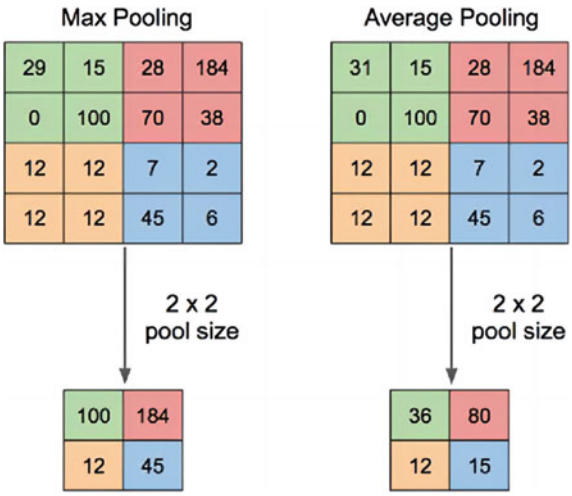


Fig. 15.2 Pooling layer operations

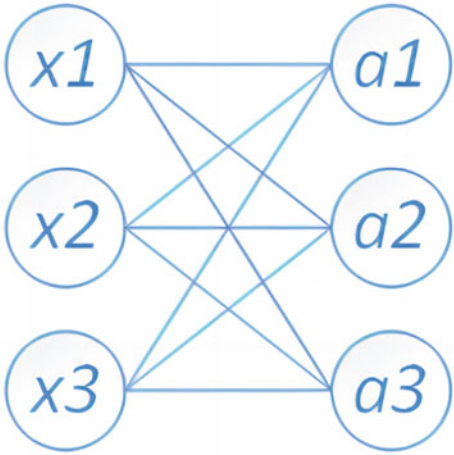


Fig. 15.3 Fully connected layer operation

A fully connected layer typically consists of several neurons, each with a set of weights and bias parameters. The fully connected layer performs a linear transformation on the input data through the weight matrix and offsets vector, then inputs the results of the linear transformation into the activation function, and finally outputs the results of this layer. The fully connected operation is shown in Fig. 15.3.

The fully connected layer operation matrix is used to represent its functional relationship, where x is the input vector and y is the output vector. In Eq. (15.1), W represents the weight matrix and b represents the bias vector. The weight matrix contains the connection weights between neurons, while the bias vector provides additional parameters for shifting the activation function. These parameters are adjusted during training to learn patterns and make predictions in the neural network Eq. (15.1) is:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{15.1}$$

In deep learning research, the full connection layer is often used in image recognition, natural language processing, recommendation systems, and other fields. However, when the full connection layer processes large-scale high-dimensional data, the number of parameters is very large, which is prone to overfitting problems, and the training time will be very long. In order to solve these problems, many optimization methods have been proposed, such as introducing regularization methods and reducing the number of model parameters [13].

4. The activation function is a basic component in the neural network, which is used for nonlinear conversion of the input signal to enhance the expression ability of the network [14]. This paper selects two activation functions, ReLu activation function and softmax activation function

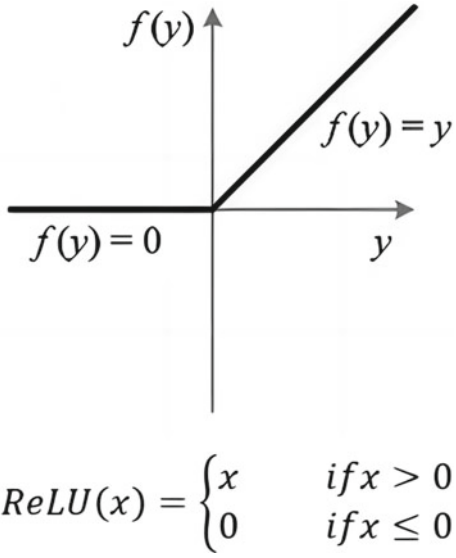


Fig. 15.4 ReLu activation function diagram

The characteristic of the ReLu activation function is that it is simple and easy to implement, does not cause gradient disappearance, and has good computational efficiency. In convolutional neural networks, convolutional layers are usually combined with the ReLu activation function to enhance the nonlinear fitting ability of the model. The function image is shown in Fig. 15.4, with the expression (15.2):

$$\mathrm{ReLu}(x) = \max(0, x). \tag{15.2}$$

The main function of the softmax activation function is to transform the complex neural network problem into a probability distribution problem. In probability distribution, it is necessary to map the output results as probabilities for each category and map them as a probability distribution, so that the total probability is 1. The output of the softmax activation function can be regarded as the prediction probability of each category, so the cross-entropy loss function can be used to train the neural network to achieve the purpose of classification. The function expression is Eq. (15.3):

$$S_i \frac{e^i}{\sum e^i}. \tag{15.3}$$

The backpropagation algorithm is an optimization algorithm in deep learning models, which can transfer errors layer by layer forward by calculating the gradient information of errors, thereby achieving adjustment of model parameters. The basic idea is to use the chain rule to calculate the error gradient of each neuron, and then update the weight and bias of the model through the gradient descent method.

Specifically, during the calculation of specific errors, the error gradient can be seen as the derivative of the current neuron error to the previous layer of neurons. After the calculation is completed, the gradient descent method updates parameters based on the direction and size of the gradient, gradually reducing the error and making the model's prediction results more accurate.

Although the backpropagation algorithm has been widely applied in the field of deep learning and has achieved many successful applications, it still has some shortcomings. For example, due to the computational complexity of the chain rule, backpropagation algorithms are prone to problems such as vanishing or exploding gradients in deep neural networks, resulting in models that cannot converge or have poor training results. In addition, backpropagation algorithms are also susceptible to the influence of local minima, resulting in the inability to obtain globally optimal solutions.

In recent years, some improved methods for backpropagation algorithms have been proposed to address the aforementioned issues, such as using random gradient descent, batch normalization, residual connection, and other techniques to optimize the model training process, thereby improving the performance and stability of the model [15].

15.2.3 Image Classification Networks

AlexNet is a deep convolutional neural network model proposed by Krizhevsky et al. in 2012, which won the ImageNet image recognition competition. AlexNet is mainly composed of five convolution layers, three full connection layers, and pooling layers, and has strong nonlinear mapping and expression capabilities. Krizhevsky et al. have greatly improved the classification accuracy by improving some key technologies in the deep convolution neural network, such as the use of the ReLU activation function, local response normalization technology, and dropout technology [16].

In AlexNet, the first layer is a convolutional layer used to extract low-level features of the image, and the second layer is a pooling layer used to reduce the dimensionality of the feature map and increase the robustness of the network. The next three convolutional layers and two pooling layers are used to further extract higher-level features. Finally, three fully connected layers and a softmax classifier are used to make the final classification decision.

In the training process of AlexNet, the backpropagation algorithm is used to optimize the network weight to minimize the loss function. In addition, Krizhevsky et al. also used data augmentation techniques to expand the training dataset and accelerated the training process using multi-GPU parallel computing technology.

Overall, the application and improvement of key technologies in AlexNet provide important ideas and methods for the development of deep learning in fields such as image classification.

VGG (Visual Geometry Group) is a deep learning architecture proposed by the Visual Geometry Group of Oxford University in 2014 [17]. It uses a very small convolutional kernel (3×3) and a deeper network structure, aiming to improve the accuracy of image classification. The main idea of VGG is to replace fewer large convolutional layers with multiple smaller ones to increase network depth and to achieve higher-level feature extraction by stacking multiple similar convolutional layers. In addition, VGG also introduces the idea of reusing the same type of layer, making the network structure clearer and easier to understand.

The innovation of VGG lies in the use of small convolutional kernels and deeper network architecture to improve accuracy, but there are also some drawbacks. VGG16 has a large number of parameters, resulting in a high computational cost and time budget for training and prediction. In addition, due to the relatively simple relationship between convolutional layers and fully connected layers, the VGG architecture may have limitations when dealing with more complex tasks.

GoogleNet (also known as Inception V1) is an image classification model based on deep convolutional neural networks, developed by researchers from Google and published at the CVPR (Computer Vision and Pattern Recognition) conference in 2014. The main contribution of GoogleNet is the introduction of the Inception module into its network structure, which enables the network to maintain fewer parameters while achieving higher recognition accuracy. In GoogleNet, an Inception module consists of multiple branches, each with different convolutional kernel sizes and stride sizes, capturing features at different scales. The outputs between these branches are concatenated in the channel direction to form a richer feature representation. By stacking multiple Inception modules, GoogleNet has better performance and accuracy in processing multi-scale feature information. At the same time, Google Net also uses multiple auxiliary classifiers to improve the training effectiveness of the network, which has become an important idea for subsequent network design [18].

ResNet was proposed by Kaiming He et al. in 2015. ResNet solves this problem by extracting information from multiple convolutional blocks, allowing the network to have deeper layers. The design of residual blocks is achieved by adding a cross-layer residual connection between two adjacent convolutional layers, allowing information to be skipped directly from one layer to another, thereby avoiding information loss.

15.3 Proposed Methodology

15.3.1 Image Processing Based on Keras

15.3.1.1 Digital Image

Digital images can be represented by matrices, and the positions of pixels can be represented by the row–column coordinates of the matrix. The values of each pixel represent the brightness or color information of the image at that position. For example, a digital image with M rows and N columns can be represented as a matrix of $M \times N$ [19], with the expression shown in 4. There are two common digital image representation methods: grayscale image and color image. A grayscale image refers to a pixel with only one value, representing the brightness information of that pixel; a color map, on the other hand, has three channels per pixel, representing the intensity values of red, green, and blue color information [20]. In a computer, a digital image can be represented as a three-dimensional array, where the first dimension represents the number of rows, the second dimension represents the number of columns, and the third dimension represents the number of channels. The commonly used image formats include JPEG, PNG, BMP, etc., which use different compression and encoding methods to reduce the size of image files. Algorithms and models in the fields of image processing and computer vision often require preprocessing and representation of digital images for subsequent analysis and processing.

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0, N-1) \\ f(1,0) & f(1,1) & \cdots & f(1, N-1) \\ \cdots & \cdots & \cdots & \cdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1, N-1) \end{bmatrix} \tag{15.4}$$

Digital image processing technology is a technology that studies how to obtain, process, and analyze digital images. Its purpose is to improve image quality, extract feature information, and achieve image analysis and recognition [21], which can be divided into the following aspects:

(1) Image enhancement: Histogram equalization, contrast stretching, adaptive filtering, image sharpening techniques are referring to the process of processing images to improve their quality or make them more suitable for specific applications.
(2) Image segmentation: K-means clustering is the process of dividing a digital image into multiple different regions, which typically have similar features such as color, texture, and brightness. Image segmentation can be used in fields such as image analysis and understanding, object recognition, and tracking.
(3) Feature extraction: CNN refers to the extraction of meaningful features from digital images for classification, recognition, or analysis. Common feature

extraction methods include edge detection, corner detection, texture analysis, and color histogram.

(4) Image recognition: AlexNet refers to the process of automatically classifying digital images. Common image recognition methods include feature-based methods and deep learning-based methods.

(5) Object detection: Viola–Jones algorithm (Haar cascades), R-CNN (Region-based Convolutional Neural Networks), Fast R-CNN, Faster R-CNN, YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector) refer to the detection of specific target positions and quantities in digital images. Common object detection methods include the sliding window method, region proposal method, and deep learning method.

(6) Image reconstruction: Compressed sensing, total variation (TV) regularization, sparse representation, iterative methods (e.g., iterative back-projection) refer to the use of limited data to reconstruct high-quality images.

15.3.1.2 Image Normalization

Normalize the pixel values of the image to a range of 0–1 [22]. Specifically, the minimum pixel value is mapped to 0, the maximum pixel value is mapped to 1, and the middle pixel values are mapped proportionally. Normalization can enable different images to have the same pixel value range, making it convenient for subsequent processing.

In image processing, grayscale value is usually a key parameter representing image brightness. Due to factors such as lighting and contrast, the grayscale range of images may vary, making it inconvenient to compare or process them. Therefore, it is necessary to unify their grayscale ranges [23].

(1) Read-in image: Read in the image to be processed from a file or other source.

(2) Histogram equalization: The effect of transforming the gray histogram of an image so that each image has the same grayscale.

(3) Statistical grayscale value range: Traverse the entire grayscale image, and calculate the maximum and minimum grayscale values of all pixels.

(4) Grayscale range mapping: Maps the grayscale values of each pixel proportionally to a unified grayscale range.

(5) Output Image: Output the processed image to a file or other destination.

15.3.1.3 Image Enhancement Processing

ImageDataGenerator is a tool for data augmentation in the Keras deep learning framework, mainly used to generate randomly transformed image data and generate more sample data [24].

15.3.2 Sequential Model Training and Optimization

15.3.2.1 Selection and Partitioning of Data

Choosing a high-quality dataset is crucial. The dataset constructed in this article is sourced from the open-source dataset, Kaggle garbage dataset obtained online. The bad images included were cleaned and included 2527 images, totaling six categories. The dataset was divided into two groups in an 8:2 ratio: the training set and the testing set.

15.3.2.2 TensorFlow Deep Learning Development Framework

TensorFlow, a deep learning development framework, is an open-source deep learning framework developed by the Google Brain team. It is widely used in deep learning research and development in industry and academia [25]. The advantages of using the TensorFlow framework in this article compared to other deep learning frameworks are:

(1) Easy to use: Provides a concise and easy-to-use advanced API, such as Keras, which can easily build various deep learning models and is easy to expand. In addition, TensorFlow also provides many pretrained models and tools, making it easier to browse the detailed processes and results during the visualization process.
(2) Easy to deploy: Supports a variety of different platforms and devices, including PCs, servers, mobile devices, and embedded devices. At the same time, TensorFlow also provides a set of tools and libraries that can help developers more easily deploy models to various devices and platforms.
(3) High performance: With multiple hardware platforms and operating systems, such as CPU, GPU, and TPU, it is convenient to train and infer deep learning models on different hardware platforms. At the same time, TensorFlow also supports multiple programming languages, making it easy to conduct deep learning and development in different programming environments.
(4) Community support: With strong community support and developers and users, rich development documents and tutorials make it easy for users to access help and learning materials [26].

15.3.2.3 Sequential Model Construction

The basic idea of the two-dimensional convolutional layer is to extract image features on the input two-dimensional matrix by sliding a small convolutional kernel (a set of weights). This process can be seen as the operation of feature extraction and dimensionality reduction on the input image or feature map. The Conv2D class is a two-dimensional convolutional layer that can perform convolution operations

on two-dimensional data such as images, that is, spatial convolution of images. Its advantage lies in its ability to automatically learn local features of images and extract higher-level features by stacking multiple convolutional layers.

The role of MaxPooling2D in convolutional neural networks is to preserve image features while reducing spatial resolution, while also improving the stability and generalization ability of the model. Specifically, the MaxPooling2D layer preserves the maximum value in each pooling window as the corresponding position value in the downsampled feature map.

In addition, there is a flatten layer (but not a pooling layer) between the convolutional layer and the fully connected layer, which is a data transformation layer used to flatten multi-dimensional tensors (such as the output of the convolutional layer) into one-dimensional tensors for easy processing by subsequent fully connected layers. The pooling layer is used to downsample the input feature map, reduce the size of the feature map, reduce the complexity of the model, and to some extent improve the robustness of the model.

This layer has strong expressive ability in feature extraction and classification tasks. Although its structure is simple, it can be connected to upper level neurons to express complex relationships of different input features, thereby improving the prediction accuracy of the model.

15.3.2.4 Backbone Network Selection

The backbone network selected in this article is the AlexNet classification network mentioned earlier. During the learning process, the weight is adjusted using optimization functions, which can effectively fit the output of the model and minimize losses. The most common optimizers are Adam and SGD. The main idea of the Adam optimizer is to combine the momentum gradient algorithm and the adaptive learning rate algorithm. It can consider both the average value of the historical gradient and the variance of the historical gradient in the gradient descent process, which can converge to the optimal solution faster [27], and usually requires less hyperparameter adjustment. Adam is suitable for processing large-scale data and parameters and typically converges to local optima faster than SGD. Therefore, the Adam optimizer is selected in this paper to calculate the gradient and optimize the loss function.

The selection of loss function is related to tasks. Different tasks need to choose different loss functions. In neural networks, there are two common types: one is the mean square error function for regression problems, and the other is the cross-entropy function for classification [28].

Jun variance loss function: it measures the sum of squares of the difference between the predicted value and the real value. The relation is expressed by Eq. (15.5):

$$J = \frac{1}{2m} \sum_{i=1}^{m} (z_i - y_i). \tag{15.5}$$

In Eq. (15.5) the variables have the following interpretations: m denotes the number of samples in the dataset, z represents the predicted value, y represents the real value, and Σ indicates the summation operation, which calculates the sum of squared differences between the predicted and real values across all samples.

Cross-entropy loss function: Cross-entropy is a concept in information theory, which is used to measure the difference between two probability distributions. In the classification problem, we regard the prediction result as a probability distribution and the real label as another probability distribution. The cross-entropy loss function is used to measure the difference between the two distributions. The relation is expressed by Eq. (15.6):

$$H(p, q) = -\sum_{x} p(x) \log q(x). \tag{15.6}$$

In Eq. (15.6), the variables have the following interpretations: $H(p, q)$ represents the cross-entropy loss value between two probability distributions p and q. Σ denotes the summation operation, which calculates the sum over all elements. $p(x)$ represents the probability of an event x according to the true distribution, while $q(x)$ represents the probability of the same event x according to the predicted or estimated distribution. The logarithm (log) is applied element-wise to the $q(x)$ values. Overall, the equation measures the dissimilarity between the true and predicted probability distributions based on their event probabilities.

15.4 Model Training

Measure the advantages and disadvantages of the model through the following indicators.

The second category confusion matrix is given in Table 15.1.

The evaluation indicators of the model are given in Table 15.2:

The network classification selected for the experimental model is the AlexNet, which is based on TensorFlow and runs in the environment of win10 + cuda9.1 + cudnn7 + tensorflow gpu-1.12.0, pythoch1.4.0 + keras-2.2.4.I n the training phase, the size of the input image is uniformly adjusted to a ratio of 300 × 300, and then, through data enhancement operations such as normalization and gray unification, the model is trained and optimized using the cross-entropy function and Adam algorithm. The initial learning rate is set to 0.01. After 15 epochs, the learning efficiency of this

Table 15.1 Two-category confusion matrix

	The true value is positive	The true value is negative
The predicted value is positive	TP (true positive)	FP (false positive)
The predicted value is negative	FN (false negative)	TN (true positive)

Table 15.2 Model evaluation indicators

	Formula description	Description
Accuracy	$\text{Acc} = \frac{\text{TP+TN}}{\text{TP+TN+FN+TN}}$	The ratio of sample size to total number for correct classification predicted by the model
Recall	$R = \frac{\text{TP}}{\text{TP+FN}}$	The number of positive classes correctly predicted by the model in the sample
Precision	$P = \frac{\text{TP}}{\text{TP+FP}}$	The proportion of the true sample size to the positive sample predicted by the model

method decreases by 0.005, and 100 epochs are learned in total. In order to enhance the model's ability to learn image features and improve accuracy in recognition, this experiment sets up a training set and a validation set and assigns the same participation for comparison. The comparison diagrams of the accuracy loss functions are shown in Figs. 15.5 and 15.6, respectively.

Through the above comparison results, it can be found that the training set and verification sets are relatively close in the accuracy comparison results, which verifies the feasibility of the model. However, in the comparison of the loss function, when epoch > 40, there is a relatively unstable trend.

A garbage classification and recognition system have been designed and implemented to assess the model's classification accuracy. The specific process is that the garbage images collected by the front-end system are preprocessed and feature extracted, and then transmitted to the backend classification system for classification. The backend system classifies the data transmitted from the front end through a trained classifier and outputs the classification results. If the classification results do not meet the user's needs, the backend system will feed back the classification results to the front-end system, so that the front-end system can re-collect or process

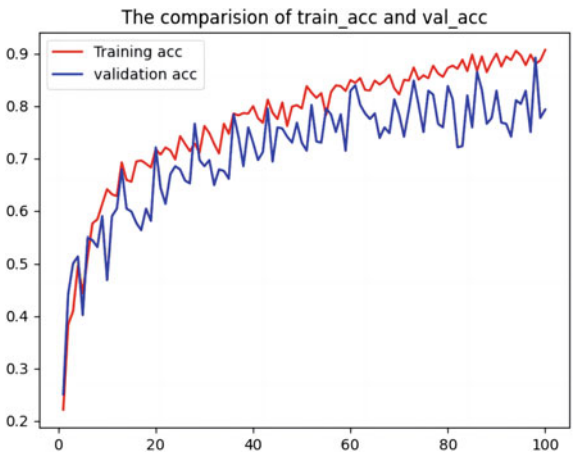


Fig. 15.5 Comparison of accuracy trends between two models

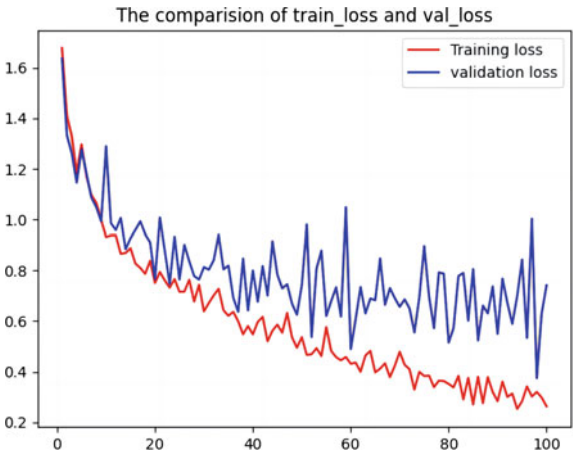


Fig. 15.6 Comparison of loss trends between two models

garbage images. Through the collaboration between the front-end and backend, the garbage classification and recognition system can efficiently and accurately classify garbage, achieving effective garbage classification management.

In the front-end input section, this article mainly adopts PyQt5 based on Python programming language to complete the system implementation. PyQt5 can create a modern user interface and improve the user experience; equipped with powerful event processing and signal slot mechanisms, it can achieve complex interaction logic; supporting multi-threaded programming can improve system performance and response speed; it can easily integrate other Python libraries and third-party tools to expand the functionality of the system; it has good cross-platform performance and can run on different operating systems.

15.5 Conclusion and Future Work

The garbage classification and recognition system based on deep learning proposed in this paper can efficiently and accurately classify garbage by training and testing different types of garbage images, with an accuracy rate of over 90%. Its main purpose is to automatically classify garbage. The system uses convolutional neural networks (CNN) for feature extraction and classification of garbage images and uses different types of garbage image datasets for training and testing.

The garbage classification system proposed in this paper includes the following parts: garbage image collection and preprocessing: By uploading garbage images and performing image preprocessing, including adjusting image size, cropping, cropping, and other operations, it facilitates subsequent feature extraction and classification. Design of deep learning model: The system adopts a CNN-based deep learning model for feature extraction and classification of garbage images. This model includes multiple convolutional layers, pooling layers, and fully connected layers, and uses a

softmax classifier for garbage classification. Implementation of garbage classification system: The paper implements a garbage classification system based on Python language, using TensorFlow deep learning framework and OpenCV image processing library for development and implementation. The main advantage of this system is that it adopts a deep learning model for feature extraction and classification of garbage images, with a high degree of automation and intelligence, which can effectively improve the efficiency and accuracy of garbage classification. In addition, the system is developed and implemented using Python language, which has high scalability and flexibility, making it convenient for secondary development and application. The system still has some limitations, such as being unable to recognize some special garbage and having an impact on the garbage classification effect under different angles and lighting conditions. In addition, there may be some deviations in the training and testing datasets of the system, which require further expansion and optimization.

Future research directions mainly include focusing on fitting problems in systems, optimizing the structure and parameters of deep learning models, and increasing the recognition ability for special garbage. In addition, multiple data sources and technological means can be considered to further improve the automation and intelligence of garbage classification, achieving more accurate and effective garbage classification. In summary, the garbage classification and recognition system based on deep learning has broad application prospects and important environmental significance, which needs further improvement and optimization.

Acknowledgements This research was funded by the National Natural Science Foundation of China (No.62262019), the Hainan Provincial Natural Science Foundation of China (No.823RC488, No.623RC481, No. 620RC603), the Haikou Science and Technology Plan Project of China (No.2022-016).

References

1. Qingxia, Z., Junyan, H.: Research on the factors influencing the learning behavior of public waste classification knowledge. J. Hebei Univ. Environ. Eng. 1–4 (2022)
2. Abuga, Y., Shuhui, L., Jinghua, W., Shanyang.: Optimization of computational cost for non-invasive load identification algorithm based on image classification network. Power Autom. Equip. 1–8 (2023)
3. Shangwang, B., Mengyao, W., Jing, H., Zhibo, C.: A fine-grained image classification network for multi region attention. Comput. Eng. 1–11 (2021)
4. Dongxu, J., Zhipeng, Z.: Exploration of the transformation of public goods supply mode for urban residents' household waste classification. Shanghai Urban Manag. **32**(02), 88–96 (2023)
5. Shuli, G., Baiyu, P., Kehao, J.: Research on the classification of domestic waste in Shandong Province. Clean World **39**(02), 172–174 (2023)
6. Fang, L., Qian, L., Minjun, T., Meixia, X.: Exploration of the development of garbage classification in China. Light Ind. Stand. Qual. **1**(01), 113–114 (2023)
7. Bei, W., Li, X.: Research on intelligent image recognition and classification of crops based on convolutional neural networks. Agric. Mechanization Res. **45**(12), 20–29 (2023)

8. Tengfei, S., Jiaxiu, S., Zonghang, W.: Research on image classification improvement methods based on convolutional neural networks modern information technology, **7**(05), 109–112 (2023)
9. Caijian, Y., Jun, C., Shiwei, W.: Household waste image classification based on lightweight convolutional neural networks. Softw. Eng. **26**(03), 41–65 (2023)
10. Huan, C., Liping, J., Huaiyu, Q.: Design of an intelligent garbage classification and identification system based on FPGA. Electron. Compon. Inf. Technol. **6**(12), 116–121 (2022)
11. Linesh, R., Mahendra, G.K., Rakshit, A., Vinayak, S., Manoj, S.: Comparison of garbage classification frameworks using transfer learning and CNN. Int. J. Soc. Ecol. Sustain. Dev. (IJSESD) **13**(9) (2023)
12. Mutu, C., T. Rui, T., Leilei, S., Yueqin, F.: Design of an intelligent garbage classification system based on deep learning. Electron. Testing **36**(17), 12–30 (2022)
13. Hao, D., Wang, J.: Design of an intelligent garbage classification system based on convolutional neural networks. J. Hunan Univ. Sci. Technol. **43**(03), 28–32 (2022)
14. Ying, W.: Research on classification algorithms for household garbage images based on deep learning. Chongqing University of Technology (2022)
15. Wang, L., Jiang, J., Li, N., Yin, R.: Design and implementation of garbage classification system under the "double carbon" target. Int. J. Front. Sociol. **4**(6) (2022)
16. Hairong, S., Yingjie, Z.: Photovoltaic infrared thermal image recognition method based on scatter plot AlexNet network. J. Solar Energy **44**(01), 55–61 (2023)
17. Yangyang, J., Pingzhi, L., Ailong, L., Songlin, L.: A classification method for map building shapes supported by AlexNet. J. Earth Inf. Sci. **24**(12), 2333–2341 (2022)
18. Bing, J., Huiying, W., Chen, C., Xingli, Z., Xianlong, W.: Research on facial emotion recognition technology based on AlexNet convolutional neural network. Comput. Knowl. Technol. **18**(27), 24–26 (2022)
19. QinBin, B., Ho, L.: Research on garbage classification based on convolutional neural network. J. Wirel. Commun. Technol. **28**, 51–56 (2019)
20. Chaohui, H., Pengcheng, Z.: Household waste classification based on deep learning. Comput. Knowl. Technol. **18**(03), 99–100 (2022)
21. Yoshimoto, Y., Kiyogawa, T., Hota, T., Ogasawara, T.: Classification of waste materials in dense container packaging using thermal imaging techniques. J. Nippon Inst. Sci. Technol. **40**(6) (2022)
22. Lanning.: Design method and implementation of image recognition based on android platform. Electron. Technol. Softw. Eng. **19**, 61–64b (2021)
23. Kangjian, T., Zhan, W., Wenzao, L.: Research and application of garbage image classification model based on convolutional neural network. J. Inf. Eng. Univ. **36**(4), 374–379 (2021)
24. Yufeng, C., Jianwen, C., Jiayi, H.: Radiological image recognition of viral pneumonia based on deep learning framework Keras. Electron. Compon. Inf. Technol. **5**(4), 148–150 (2021)
25. Pu, T., Chen, H.: Optimization and application of handwritten digit recognition model based on Tensorflow deep learning framework. Autom. Technol. Appl. **39**(12), 110–114 (2020)
26. Hengtao, W.: Integrated image recognition system based on TensorFlow, Keras, and OpenCV. Electron. Testing **24**, 53–64 (2020)
27. Yu, W., Mengjia, W., Weihong, Z.: Campus garbage image classification based on CNN and group normalization. J. Jilin Univ. (Inf. Sci. Edn.) **38**(6), 744–750 (2020)
28. Reyimu, G.M., Furong, C., Chen, L., Xiuhong, Y.: Comparison of mainstream deep learning frameworks. Electron. Technol. Softw. Eng. **1**(7), 74–89 (2020)

Chapter 16
Research on Portable Methane Telemetry System Based on TDLAS

Jianguo Jiang, Junkai Hu, and Chunlei Jiang

Abstract This work studies the key technology of a non-contact and long-distance methane-measuring instrument based on tunable diode laser absorption spectroscopy (TDLAS) due to the difficulties in real-time monitoring of methane leaks over a large area and the challenges of installing sensors. Using Beer–Lambert law as the theoretical basis, the spectral line information of methane gas recorded in the high-resolution transmission (HITRAN) database is compared to select the mature light source technology of the 1653.72 nm absorption line. Through simulation and analysis of the optical path within the detection distance of 0–100 m, the size of the optical system is designed to be minimized, and the sensor can be installed in portable equipment or monitoring probes. A reference gas chamber is specially designed in the optical components to provide a feedback optical signal for the system to eliminate the fluctuation of the light source signal and achieve automatic peak searching. The detection accuracy within ten meters is improved from 200 to 130 ppm, and the dynamic monitoring accuracy is maintained at 0.3–3.6% with a detection speed of one second. After multiple experiments, the system is proven to be stable and reliable with fast response speed, which meets the requirements of real-time methane gas detection applications.

16.1 Introduction

Methane has a wide range of applications, but when the concentration reaches a certain level and mixes with oxygen, it can cause explosions, fires, and other safety accidents, seriously endangering life and property safety [1]. In order to avoid huge losses of personnel and property during production and prevent serious atmospheric pollution caused by methane gas leaks [2, 3], monitoring methane distribution and leakage is of great significance in preventing accidents and leaks.

J. Jiang · J. Hu (✉) · C. Jiang
Northeast Petroleum University, Daqing Heilongjiang 163318, China
e-mail: hujunkai_nepu@163.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_16

Traditional detection methods [4–7] have serious drawbacks in terms of inability to remote sensing and a large number of blind spots in regional monitoring inspections. Compared with traditional detection methods, tunable diode laser absorption spectroscopy (TDLAS) technology has the advantages of high sensitivity, fast response, and detection of large space measurements [8–10]. Based on the TDLAS technology with an open-cavity detection method, this paper builds a detection system that can achieve remote sensing and improves the open optical path design. The added auxiliary gas chamber and reflector are used to achieve automatic peak searching through feedback signals, effectively improving the system's response speed and measurement accuracy.

16.2 Working Principle of Telemetry System

16.2.1 Beer–Lambert Law

When a specific light source is irradiated onto methane gas, the vibration of the molecules in the ground state of the methane gas changes, which is essentially caused by the change in dipole moment of the molecule due to the photon energy. Radiant frequencies of the light cause strong absorption of energy by the gas molecules and lead to transition, resulting in a strong absorption peak at a specific frequency on the spectrum of the light source. This shows the relationship between the strength of the absorption of a substance to a certain wavelength of light and the concentration of the absorbing substance and the thickness of its gas layer. When the laser beam passes through the gas layer, the difference between the outgoing and incoming light intensity is directly proportional to the amount of gas molecules, and the concentration of the gas can be calculated by determining the variation in the intensity of the incoming and outgoing light. This relationship is described by Beer–Lambert law as shown in the equation.

$$\mathrm{I}_t(\lambda) = \mathrm{I}_0(\lambda)\exp(-\alpha(\lambda)\mathrm{LC}) \tag{16.1}$$

where I_t is the output light intensity, I_0 is the input light intensity, C represents the concentration of the gas to be measured, L represents the optical path length. $\alpha(\lambda)$ represents the gas absorption coefficient, and it is expressed as:

$$a(\lambda) = S(T)\varphi(\lambda)N \tag{16.2}$$

$S(T)$ is the linear intensity of the gas absorption line; $\varphi(\lambda)$ is linear absorption function; N is the molecular density of the gas.

The linear intensity of the gas absorption spectral line can be obtained by querying the high-resolution transmission (HITRAN) molecular absorption database [11]. Common line shape functions include Lorentzian, Gaussian, and Voigt functions [12,

13]. Based on the characteristics of these three functions, when the remote sensing system works under normal temperature and pressure, the calculation results of the Lorentzian function tend to be closer to the Voigt function. In this state, the error of the function is small and the trend is good, so the Lorentzian line shape function is selected as the line shape absorption function for modeling and simulation. Converting the concentration according to Eq. (16.1) gives Eq. (16.3):

$$c = \frac{1}{a(\lambda)PL} \ln \frac{I_0}{I_t} \tag{16.3}$$

According to Eq. (16.3), the gas concentration is related to the absorption coefficient, pressure, optical path, and incident and transmitted light intensities. After the gas to be measured is determined, the absorption coefficient of the gas can be obtained by querying the HITRAN database [11]. Then, by digitally measuring the light intensity of the incident and transmitted light, the gas concentration within the optical path can be calculated. Using this theorem, selective measurement of methane gas mixed in the air can be achieved. The Beer–Lambert law provides the most basic theoretical support for gas concentration detection.

16.2.2 Selection of CH_4 Absorption Line

Selecting appropriate methane (CH_4) absorption lines can effectively improve the measurement accuracy of the methane leak telemetry system and reduce the impact of background interfering gases on the system detection accuracy. The selection of absorption lines mainly considers two aspects. First, interfering gases such as vapor (H_2O) and carbon dioxide (CO_2) in the air. Since nitrogen (N_2) and oxygen (O_2) have different elements from methane, the spectral absorption lines differ greatly, and nitrogen and oxygen can be excluded from the interference source. Second, the laser must meet the limitations of current technological processes, and the laser component costs are expensive. Selecting lasers with higher frequencies can reduce the cost of portable devices. Refer to the HITRAN 2022 database [11] to obtain all absorption lines in the visible and infrared wavelength range of 10–26,315 cm^{-1} as shown in Fig. 16.1, excluding all isotopes and locking the CH_4 absorption line at the peak point of the absorption line strength parameter of 6046.96 cm^{-1} (1653.7235 nm) and 1.455e-21 cm^{-1} mol^{-1} cm^{-2}.

16.3 Detection System Design

According to Beer–Lambert's law, methane gas can also selectively absorb specific frequencies of lasers. By using direct absorption spectroscopy (DAS), the absorption spectrum of the target gas molecule can be obtained, and concentration analysis

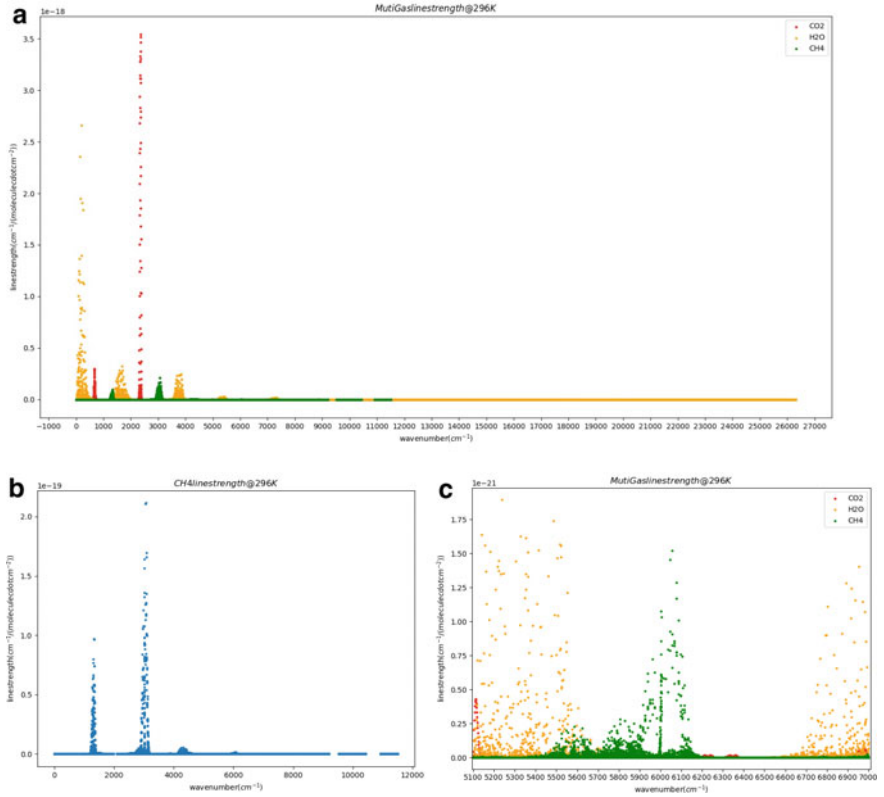


Fig. 16.1 Absorption lines of CH_4 in different ranges. **a** Absorption lines of H_2O, CO_2, and CH_4 in the range of 0–26,315 cm^{-1}; **b** absorption line of CH_4 in the range of 0–26,315 cm^{-1}; **c** absorption spectrum of CH_4 in the range of 5100–7000 cm^{-1}

can also be performed [14]. However, this method is only applicable in low-noise environments, and complex interference sources in open optical paths cannot be eliminated, so it cannot be effectively applied to remote sensing systems. The principle of wavelength modulation spectroscopy (WMS) technology was proposed by Reid and Labile [15], which extracts the second harmonics (2f) signal by a lock-in amplifier to filter out most of the interference noise. Compared with direct absorption spectroscopy, it more intuitively displays the relationship between gas concentration and signal peak, greatly reducing the noise of the detection system. The detection system design diagram is shown in Fig. 16.2, which consists of an optical part for methane detection and an electronic circuit for TDLAS signal processing. The function of low-pass Filter (LPF) is to obtain a higher signal-to-noise ratio (SNR) for the phase-locked amplifier.

Distributed feedback laser (DFB) is the most critical component in TDLAS technology, and the entire hardware system design is actually centered around the DFB laser. The methane absorption peak is very narrow, and there is no laser that can

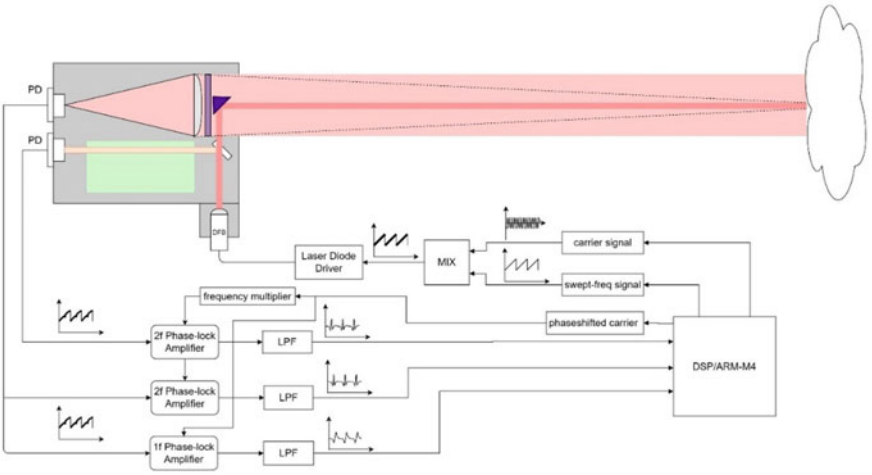


Fig. 16.2 Diagram of wavelength modulation technique for remote sensing system

continuously output precise 1653.7 nm laser. The DFB laser has the characteristic of wavelength controlled by the driving current. The embedded component generates a 38 kHz modulation carrier and a 5 Hz sweep signal, which is modulated by a mixer. The DFB emission current is controlled by the laser driver, with a driving current of 25–40 mA. The laser wavelength is adjusted by controlling the driving current of the DFB laser until it coincides with the absorption peak. As shown in Fig. 16.3, to ensure the stability of the light source, the DFB laser must be precisely temperature-controlled. The LPF in the figure plays the same role as in Fig. 16.2.

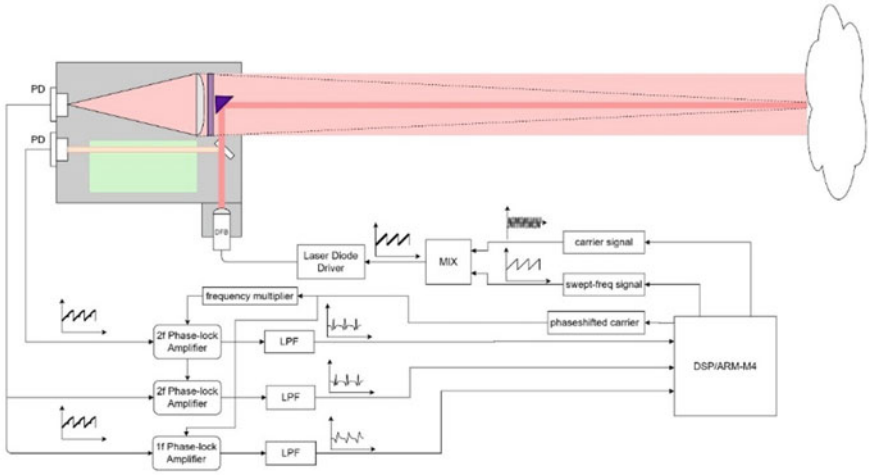


Fig. 16.3 Temperature–wavelength relationship curve

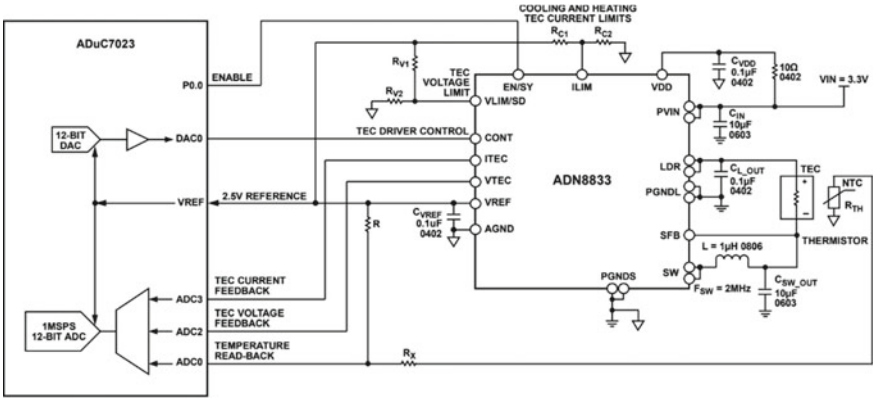


Fig. 16.4 Schematic diagram of ADN8833 drive circuit

After cost and control performance comparison, ADN8833 controller was ultimately chosen as the thermos-electric cooler (TEC) controller for the laser tube. ADN8833 reads the feedback voltage of the thermistor and provides a driving voltage to the TEC, with the microcontroller monitoring and controlling the thermal loop as shown in Fig. 16.4.

16.4 Optical System Design and Simulation

16.4.1 Optical Component Design

The infrared light emitted by the DFB is reflected by a total internal reflection right-angle reflector and directed toward a white diffuse reflection target. The diffuse light spot at the target forms a real image at 50.04–55 mm behind the collecting lens. The signal strength of the spot on the photodiode (PD) sensor mainly depends on the distance to the target and the scattering loss. The installation positions of the laser optical components are shown in Fig. 16.5, with the collimating lens made of K9L material and the filter using a 1653 nm narrowband filter.

16.4.2 Simulation and Verification of Optical Systems

When a high concentration of methane gas appears in the target area, the laser is emitted from the photosensitive window, passes through the measured space, and reaches the target. The light spot on the target is projected onto the PD1 through the collecting lens. The optical design of the telemetry system needs to consider

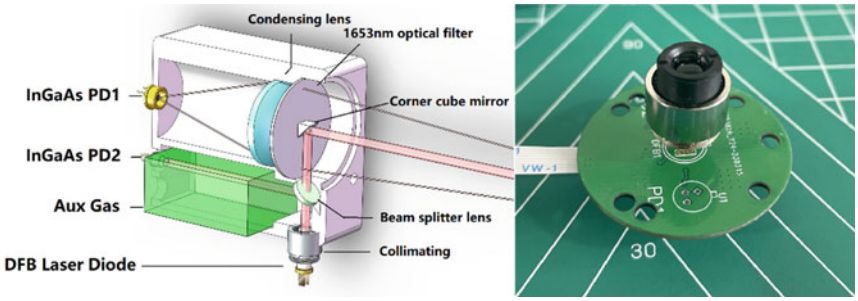


Fig. 16.5 3D design of the optical system and installed collimator DFB components

the position and size of the imaging spot on the photosensitive element surface of the lens for different placement positions of the reflecting target surface. To avoid uncontrollable factors beyond the design range during the actual adjustment of the components, ray-optics software is used to simulate the optical path and determine the optical path of the sensor photosensitive interface under each target position. Due to the limitation of the overall design size, the focal length of the lens is limited to 50 mm and the lens diameter is 50 mm, and the result is shown in Fig. 16.6.

The effective reception efficiency of the fiber at the receiving surface of the element is calculated according to the optical imaging calculation principle. Besides, based on the calculation of a 50 mm focal length and 50 mm diameter lens, the optimal measurement distance is 0.1–4 m, and the phototube can receive a maximum of 300 uA photocurrent, which is of great significance for the design of the photoelectric receiving circuit and the selection of the analog to digital converter (ADC) (Fig. 16.7).

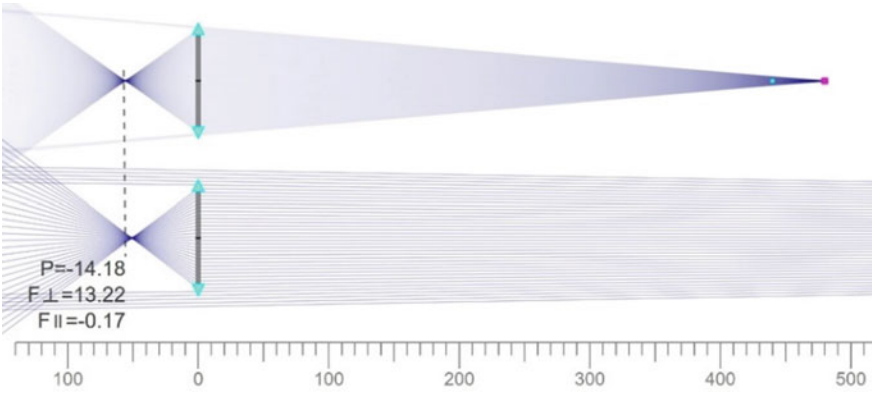


Fig. 16.6 Imaging optical path of the condenser at a distance of 0.5 and 50 m from the reflective surface

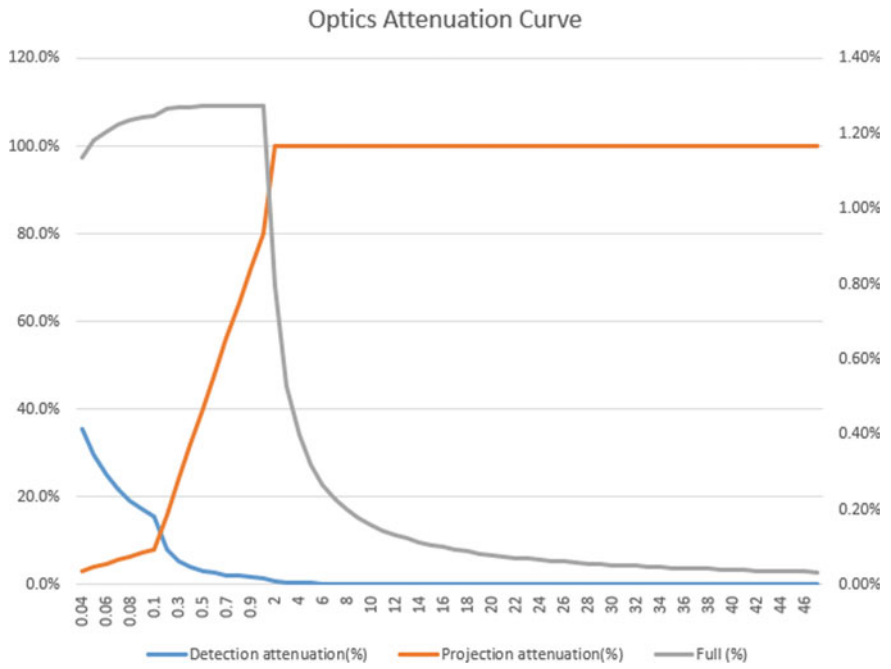


Fig. 16.7 Effect of imaging distance on the light intensity received by the sensor

16.5 Methane Concentration Detection Experiments

Experimental method: Using nitrogen as the balance gas, different concentration gas bags were obtained through a gas distribution device at 1 atm (1 atm = 101,325 Pa) to simulate methane plumes in actual detection scenarios. The arrangement of the gas plumes in the experiment is shown in Fig. 16.8.

Signal acquisition part is ultimately using NI PXle-6366 as the acquisition card for observing processed waveforms. The upper computer uses PCI-1DA 1.5.1 for analysis. The open-cavity remote sensing system differs from the inhalation detection device. The concentration calculation results will be expressed in ppm.m form. According to the Beer–Lambert law, the longer the optical path, the higher the absorption peak signal intensity. However, under diffuse reflection imaging, the farther the distance, the lower the received signal intensity, until it is difficult to distinguish. To verify the instrument stability, it is necessary to first calibrate the relationship between the voltage amplitude of the absorption spectrum line and the concentration, and then measure the same concentration gas bag at different distances and compare the data results.

Fig. 16.8 Gas bag simulating a leaked gas plume

16.5.1 Measurement Calibration

Nitrogen was used as the background gas for the measurement experiment. A gas distribution system was built using a precision flow meter, and five gas samples of different concentrations, namely 987, 1000, 1500, 5000, and 10,000 ppm, were prepared. The second harmonic signals of the extracted methane gas at different concentrations are shown in Fig. 16.9.

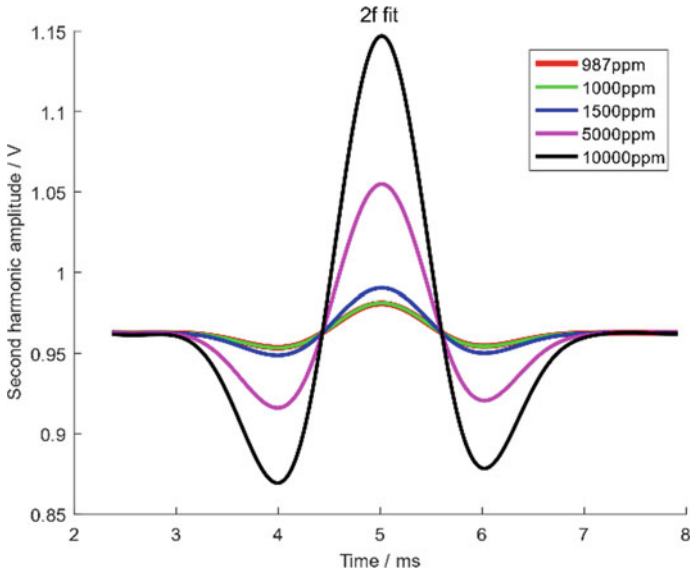


Fig. 16.9 2f signal intensity curves at different concentrations

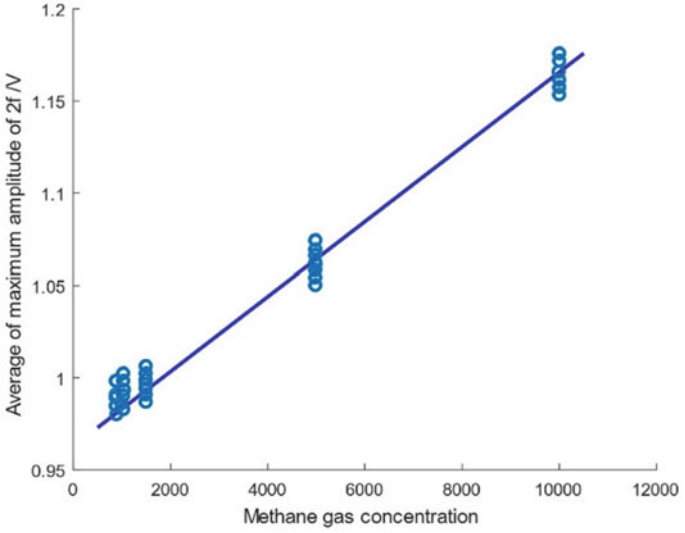


Fig. 16.10 Relationship between methane gas concentration and 2f signal intensity

Multiple measurements at different concentrations are required to obtain the median average value of ten sets of second harmonic signals for each concentration. There is a good linear relationship between the light intensity and the electrical signal generated by the sensor. By linear fitting, the data relationship between the median average value of the second harmonic amplitude (Amp) of the received light intensity signal and the methane gas concentration is obtained (see Fig. 16.10).

The fitting curve equation is Eq. (16.4):

$$C = 4484.0514 \cdot \max(\text{Amp}) - 0.9628 \tag{16.4}$$

According to the Beer–Lambert law, the absorption line strength is positively correlated with concentration. The 2f signal does not change the linear relationship between line strength and concentration. The calibration parameters obtained by linear fitting can be effectively used for the detection device.

16.5.2 Gas Measurement Accuracy Experiments

Standard concentrations of methane gas at 81.6, 90.9, 136.4, 454.5, and 909.1 ppm.m are prepared, and the signal concentrations are measured and calculated to determine the errors between the true and measured gas concentrations as given in Table 16.1. Based on observation, the relative errors between the true and measured gas concentrations are all within 7%. Furthermore, as shown in Fig. 16.11, the 2f harmonic amplitude still maintains a good linear relationship with the methane gas concentration.

Table 16.1 Relative error between proportioned concentration and measured concentration of standard methane gas

Item	Gas strength/ppm.m				
	81.6	90.9	136.4	454.5	909.1
Average 2f Amp/V	0.9805	0.9845	0.9953	1.0668	1.1634
Measuring concentration/ppm.m	79.1339	97.4669	145.7474	465.9606	899.0523
Relative error (%)	3.0	7.2	6.9	2.5	1.1

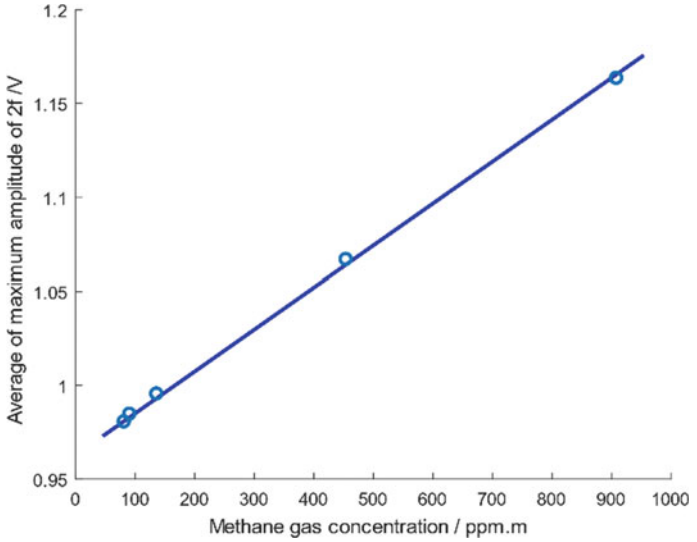


Fig. 16.11 Actual measurement concentration measurement linearity

According to the data in the table, the relative error between the real concentration of the measured gas and the concentration calculated by the measured signal is within 7.2%. It can be seen from the figure that the amplitude of the second harmonic and the concentration of methane gas still maintain a good linear relationship.

16.6 Summary

With TDLAS technology as the research background for remote measurement of methane gas concentration, wavelength modulation technology is used to suppress noise and improve detection accuracy. Furthermore, the mechanical structure of the sensor and the positioning of optical components are determined through simulation. Specifically, a precise flow meter was used in the experiment to mix samples for calibration, and through multiple experiments, a linear mathematical relationship

between the peak-to-peak value of the second harmonic signal and the methane gas concentration was determined. The experiment was repeated with other standard concentration gases to determine the actual error range of the detection system is less than 8%, which meets the requirements for industrial safety monitoring 16, proving the engineering application value of the device.

Acknowledgements This article was supported by grants from the Funding. Natural Science Foundation of Heilongjiang Province (LH2021F008).

References

1. Greenfield, P.F., Batstone, D.J.: Anaerobic digestion: impact of future GHG mitigation policies on methane generation and usage. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. **52**(1–2), 39–47 (2005)
2. Akimoto, H.: Global air quality and pollution. Science (2003)
3. Hodgkinson, J., Tatam, R.P.: Optical gas sensing: a review. Meas. Sci. Technol. **24**, 012004 (2013)
4. Dosi, M., Lau, I., Zhuang, Y., Simakov, D., Fowler, M.W., Pope, A.: Ultra-sensitive electrochemical methane sensors based on solid polymer electrolyte-infused laser-induced graphene. ACS Appl. Mater. Interfaces. **11**(6), 6166–6173 (2019)
5. Wan, H., Yin, H.Y., Lin, L., Zeng, X.Q., Mason, A.J.: Miniaturized planar room temperature ionic liquid electrochemical gas sensor for rapid multiple gas pollutants monitoring. Sens. Actuators, B Chem. **255**(1), 638–646 (2018)
6. Sha, M., Ma, X., Li, N., Luo, F., Fayer, M.D.: Dynamical properties of a room temperature ionic liquid: using molecular dynamics simulations to implement a dynamic ion cage model. J. Chem. Phys. **151**(15), 154502 (2019)
7. Zeng, X.: Ionic liquids: solvents and electrolytes for chemical sensor development. In: The 14th Beijing Conference and Exhibition on Instrumental Analysis (BCEIA 2011)
8. Shen, S., Li, W., Wang, M., Wang, D., Li, Y., Li, D.: Methane near-infrared laser remote detection under non-cooperative target condition based on harmonic waveform recognition. Infrared Phys. Technol. **120**, 103977 (2022)
9. Lu, J.Y., Hong, L., Dong, Y., Zhang, Y.S.: A new wavelet threshold function and denoising application. Math. Probl. Eng. **2016**(5), 1–8 (2016)
10. Benoy, T., Lengden, M., Stewart, G., Johnstone, W.: Recovery of absorption line shapes with correction for the wavelength modulation characteristics of DFB lasers. IEEE Photonics J. **8**(3), 1–17 (2016)
11. Mellau, G.C., Makhnev, V.Y., Gordon, I.E., Zobov, N.F., Tennyson, J., Polyansky, O.L.: An experimentally-accurate and complete room-temperature infrared HCN line-list for the HITRAN database. J. Quant. Spectrosc. Radiat. Transfer **279**, 107666 (2021)
12. Wahlquist, H.: Modulation broadening of unsaturated Lorentzian lines. J. Chem. Phys. **35**(5), 1708–1710 (1961)
13. Hu, C., X. Chen, Z. Li.: The gas temperature compensation research based on TDLAS technology. In: Chan, A.P.C., Hong, W.C., Mellal, M.A. (eds.) The 3rd International Conference on Material, Mechanical and Manufacturing Engineering 2015, AES, Atlantis Press, Netherlands (2015)
14. Reid, J., Labrie, D.: Second-harmonic detection with tunable diode lasers—comparison of experiment and theory. Appl. Phys. **26**, 203–210 (1981)
15. Standardization Administration of the People's Republic of China. GB/T 33672–2017 Cavity Ring-Down Spectroscopy system for measurement of atmospheric methane (2017)

Chapter 17
Recommendation Algorithm Based on Wide&Deep and FM

Songkun Zheng, Xian Li, Xueliang Chen, and Xu Li

Abstract Online learning is more and more popular because it is not limited by time and space. How to choose a suitable course from thousands of online courses is a great challenge faced by online learners, and online course recommendation came into being. The personalized recommendation algorithm analyzes the user's preferences by collecting some previous historical records of the user and other information, and generates recommendations for the user. Since Wide&Deep was proposed, due to its inherent ease of implementation, adaptability, and versatility, this approach has gained significant traction across various industry sectors. But its feature intersection method is not efficient. Sufficient feature engineering is required to provide informative features that can effectively distinguish objects. In this paper, the WD-FM model is proposed by combining Wide&Deep and factorization machine, and good results have been achieved through experimental demonstration.

17.1 Introduction

With the advent of the era of data explosion, it is necessary to correctly solve the previous problem of information scarcity and the current problem of information overload. In the face of massive data, it is very difficult to obtain the necessary information accurately and effectively. Therefore, information filtering techniques should be used. Currently, information filtering technology is mainly divided into search engine technology and recommendation system technology [1].

S. Zheng · X. Li · X. Chen
Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang, China

University of Chinese Academy of Sciences, Beijing, China

S. Zheng
e-mail: zhengsongkun20@mails.ucas.ac.cn

X. Li (✉)
Beijing Zhongke Zhihe Digital Technology Co., Ltd., Beijing, China
e-mail: lixu181@mails.ucas.ac.cn

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_17

Classification retrieval and search engines alleviate the problem of information overload. When the information classification is inaccurate or the user enters too few keywords, the user's retrieval time will be increased and the retrieval results will be affected. At present, many fields have begun to introduce personalized recommendation systems, and it is particularly important to convert large amounts of data into valuable information [2].

As an important tool for information filtering, personalized recommendations are a potential solution to the current information overload problem. The personalized recommendation algorithm analyzes the user's preferences by collecting some previous historical records of the user and other information, and generates recommendations for the user [3, 4]. The recommendation system can provide services that meet the individual needs of different users, improve the efficiency of users in finding knowledge from information, thereby effectively retaining users and making the website invincible.

The results of this paper can be summarized in the following two points:

1. We propose a model called WD-FM, which consists of Wide&Deep and FM (factorization machine).
2. We propose a new resource recommendation method based on WD-FM model. It has the function of feature memory, and separates low-order features and high-order features, and finally inputs them into the same output layer to improve the recommendation accuracy.

17.2 Related Work

Recently, researchers have done a lot of research in the field of recommendation, and the recommendation system model represented by collaborative filtering used in traditional recommendation systems [5]. Because it is difficult to solve problems such as data sparsity and cold start, the recommendation effect is unsatisfactory, especially when dealing with the huge amount of data generated at present.

Based on the assumption that user preferences are influenced by a small number of latent factors, matrix factorization is widely used in recommender systems and that an item's rating depends on how each of its characteristic factors is applied to user preferences.

MF (matrix factorization) decomposes the user item scoring matrix into the product of two or more low-dimensional matrices to achieve dimensionality reduction, and uses the low-dimensional spatial data mainly for non-negative matrix decomposition and matrix generalization decomposition Among them, the non-negative matrix factorization (NMF) method is to decompose the user's rating matrix $R_{n \times m}$ into two real-valued non-negative matrices $U_{n \times k}$ and $V_{k \times m}$, so that $R \approx U^{\mathrm{T}} V$.

Matrix decomposition is used for modeling. Usually, build a matrix of 1/0 values from the interaction data, and then decompose the matrix into two lower-dimensional matrices. One of the matrices has the same number of rows as the number of users, and each row represents a latent feature vector of a user. For example, in the literature [6], an $N \times D$ matrix C is used to represent the performance of users on the forum, where each row represents the user n who has posted at least once on the forum, and each column d represents the defined in the text A class label among the five behavioral dimensions of learners in the forum. Each entry C_{nd} of C is 1 if user n published at least one post assigned a content label of d, and 0 otherwise. Therefore, C is a matrix with a value of 1/0, and then the Bayesian non-negative matrix factorization (BNMF) method is used for C to generate user latent feature vectors. Literature [7] first regards the user's click, reading or use of resources as an interaction, thus forming a user-resource interaction matrix. The generalized matrix factorization (GMF) method is used to decompose it into hidden feature vectors of users and resources. In order to incorporate the characteristics of long-term interaction between users and resources, the model also combines long short-term memory (LSTM) to further generate fusion of users and resources. Hidden eigenvectors, and after combining the two eigenvectors, they share the same sigmoid output layer.

Under the background of big data, deep learning technology is more and more introduced into the core model design of the recommendation system, and the same is true for movie recommendation systems [8, 9]. Deep learning has brought a revolutionary impact to the recommendation system and can significantly improve the effect of the recommendation system. There are two main reasons. One is that deep learning greatly enhances the fitting ability of the recommendation model, and the other is that the deep learning model can utilize the model. The structure simulates different user behavior processes such as changes in user interest and user attention mechanisms. Deep crossing [10] was the first model proposed. Compared with multilayer perceptron (MLP), deep crossing adds an embedding layer between original features and MLP. Convert the input sparse features into dense embedding vectors, and then participate in the training in the MLP layer, which solves the problem that MLP is not good at dealing with sparse features. Convert the input sparse features into dense embedding vectors, and then participate in the training in the MLP layer, which solves the problem that MLP is not good at dealing with sparse features.

The Wide&Deep [11] recommendation model proposed by Google combines a deep MLP with a single-layer neural network, and at the same time gives the network good memory and generalization. Since its proposal, Wide&Deep has been widely used in the industry by virtue of its characteristics of easy to implement, easy to implement, and easy to transform.

17.3 Preliminary Knowledge

This section provides a brief review of how linear regression and multilayer perceptrons work.

17.3.1 Wide&Deep Model

The Wide&Deep model is divided into the wide side and the deep side. The wide side mainly uses logistic regression in the generalized linear model. Generally speaking, it is the weight multiplied by the feature plus the bias, and then thrown into the sigmoid function, and finally the probability of predicting whether or not. Overall, the wide part of the Wide&Deep model creates the interaction between features through a linear model to provide the ability to model a wide range of features. By combining the nonlinear feature extraction capabilities of the deep part, the Wide&Deep model can simultaneously consider both wide and deep features in tasks such as recommender systems, thereby improving the model's expressiveness and prediction accuracy. Its structure diagram is shown in Fig. 17.1.

The logistic regression formula is as follows:

$$y = w^T x + b \tag{17.1}$$

y represents the final prediction result, x is a vector, which is a vector of n features, and w is the weight corresponding to each feature. b represents the bias. The feature set includes the original input features and transformed features.

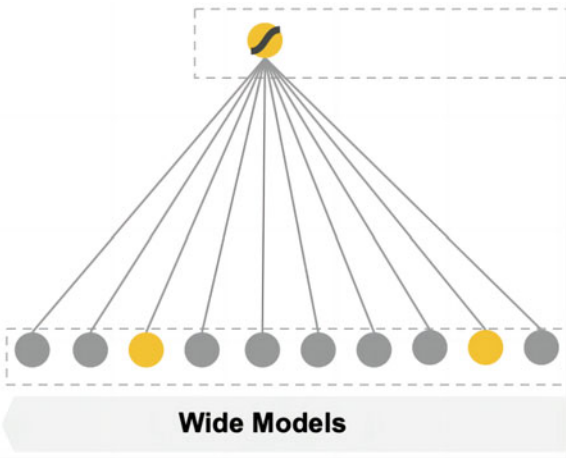


Fig. 17.1 Wide side of the Wide&Deep model

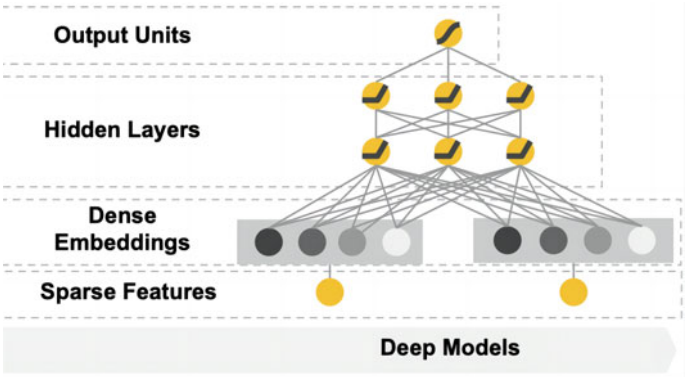


Fig. 17.2 Deep side of the Wide&Deep model

Among them, the optimizer of LR (logistic regression) is different from the past. In the past, stochastic gradient descent (SGD) was used, while the wide model used follow-the-regularized-leader (FTRL) published by Google on kdd in 2013. FTRL mainly fine-tunes the gradient. The idea is to hope that the new solution will not match the current solution. If the difference is too much, make the gradient step smaller. In addition, L1 regularization still needs to be added to make the solution found a little sparser.

The deep side is a multilayer perceptron, as shown in Fig. 17.2.

Each vector in the sparse and high-dimensional classification features is first converted into a low-dimensional and dense vector, which is called an embedding vector. The number of dimensions to join is typically on the order of O(10)–O(100). The embedding vectors are initialized randomly and trained concurrently during model training with the ultimate goal of minimizing the loss function. Low-dimensional dense embedding vectors are processed in the hidden layers of the feed-through neural network.

$$a^{(l+1)} = f\left(W^{(l)}a^{(l)} + b^{(l)}\right) \tag{17.2}$$

In the above formula, l represents the number of layers and f represents the activation function, usually a rectified linear unit (ReLU), where $a^{(l)}$, $b^{(l)}$, and $W^{(l)}$ are the activation function, bias, and model weights of layer l, respectively.

In the Wide part, after some nonlinear transformation and cross-combination of the input features, they are input into the linear model for modeling, and the interaction weights between the features are learned. The Deep part passes the input features through a series of hidden layers, and each hidden layer contains multiple neurons and performs feature transformation and abstraction layer by layer. Finally, after processing the depth part, a high-dimensional feature representation is obtained. The Wide&Deep model fuses the output of the breadth part and the depth part, which can be fused by simple weighted summation or other methods.

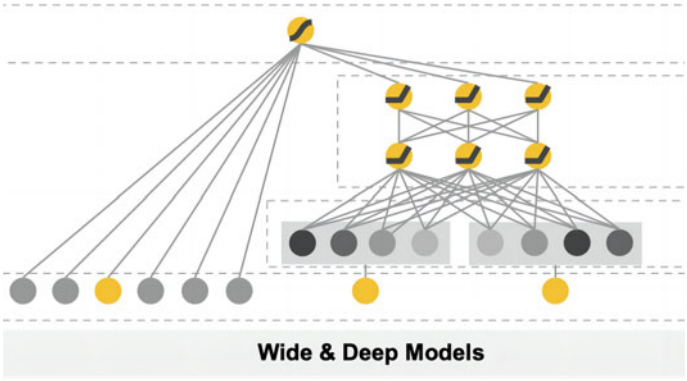


Fig. 17.3 Wide&Deep model

The combined model of wide and deep is shown in Fig. 17.3.
The prediction formula of the model is as follows:

$$P(Y=1|x)=\sigma\left(w_{\text{wide}}^{T}[x,\emptyset(x)]+w_{\text{deep}}^{T}a^{(l_f)}+b\right) \tag{17.3}$$

In the above formula, Y is a binary class label, $\sigma(\cdot)$ represents the activation function, which is a sigmoid, $\varphi(x)$ represents the cross-product transformation of the initial feature x, and b represents a bias term. W_{wide} is the weight vector corresponding to the wide side vector, and W_{deep} is the weight to activate $a(^{l}{}_{f})$ when computing.

17.3.2 FM Model

FM is a supervised learning method. It is mainly used for click-through-rate (CTR) estimation and is suitable for high-dimensional sparseness. The advantage is that it can automatically combine cross-features. The FM formula is as follows:

$$\hat{y}(x):=w_0+\sum_{i=1}^{n}w_i x_i+\sum_{i=1}^{n}\sum_{j=i+1}^{n}\langle v_i,v_j\rangle x_i x_j \tag{17.4}$$

V_i is the hidden vector of the ith dimension feature, and $<\ ,\ >$ represents the vector dot product. $\hat{y}$ is the predicted output value for the sample. w_0 is the bias term, representing the global bias of the model. w_i is the linear weight of the ith feature, used to represent the contribution of the feature to the predicted output. x_i is the value of the ith feature in the sample x.

Feature combination is a problem encountered in many machine learning modeling processes. If you model directly, you may ignore the correlation information between features. Therefore, you can improve the effect of the model by building new cross-features. In fact, it is to increase the feature intersection term. In the general linear model, each feature is considered independently, without considering the relationship between features. But in reality, there are associations between a large number of features.

High-dimensional sparse matrix is a common problem in practical engineering, which directly leads to excessive calculation and slow update of feature weights.

The advantage of FM lies in the handling of these two aspects of the problem. The first is the combination of features, through the combination of two features, the introduction of cross-features to improve the model score. The second is the high-dimensional disaster, which estimates the characteristic parameters by introducing hidden vectors.

17.4 Recommendation Model

The problem of feature combination and feature intersection is very common. In practical applications, there are many more types of features, and the complexity of feature intersection is also much greater [12].

The key to solving this problem is the ability of the model to learn feature combinations and feature intersections. This is because it is the key to determining the model's ability to predict unknown features combined with the sample and measuring its recommendation effectiveness for complex recommendation problems.

Wide&Deep does not carry out special processing on feature intersection, but directly sends independent features into the neural network, allowing them to be freely combined in the network [13]. Such feature intersection methods are not efficient. Although neural network has strong fitting ability, the premise is that there are any multilayer network and any number of neurons.

In the case of limited training resources and limited time for parameter adjustment, MLP is actually relatively inefficient for feature intersection processing. MLP connects all features together into a feature vector through the concatenate layer, there is no feature intersection, and there is no relationship between two features [14].

FM is a classic traditional machine learning model for solving feature intersection problems. FM will use a unique layer FM layer to specifically deal with the intersection between features, and there are multiple inner product operation units in the FM layer to combine different feature vectors pairwise [15]. Through the inner product operation of the two features in the FM layer, the features can be fully combined. Combine Wide&Deep with FM to generate a new model with strong feature combination ability, fitting ability and memory ability.

The model as show in Fig. 17.4.

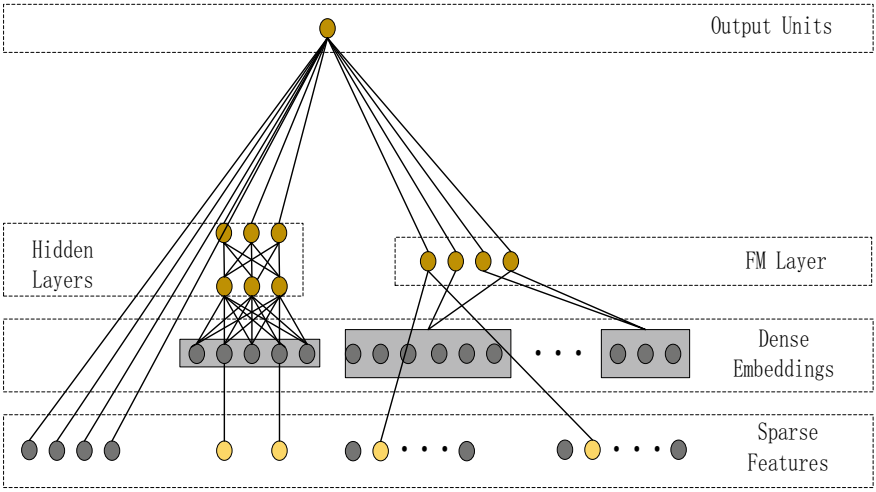


Fig. 17.4 WD-FM model

17.5 Experiments

17.5.1 Datasets

The dataset we used was the MovieLens dataset [16], on which the area under curve (AUC) and accuracy of our proposed Wide&Deep and FM model were evaluated. MovieLens: It is a non-commercial, research-oriented experimental site. The GroupLens research team created this dataset from data provided by users of the MovieLens website. This dataset consists of multiple movie rating datasets, each serving a different purpose. The collection is divided into several sub-datasets based on creation time, dataset size, etc. Each dataset varies in format, size, and purpose. This article uses the MovieLens 1 M dataset. MovieLens' 1 M dataset consists of 1,000,209 anonymous reviews of approximately 3900 movies. Users of these reviews joined MovieLens in 2000.

17.5.2 Evaluation Protocols

We divide the sample into training set and test set during evaluation, but splitting the sample is far from enough. In order to compare the quality of the model, we need to use some indicators to measure it.

We used the following evaluation metrics in the experiments in this paper: area under ROC (AUC) and logloss (cross-entropy).

The receiver operating characteristic (ROC) curve is a very commonly used indicator to measure the comprehensive performance of the model. ROC was first born in the military field, and then widely used in the medical field.

The x-axis in the ROC coordinate axis represents the false positive rate, and the y-axis in the curve represents the true positive rate.

The definitions of these two indicators are as follows:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{N} \tag{17.4}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{P} \tag{17.5}$$

In the above formula, P represents how many real positive samples there are, N represents the number of real negative samples, TP refers to how many P positive samples are predicted by the model as positive samples, and FP refers to how many N negative samples are classified as predicted by the model positive sample.

17.5.3 Baseline Algorithms

Wide&Deep: The Wide&Deep model includes two parts, namely the Wide part and the Deep part. The Wide side model is a generalized linear model, which can be expressed as $y = w^{\mathrm{T}} + b$. y represents the predicted output variable, w represents the weight vector, which is used to multiply the individual components of the input feature vector to weight them, and b stands for bias or intercept. The Deep side model is a typical deep neural networks (DNN) model.

DeepFM [17]: DeepFM is a combination of DNN and FM. On the basis of the Wide&Deep structure, FM is used to replace the LR of the Wide part, which can avoid artificially constructing complex feature projects. FM extracts low-level combined features, deep extracts high-level combined features, performs end-to-end joint training, and shares input embeddings.

17.5.4 Experimental Results

The main purpose of this experiment was to verify the accuracy of model by open resources data.

In this article, a high-performance server system with Ubuntu 20.04 LTS and a high-performance NVIDIA GeForce RTX 3080 graphics card was used. A deep learning framework is used called tensorflow-gpu-2.4.0. Use Pycharm under windows to connect to the server system remotely, and use Python language for software development. The specific experimental configuration is given in Table 17.1.

Table 17.1 Server software and hardware parameters

Name	Version
Operating system	Ubuntu 20.04 LTS
GPU	NVIDIA GeForce RTX 3080 GPU
CPU	Intel(R) Core(TM) i9-11900K CPU
Python	3.7.10
TensorFlow	2.4.0

Table 17.2 Comparison of models

Model	AUC	Logloss
Wide&Deep	0.7285	0.6065
DeepFM	0.6869	0.7721
WD-FM	0.7371	0.6010

As the amount of data increases, due to the addition of more redundant information, the accuracy of our prediction model begins to decline (Table 17.2).

In the mixed model structure of Wide&Deep, the wide side provides the model with a strong memory ability, and the deep side provides the model with a strong generalization ability. This structure allows the model to have both the advantages of logistic regression and the advantages of deep neural networks. In this way, a large number of historical behavioral characteristics can be memorized, and the ability to express has also been enhanced.

The prediction ability of the model for unknown feature combination samples depends on the feature combination and feature intersection. The wide part of DeepFM is FM, and FM can handle feature intersection well. There are multiple inner product operation units inside it. Two combinations, through this inner product operation, the features can be fully combined, and the prediction effect can be further improved. FM and Wide&Deep are combined to generate a brand new model with strong feature combination ability and strong fitting ability. Based on this, in order to make the model have memory ability, the three are combined. The memory ability can remember some rules in the data. Deep and FM can handle low-order feature combinations and high-order features well, and the prediction effect is compared with Wide&Deep and DeepFM. There are some improvements.

17.6 Conclusion

Recommendations are becoming more and more important in the era of big data. This paper proposes a new combinatorial recommendation model, called WD-FM. It combines the Wide & Deep and factorization machine (FM) models. Extensive experiments on the MovieLens dataset show that our model is improved in terms of effectiveness and accuracy.

References

1. Sedhain, S., Menon, A.K., Sanner, S., et al.: Autorec: autoencoders meet collaborative filtering. In: Proceedings of the 24th international conference on World Wide Web, pp. 111–112. United States (2015)
2. Mansur, F., Patel, V., Patel, M.: A review on recommender systems. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), IEEE, pp. 1–6. India (2017)
3. Ponnam, L.T., Punyasamudram, S.D., Nallagulla, S.N., et al.: Movie recommender system using item based collaborative filtering technique. In: International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), IEEE, pp. 1–5. India (2016)
4. Gupta, M., Thakkar, A., Gupta, V., et al.: Movie recommender system using collaborative filtering. In: International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, pp. 415–420. India (2020)
5. Alhijawi, B., Kilani, Y.: The recommender system: a survey. Int. J. Adv. Intell. Paradigms **15**(3), 229–251 (2020)
6. Gillani, N., Eynon, R., Osborne, M., et al.: Communication communities in MOOCs. arXiv preprint arXiv:1403.4640 (2014)
7. Li, J., Chang, C., Yang, Z., et al.: Probability matrix factorization algorithm for course recommendation system fusing the influence of nearest neighbor users based on cloud model. In: International Conference on Human Centered Computing, Springer, Cham, pp. 488–496. (2018)
8. Qu, Y., Cai, H., Ren, K., et al.: Product-based neural networks for user response prediction. In: IEEE 16th International Conference on Data Mining (ICDM), IEEE, pp. 1149–1154. Barcelona, Spain (2018)
9. Jais, I.K.M., Ismail, A.R., Nisa, S.Q.: Adam optimization algorithm for wide and deep neural network. Knowl. Eng. Data Sci. **2**(1), 41–46 (2016)
10. Shan, Y., Hoens, T.R., Jiao, J., et al.: Deep crossing: web-scale modeling without manually crafted combinatorial features. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 255–262 (2016)
11. Cheng, H.T., Koc, L., Harmsen J, et al.: Wide & deep learning for recommender systems. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. pp. 7–10 (2016)
12. Yuan, W., Wang, H., Hu, B., et al.: Wide and deep model of multi-source information-aware recommender system. IEEE Access **6**, 49385–49398 (2018)
13. Wang, H., Wang, N., Yeung, D.Y.: Collaborative deep learning for recommender systems. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1235–1244 (2015)
14. Xu, J., Hu, Z., Zou, J.: Personalized product recommendation method for analyzing user behavior using DeepFM. J. Inf. Process. Syst. **17**(2), 369–384 (2021)
15. Chen J, Sun B, Li H, et al.: Deep CTR prediction in display advertising. In: Proceedings of the 24th ACM international Conference on Multimedia, pp. 811–820 (2016)
16. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. ACM Trans. Interact. Intell. Syst. (TIIS) **5**(4), 1–19 (2015)
17. Guo, H., Tang, R., Ye, Y., et al.: DeepFM: a factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247 (2017)

Chapter 18
Simulation of E-Commerce Big Data Classification Model Based on Artificial Intelligence Algorithm

Yanfang Li and Sigen Song

Abstract In the era of artificial intelligence (AI), it has become an essential tool for e-commerce platform to analyze the real consumption potential of its users. In the classification and mining of e-commerce BD (big data), the low accuracy of the algorithm is the main reason that restricts the development of e-commerce. For this reason, in order to solve the problem that e-commerce needs to classify and process data quickly, this paper combines clustering theory, fuzzy logic theory, and artificial neural network (ANN) theory to design an e-commerce BD classification model based on AI algorithm. The simulation results show that compared with support vector machine (SVM) and fuzzy c-means (FCM) algorithms, our proposed fusion algorithm has better performance in clustering effect and convergence speed, and the error rate is low. Through the research of this project, more accurate data mining (DM) can be realized, thus improving the accuracy of the existing DM algorithm.

18.1 Introduction

It is necessary to study effective data classification algorithms, optimize the ability of data feature extraction, and achieve optimal scheduling and access to BD (big data). Online social networking, artificial intelligence (AI), e-commerce, and other industries have great demand for data mining (DM). As a new computer science and technology, DM has developed rapidly under the impetus of social demand [1]. In the AI era, it has become a necessary condition for e-commerce platforms to analyze the actual consumption potential of their own users to obtain high profits [2].

The main tasks of DM include classification analysis, cluster analysis, correlation analysis, and sequence pattern analysis, among which classification analysis has always been one of the hot spots in DM research because of its special position. The development history of classification algorithms can be traced back to the last century.

Y. Li (✉) · S. Song
School of Economics and Management, Shanghai Institute of Technology, Shanghai 200235, China
e-mail: lyf19990409@126.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_18

There are many kinds of algorithms, such as logistic regression, random forest (RF), support vector machine (SVM), decision tree, Bayesian, *K*-nearest neighbor (KNN), and neural networks which have developed rapidly in recent years, which can be used to solve classification problems [3–6]. In literature [7], the problem that each tree contains samples belonging to a common class caused by a large number of data and diversity in a large database may be unreasonable, and it is solved by introducing accuracy or classification closed value. In addition, the noise and isolated points in the training set will also affect the final classification results. Literature [8] expounds the application of automatic text classification technology to the preprocessing of corpus linguistics. In the subsequent domestic research and discussion, the main work focused on the research and improvement of some classical algorithms [9, 10].

The above method is to model the user data and then mine the data, which makes the anti-interference in the mining process more stable, but its practical application field is limited. The algorithm proposed in this paper maintains high accuracy in complex environment and has high applicability.

Intelligent innovation technology will lead the e-commerce industry to digital development, and AI technology will become an important power source for e-commerce platforms to acquire loyal users. In the process of BD classification and mining, its low accuracy is an important factor limiting its application. For this reason, in order to solve the problem that e-commerce needs to classify and process data quickly, this paper combines clustering theory, fuzzy logic theory, and artificial neural network (ANN) theory to design an e-commerce BD classification model based on AI algorithm. The algorithm proposed in this paper maintains high accuracy and high applicability in complex environment.

18.2 Research Method

18.2.1 Classification of DM

Commodity classification is a common problem in the development of electronic commerce. Classification is to extract a set of patterns that can describe the common characteristics of data from data, and then classify the data into attributes. When solving a classification problem, a single classification algorithm is usually used to construct a classification model composed of multiple categories of knowledge. Because of the popularity of the Internet and the development of online advertising, the research and application of e-commerce BD in China starts with advertising media, and the advertising delivery has also changed from the traditional non-targeted to personalized delivery. Based on the concerns of users, through BD analysis, accurate advertising can be achieved, and the efficiency of advertising can be improved while meeting the needs of users.

DM is a process of how to find valuable information in massive data. It first processes the massive data collected from the Internet and then deeply digs out

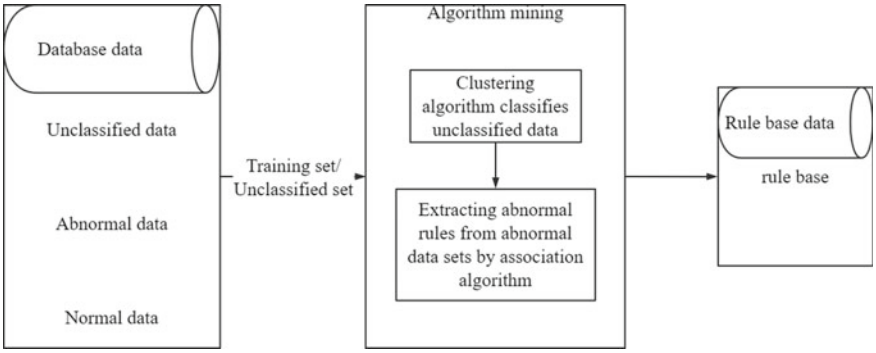


Fig. 18.1 DM process

the intrinsic value of the data through data analysis knowledge. Through this interdisciplinary comprehensive knowledge, an accurate data model is established, and corresponding algorithms are designed according to the characteristics of the data, so as to find out potential valuable information from the data and find out the relationship between things [11]. The schematic diagram of DM process is shown in Fig. 18.1.

The core step of DM is DM. We also mentioned the definition of DM above. The so-called DM is to obtain or mine valuable information from massive collected data. DM needs a lot of knowledge, which is interdisciplinary. It needs to integrate the knowledge of many disciplines, including but not limited to machine learning, data analysis, and overall planning, to get the results we want. The final result can help us make decision-making support results that are beneficial to ourselves, such as avoiding risks, looking for opportunities, and adjusting strategies in time.

Data can also be normalized, especially when learning by using neural networks or methods that include distance measurement. Normalization involves scaling a given attribute in order to reduce it to a smaller range. There are many classification algorithms for DM, such as inductive classification of decision tree, statistical Bayesian classification, neural network, KNN classification, genetic algorithm (GA), association-based classification, and fuzzy logic technology.

We can get the data on the Internet through crawler technology, and discover the regular potential value information from the disorderly data through DM, which can make users make decisions more conveniently and accurately. At present, businesses cannot do without data, and they all hope to get valuable information from massive data. Therefore, in this needed environment, more scholars will devote themselves to this field for research.

18.2.2 E-Commerce BD Classification

With the increasing popularity of Internet applications, people's basic necessities of life have undergone earth-shaking changes. For e-commerce, the unique convenience of the Internet has changed the traditional sales methods to a certain extent and promoted the development of the industry. Although the business scale of e-commerce platform built by commercial banks is increasing, its overall operation effect is still not ideal, e.g., it is not as good as the mainstream traditional e-commerce platform in terms of platform construction, platform functions, user experience, number of contracted merchants, number of active users, and user transaction volume. Enterprises can find out the pain points of customers' needs and use Internet information dissemination channels such as media and APP to launch accurate marketing, thus effectively improving marketing efficiency [12].

With the continuous improvement of intelligent computing methods in recent years, various intelligent computing methods are combined with ANN to form a hybrid intelligent system, thus giving full play to the advantages of various methods and making them occupy a place in DM classification calculation. Aiming at the problem that e-commerce needs to classify and process data quickly, this paper combines clustering theory, fuzzy logic theory, and ANN theory to design an e-commerce BD classification model based on AI algorithm. This method has high accuracy and practicability in practical application. Neural network is a group system composed of several simple neurons in some way. It uses nonlinear mapping theorem to express the implication relationship between input and output data in samples by learning weights.

In this paper, the strategy of e-commerce BD classification model is to use improved GA to preliminarily optimize the weights and closed values of neural network, and then use the improved algorithm to refine the preliminary results, so as to complete the training of neural network. The fuzzy weights are synthesized as simplified input data, and these simplified data are brought into the new neural network to continue learning, thus completing the whole learning process. The specific framework is shown in Fig. 18.2.

Although the emergence of BP algorithm makes up for the deficiency that it is difficult to determine the weight of neural network in practical application, the forward multilayer neural network with strong recognition ability can be applied. However, BP algorithm is essentially a gradient descent algorithm, so it inevitably has some defects, such as easy to fall into local minimum and slow training speed. For the problem that is easy to fall into local minimum, a parallel and multi-point search algorithm, GA, can be used to improve it.

GA has been used in the training of neural networks for some time. It can be used to learn the weights and topological structure of neural networks, and has achieved good results. However, there are still some problems. GA may have the problems of early convergence and gene deletion, so it cannot completely guarantee the shortening of training time and the global search. Differential evolution (DE) method has been proved to have a good optimization effect on real-valued and multi-mode objective

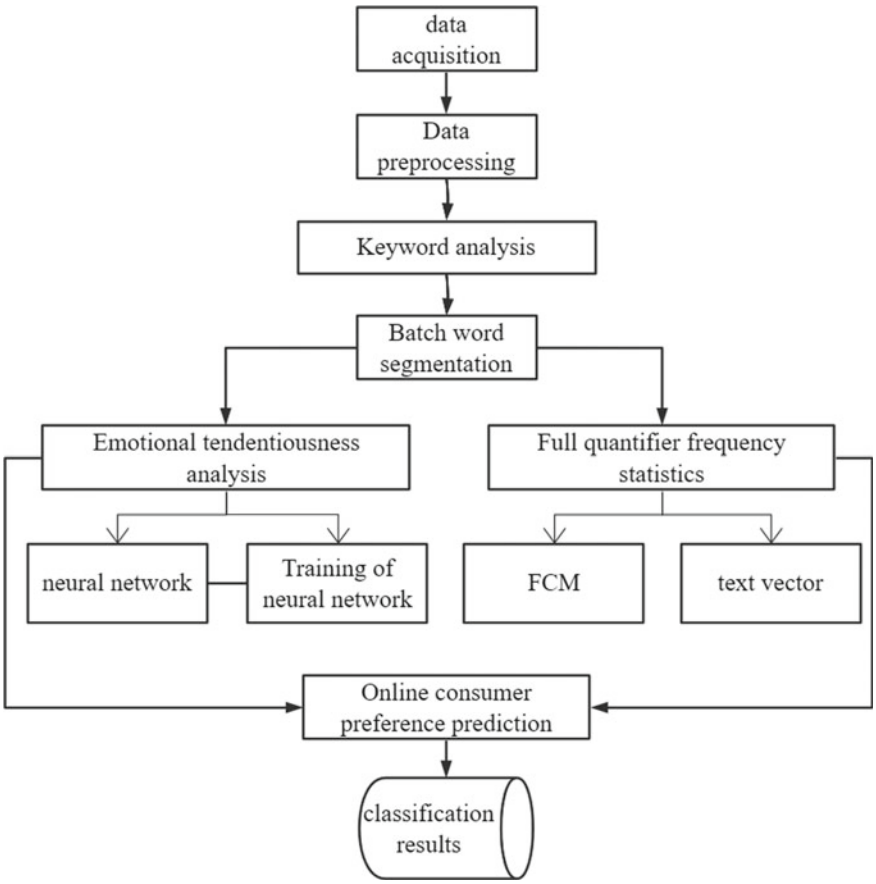


Fig. 18.2 Overall framework of e-commerce BD classification model

function optimization test problems. Besides its good convergence property, DE algorithm is very simple to understand and implement, with few control variables and remains unchanged throughout the optimization process.

In this paper, DE algorithm is used to train the weights of neural networks. Considering the deficiency of DE algorithm in local optimization, we introduce the Levenberg–Marquardt (LM) algorithm for local optimization, first use DE algorithm for global optimization to get a sub-ideal solution, and then use LM algorithm with stronger local optimization to further search and get an ideal solution, thus overcoming the shortcomings of DE local optimization and LM global poor.

In the G generation of a population, there are D-dimensional parameter vectors $X_{iG}(i = 1, 2, \cdots, N-1)$, which remain unchanged in the process of N optimization. The perturbation vector is generated by the following formula:

$$V_{i,G+1} = X_{r1,G} + F\left(X_{r2,G} - X_{r3,G}\right) \tag{18.1}$$

where $r1, r2, r3$ is an integer between $[0, N-1]$, which is different from each other and different from the current vector number i; F is a weighting coefficient, which controls the amplification of the difference between the two vectors of serial number $r2, r3$.

In order to make the disturbance vector more representative in the parameter space, the crossover process is introduced to get the next generation of the population:

$$U_{i,G+1} = \left(U_{0i,G+1}, U_{1i,G+1}, \cdots, U_{(D-l)i,G+1}\right) \tag{18.2}$$

where l represents the number of parameter exchanges, choose between $[1, D]$, and the size of l is controlled by the crossover probability. After the crossover operation, the objective function of the new vector $U_{i,G+1}$ and the predetermined vector $X_{i,G}$ is compared. If the new vector has a lower objective function, the predetermined vector is replaced by the new vector, otherwise, the predetermined vector is retained.

The purpose of clustering analysis is to divide the sample set into classes, which requires the internal compactness of classes, that is, the similarity of sample points in them is high and the similarity between classes is as sparse as possible, so as to maximize the difference between classes. Cluster analysis is essentially a classification problem of multidimensional data. When classifying samples, you do not need to consider parameters such as attributes and quantity, just determine the similarity between samples. Fuzzy c-means (FCM) has been widely used in pattern recognition, image processing, and deep learning in recent years.

The traditional FCM classifier limits the accuracy and anti-interference ability of classification, cannot effectively remove redundant information, and has poor convergence. Aiming at the above problems, this paper proposes an efficient BD classification algorithm based on FCM chaotic probability feature clustering and improves the algorithm.

According to the global search of chaotic DE algorithm, the optimal value of clustering center is found, and the optimal control input sequence and optimal objective function of large database coupling prediction are obtained. The behavior space $p(Q_s)$ of BD mining can be expressed as:

$$p(Q_s) = a_k \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left[-\frac{(Q_s - (Q_s))^2}{2\sigma_s^2}\right] \tag{18.3}$$

Among them, a_k is the mining behavior corresponding to each type at the convergence time of large database. If $a_k = 0$, it represents the joint probability density of chaotic feature transformation of k-type BD; if $a_k = 1$, it is assumed that the set of given points in the feature space of BD is $P = \{p_i | i = 1, 2, \cdots, n\}$, where the three-dimensional coordinate of p_i is (x_i, y_i, z_i) and n is the interval function of probability density of points.

Text vectorization refers to determining the dimension of feature space after feature selection, vectorizing each document according to the determined dimension of feature space, and weighting the feature words appearing in each document, so that each vector represents each document.

TF-IDF is a common word weighting algorithm and an evaluation algorithm to measure the importance of words in a text or text set. Its calculation formula is shown in Formula (18.4).

$$\text{TFIDF} = \text{TF} * \text{IDF} = \text{TF} * \log \frac{N}{n} \tag{18.4}$$

where TF is the word frequency, IDF is the anti-document frequency, N is the total number of training documents, and n is the number of documents with the feature word w. The normalized frequency of words appearing in a document is word frequency. The smaller the anti-document frequency IDF is with the number of documents in which words appear, the smaller the n value is, but the larger the IDF value is, which indicates that the word has strong ability to distinguish categories.

18.3 Simulation Analysis

In order to evaluate the performance of the AI-based e-commerce BD classification and mining algorithm proposed in this paper, MATLAB is used for simulation test, and SVM and FCM algorithms which are most commonly used in DM are used as a control.

The data used in this paper are all taken from the customers who have shopped on an e-commerce platform. Although the amount of information contained in this dataset is closely related to the issues to be discussed in this paper, through the detailed analysis of this dataset, we can see that there are many kinds in this dataset, while others are few. Using Python Web crawler technology, we can get the data of these small commodity categories on the mainstream e-commerce platform, which can expand the dataset and enhance the classification effect of these commodity categories.

In the experiment, the maximum number of iterations is set to 100, 5 subgroups, 3 subgroups, and 20 populations. The three methods were tested for 50 times, and the average results of each index were obtained. Three algorithms are tested, and the corresponding results are given (Table 18.1).

Table 18.1 Clustering experimental results

Algorithm	Total	Correct number	Accuracy/%
SVM	150	126	84
FCM	150	122	81
Our	150	141	94

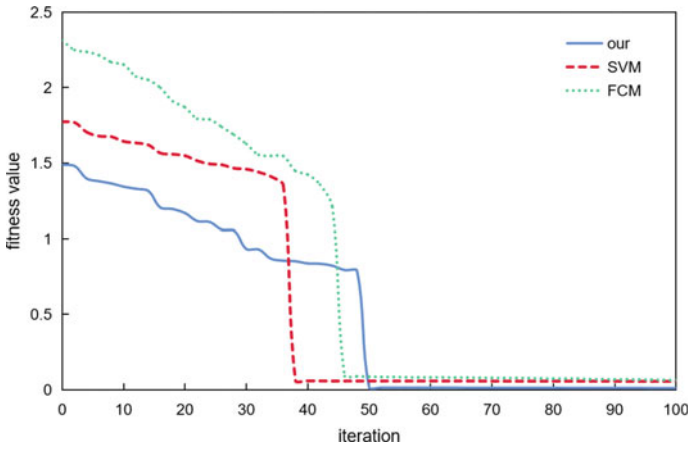


Fig. 18.3 Convergence rates obtained by different algorithms

Through the analysis of fuzzy control algorithm, it can be seen that by selecting the cluster center and fitness function in the improved FCM algorithm, the global optimization ability of FCM algorithm is improved, and the search accuracy of the algorithm is improved, thus achieving better clustering results.

Figure 18.3 is the comparison result of convergence rate obtained by applying SVM and FCM algorithm to dataset and this algorithm.

It can be seen that the convergence of SVM and FCM algorithm is poor; on this basis, the convergence rate of the proposed algorithm has been greatly improved. This method does not require high adjustment parameters, and has higher accuracy and faster convergence speed. Compared with other methods, this method has better clustering effect and faster convergence speed.

Figure 18.4 shows the comparison of the algorithm proposed in this paper with SVM and FCM algorithms in mining errors under different mining intensities.

Under the same mining intensity, compared with the other two methods, our method has a lower false alarm rate; in this paper, with the increase of depth, the rising speed of error rate slows down. The experimental results show that our method can achieve more accurate DM, thus improving the accuracy of existing DM methods to some extent.

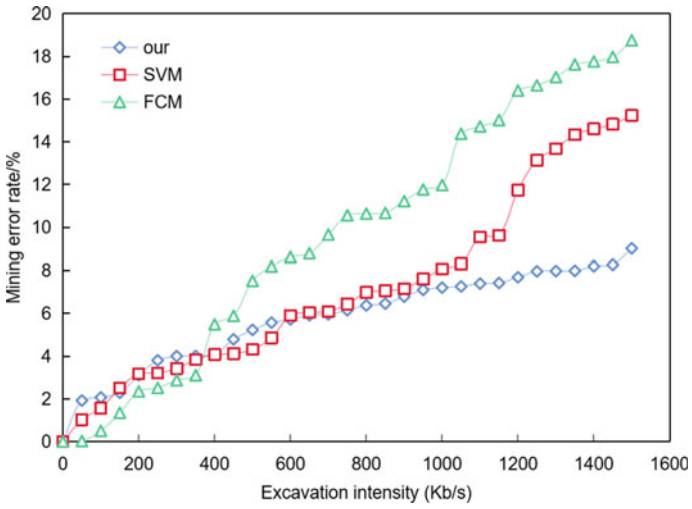


Fig. 18.4 Mining error test results

18.4 Conclusion

This project takes e-commerce big data as the research object, and studies an AI-based e-commerce BD classification model by using the theories of clustering, fuzzy logic, and neural network. Simulation results show that compared with SVM and FCM, the proposed algorithm has better clustering effect and faster convergence speed. Compared with SVM and FCM algorithms, the error rate of the proposed algorithm is lower. In this paper, with the increase of depth, the rising speed of error rate slows down. The experimental results show that this method can achieve more accurate DM, thus improving the accuracy of the existing DM methods. Because of the large amount of data processed in this paper, it consumes a lot of CPU resources and time. In order to speed up the running efficiency of the program, big data platforms such as Hadoop and Spark can be used for distributed processing of big data.

Acknowledgements This paper is supported by "Research on Computational Advertising Effectiveness Model Based on the Combination of All Media Communication Elements (21BXW082)".

References

1. Xing, W., Bei, Y.: Medical health big data classification based on KNN classification algorithm. IEEE Access **8**, 28808–28819 (2020)
2. Zhu, F., Chen, W., Yang, H., Li, T., Yang, T., Zhang, F.: A quick negative selection algorithm for one-class classification in big data era. Math. Probl. Eng. **2017**, 1–7 (2017)

3. Jiang, C., Li, Y.: Health big data classification using improved radial basis function neural network and nearest neighbor propagation algorithm. IEEE Access **7**, 176782–176789 (2019)
4. Li, H., Li, H., Wei, K.: Automatic fast double KNN classification algorithm based on ACC and hierarchical clustering for big data. Int. J. Commun. Syst. **31**(16), e3488 (2018)
5. Hu, L., Yang, S., Luo, X., Zhou, M.: An algorithm of inductively identifying clusters from attributed graphs. IEEE Trans. Big Data **8**(2), 523–534 (2022)
6. Majeed, A., Lv, J., Peng, T.: A framework for big data driven process analysis and optimization for additive manufacturing. Rapid Prototyping J. **25**(2), 308–321 (2019)
7. Satyala, N., Babu, C.S., Rao, M.S., Srinivas, K.: Ant cat swarm optimization-enabled deep recurrent neural network for big data classification based on map reduce framework. Comput. J. **65**(12), 3167–3180 (2022)
8. Sakthivel, D., Radha, B.: Adaptive model to detect anomaly and real-time attacks in cloud environment using data mining algorithm. Int. J. Perform. Eng. **10**, 17 (2021)
9. Hossain, M.A., Ferdousi, R., Hossain, S.A., Alhamid, M.F., Saddik, A.E.: A novel framework for recommending data mining algorithm in dynamic iot environment. IEEE Access **8**, 157333–157345 (2020)
10. Sun, H., Yao, Z., Miao, Q.: Design of macroeconomic growth prediction algorithm based on data mining. Mob. Inf. Syst. Article ID 2472373 (2021)
11. Wang, J., Liu, F., Jin, C.: PHUIMUS: a potential high utility itemsets mining algorithm based on stream data with uncertainty. Mathe. Prob. Eng. Article ID 8576829 (2017)
12. Pan, Q., Yang, G.: Application of mining algorithm in personalized Internet marketing strategy in massive data environment. J. Intell. Syst. **31**(1), 237–244 (2022)

Chapter 19
Research and Implementation of Data Feature Extraction Technology for Multisource Heterogeneous Data in Electric Distribution Network

Junfeng Qiao, Aihua Zhou, Lin Peng, Xiaofeng Shen, and Chenhong Huang

Abstract With the development of information technology, the degree of information of electric power enterprises is getting higher and higher, and the data generated by business systems is also growing rapidly. Data has become the new power of enterprise production and plays a vital role in the business growth of power companies. State Grid Corporation of China has brought the benefits of rapid business growth by using data, but at the same time, it is also faced with data redundancy, difficult data discovery, low efficiency, resource consumption, and other issues, as is the case in the power industry. In the power industry, the data preparation stage of the distribution network will also occupy most of the time and energy of data scientists. The high degree of specialization of data extraction and feature engineering work will result in a high threshold, which will increase obstacles for business personnel to participate in data value exploration. As long as traditional automatic data extraction and automatic feature engineering are based on business experience, data analysis and data extraction must be carried out from top to bottom, which is far from enough in the era of artificial intelligence. In order to reduce the technical threshold of data utilization, this paper fully integrates data science with power grid business, carries out research on multisource heterogeneous data extraction and automatic feature engineering technology, and realizes intelligent data management, improves data availability, improves data management level, and reduces the threshold of data utilization through automatic feature engineering methods, Finally, it supports typical business scenarios such as Internet of data management or power consumption information acquisition data management of State Grid Corporation of China.

J. Qiao (✉) · A. Zhou · L. Peng
State Grid Key Laboratory of Information & Network Security, State Grid Smart Grid Research Institute Co, LTD, Nanjing 210003, Jiangsu, China
e-mail: 1318558905@qq.com

X. Shen · C. Huang
Qingpu Power Supply Company, State Grid Shanghai Electric Power Company, Shanghai 201799, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_19

19.1 Introduction

Each department of the electric power enterprise is responsible for different businesses. The monitoring methods for equipment are also different, and the types and structures of monitoring are also diversified [1]. At the same time, the collection of electric power equipment information by each department is cross-related [2]. There is more or less duplicate data in different systems. How to share and integrate the data of each business system is also a challenge that we face [3].

This paper first studies the subject-oriented data management technology, investigates the current situation of multisource heterogeneous data extraction and access technology, and investigates the subject-oriented data management technology, such as master data management and data label system. The research supports the technical scheme of extracting data from the multisource heterogeneous data system connected to the data center and managing data by subject, so as to realize the subject-oriented source data management and modeling-oriented feature data management. It does a good job of data asset inventory before data enters the feature engineering platform, and deposits modeling knowledge and business knowledge after feature engineering processing.

Secondly, the standard feature engineering operator system will be realized according to the feature engineering algorithm of electric data. Then, combine the data characteristics of the electric power industry to realize automation of important processes such as data feature derivation and feature selection. Finally, in order to obtain tools that can be automatically linked with the upstream data storage system and downstream BI applications or AI models, research the service task scheduling technology.

Finally, the standardization, flow, and operator methods of feature engineering algorithms are studied. Investigate the research and application status of feature engineering automation technology at home and abroad, analyze the automation methods of key links such as feature derivation and feature selection based on the characteristics of power industry data, and study the feature engineering process automation technology for power data middle office and power industry data application and AI modeling application.

19.2 Related Work

Data feature engineering plays a very important role in machine learning, which generally includes three parts: feature construction, feature extraction, and feature selection [4]. Feature construction is troublesome and requires some experience. Both feature extraction and feature selection are aimed at finding the most effective features from the original features [5]. The difference between them is that feature extraction emphasizes obtaining a group of features with obvious physical or statistical significance through feature transformation; feature selection is to select a set of

feature subsets with obvious physical or statistical significance from the data feature set.

Data feature construction refers to the manual construction of new features from the original data. It takes time to observe the original data and think about the potential form and data structure of the problem [6]. Data sensitivity and machine learning experience can help feature construction, which requires the analysis process to find some features with physical significance from the original data. Assuming that the original data is table data, you can generally use mixed attributes or combined attributes to create new features, or decompose or segment the original features to create new features [7].

Data feature extraction generally refers to the extraction of original data before feature selection. Its purpose is to automatically build new features and convert the original data into a group of features with obvious physical significance (such as Gabon, geometric features, and texture features) or statistical significance [8]. Common methods include domain independent feature extraction methods (PCA, ICA, LDA, auto-encoder, etc.), domain-related feature extraction methods (SIFT, Gabon, HOG, etc., in image, word bag model in text, word embedding model, etc.) [9]. Principal component analysis (PCA) is the most classical method for dimension reduction. It aims to find the principal components in the data and use them to characterize the original data, so as to achieve the goal of dimension reduction. Through PCA, features with small variance can be discarded [10]. Here, the feature vector can be understood as the direction of the new coordinate axis in coordinate transformation. The eigenvalue represents the variance on the corresponding feature vector. The larger the eigenvalue, the greater the variance, and the greater the amount of information. This is why the eigenvectors corresponding to the first n largest eigenvalues are selected, because these features contain more important information.

19.3 Research on Data Feature Extraction Technology for Multisource Heterogeneous Data in Electric Distribution Network

19.3.1 Fusion of Electric Power Multisource Heterogeneous Data

In this paper, we design a data extraction interface for multiple data systems connected to the data center. Considering universality, three interfaces are implemented.

The relational database data is automatically extracted. The specified table in the specified database is extracted through JDBC connection, or the specified field of the specified table is selected through user-defined SQL; achieve the goal of importing data from relational databases to feature platforms. The automatic extraction of data in the distributed data storage system realizes the pulling of distributed data files and achieves the goal of importing data from the distributed outgoing system to the

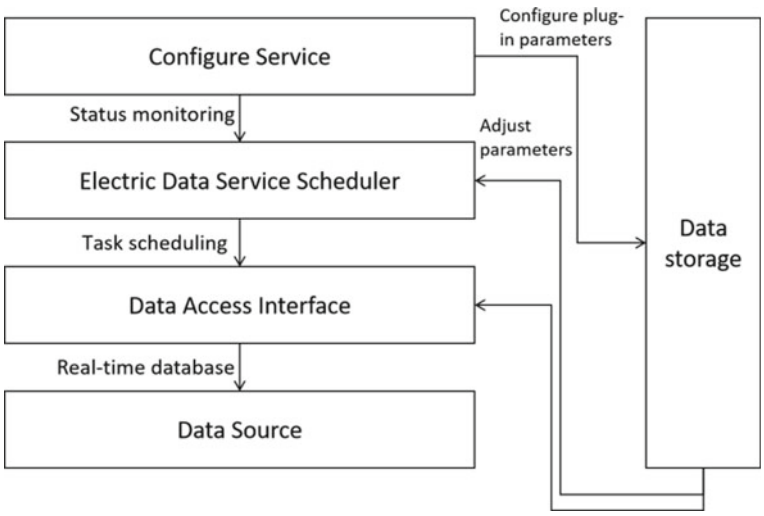


Fig. 19.1 Multisource heterogeneous electric data access process

feature platform. Unstructured data is automatically extracted, and fuzzy matching method is used to match the fields on the line and extract line by line.

As is shown in Fig. 19.1, the process of multisource heterogeneous data access includes extracting data from the data source, accessing the data to the data call interface through the data extraction tool, configuring the scheduler according to the scheduling task, and finally monitoring the status of the data scheduling task, controlling the task process, and realizing the configuration service.

(1) Data access consists of five parts: configuration service, access scheduler, access plug-in, data storage, and data source. The core is configuration service, configuration scheduler, and access socket.
(2) The configuration service includes three sub-modules: access status monitoring, access start stop control, and access configuration. The access configuration module is used to configure the access task scheduling parameters, data analysis rules, measuring points, and measuring point value creation rules and other parameters; The access start stop control module controls the start and stop of access tasks by sending commands to the access scheduler; access status monitoring is used to monitor the operation status of the access scheduler.
(3) The access scheduler is responsible for loading tasks and scheduling and executing access tasks according to the task scheduling parameters. The task scheduler starts to periodically check the task scheduling parameters. When it finds that a task meets the execution conditions, it immediately starts the process and executes the task association plug-in.
(4) The access plug-in accesses the data to be accessed according to the configured data access parameters, parses the data according to the data analysis rules, creates the measuring points on the platform according to the measuring points

and measuring point value creation rules, and parses the data into measuring point values and stores them in the corresponding measuring points.

When completing data access and data processing, both structured and unstructured data use different processors. This requires selecting the corresponding processor according to the format of the data source, and specifying a time point. The time can be set to dynamic time such as hourly and daily. After data processing, store the data to the destination. Then, conduct data processing on the front-end data source acquisition, perform regular data cleaning, progress check, error recovery, and exception recording on the data processor, and store the data in the corresponding target database.

19.3.2 Data Feature Extraction Technology for Electric Industry Data

According to the analysis of load characteristics, it is necessary to extract the power and daily frozen electricity data of stations and users in the power consumption information acquisition system; account data of the marketing system; meteorological data, etc., and integrate relevant data. An example of the extracted station area and user related data is given in Table 19.1.

After data processing, the feature vectors are extracted according to the demand of load analysis. This vector is a generalized vector, which can be a one-dimensional vector or a multidimensional vector. This time, it is set as a one-dimensional vector, and the feature vector format is set according to the planning data model standard. The fusion rules are as follows: when there is a consistent description of an attribute in two vectors, the fused vector takes one of the data; when there is an inconsistent

Table 19.1 Substation area data of electric consumption information acquisition system

Number	Attribute	Field content
1	Internal logical number	7,400,000,161,518
2	Assessment unit no	0,690,100,034,721
3	Assessment unit name	Address information of the assessed user
4	Substation name	Details of middle voltage and high voltage substations
5	Electric line name	The specific name of electric line is in the equipment management system
6	Station area no	0,690,100,034,721
7	Electric customer no	8,104,458,164

description of an attribute in two vectors, it shall be selected according to the data access priority criterion; when a vector has an attribute description that another vector does not have, the fused vector needs to retain this data. The fusion of three data can be seen as the result of two data fusion and then the third data fusion. Similarly, it can be extended to the fusion of multiple data.

The above three data tables are fused, and each oil cartographic data is used as the primary key value to supplement the manufacturer and number of the sensor that collects this cartographic data, as well as the manufacturer, connection mode, power supply company, commissioning date, voltage level, and other information of the transformer installed with this sensor. Thus, a data information base in a unified format can be formed, which contains the information of all sensors and transformers related to oil cartographic data. The data with unified format can be read and processed by using the data processing program and setting the corresponding call command as required.

19.4 Implementation of Data Feature Extraction Technology for Multisource Heterogeneous Data in Electric Distribution Network

Electric data feature extraction technology is closely related to specific data applications or modeling tasks. In the traditional machine learning methodology, the introduction of business knowledge can better perform feature extraction. Although violent dimension upgrading does not rely on business knowledge, the cost of computing resources consumed is high. In order to deal with the multidimensional data of power big data across disciplines, this paper conducts research from three aspects.

The task of data feature extraction is often aimed at specific power data applications or modeling tasks, e.g., a modeling task involves the input of 160 original fields. It needs to filter out the original fields closest to the task and process them into features through a series of data process, feature transformation, feature derivation, and feature selection. Considering the specificity of the modeling task, it is difficult to provide a universal and fully automatic feature processing logic. Therefore, this chapter provides rich feature operators and implements graphical interface components. Each component implements a specific feature engineering action to ensure the certainty of the action execution results. The user dominates the workflow editing. However, it also provides fully automated feature derivation and feature selection algorithms, provides reference example workflows for typical scenarios, and realizes the generation of feature engineering workflows in a semi-automatic manner.

19.4.1 Access and Governance of Electric Data Resources

Before data extraction, the cleaning tools are used to filter and clean the collected original data, and then, logical operations are carried out according to the calculation rules of different data. Finally, they are classified and stored in the corresponding data worksheet to obtain high-quality source data, which is conducive to solving the problem of data storage redundancy and reducing the data storage capacity.

Data cleaning mainly includes correcting errors, deleting duplicates, unifying specifications, correcting logic, transforming structures, data compression, supplementing incomplete/null values, and discarding data/variables. Among them, the most important is the data transformation and construction.

Data transformation is an important step in the data cleaning process and a standard processing of data. Almost all data processing processes involve this step. Common contents of data conversion include: data type conversion, data semantic conversion, data range conversion, data granularity conversion, table/data splitting, column conversion, data discretion, data discretion, refining new fields, attribute construction, data compression, etc. The input data-frame is duplicated according to the specified column. Example parameters are given in Table 19.2.

Create a data-frame that contains only rows that meet the given criteria. Conditions are used in WHERE conditions in Spark SQL. The column order of the output data-frame will be retained, which is the same as the column order of the input data-frame. A transformer is also returned, which can be applied to another data-frame later using the transform operation.

To remove the extreme value of the column, the conventional processing concept is to determine the upper and lower threshold of this data indication, and then set the data that exceeds (or is below) this limit as the threshold to improve the accuracy of the data conclusion. Missing value processing: fill in missing values in batches. The filled columns must be numerical types. Note: the mean and median only support double types. A transformer is also returned, which can be applied to another data-frame later using the transform operation.

As given in Table 19.3, it specifies the maximum number of different documents where a word may appear in the vocabulary which terms that exceed the threshold will be ignored. If it is an integer greater than or equal to 1, specify the maximum number of documents that may appear in the word, and specify the maximum proportion of documents that may appear in the word. Filter to ignore rare words in the document. For each document, words whose frequency/count is less than the given threshold are ignored. If it is an integer greater than or equal to 1, it specifies a count of the

Table 19.2 Results of electric data process

Parameter	Description
Select rename column	Select “device” by name
New column name	Select “device_type” by name
Report type	Extended report, generally select metadata report to save time

Table 19.3 Electric text data process result

Item	Description
Column	Select input columns
Column name	Columns generated after model transformation
Number of characteristic	The number of features must be an integer
Binarization	If it is true, all nonzero counts are set to 1, which is useful for discrete probability models
Advanced parameter options	Override previously set options
Parameter	English name of spark algorithm parameter
Value	Set the value of the spark algorithm parameter

number of times the word must appear in the document. If this is double precision, it specifies the proportion of the word count in the document.

19.4.2 Implementation of Feature Extraction Technology for Electric Data

The first step of feature extraction is to process the data vector, extract a single vector value, extract a single index value from the vector column, the number is specified by the index parameter, and the output column is of double type. This operation also returns a transformer, which can be applied to another data-frame later using the transform operation. Then, mark the vector size and add metadata of the vector size for the vector column with missing metadata. The vector type in the vector combination needs to provide metadata of the vector. The vector metadata size result is not displayed in the input and output, but added as a schema to mark the vector metadata.

Secondly, multiple columns or stored vector columns are merged into a vector column in order. The vector composition component adds the values of the input columns to a vector in the order of the columns in the data-frame and finally extracts the elements of the vector to form a new vector.

In the date processing step, first decompose the date and time, and extract the part of the timestamp from the specified timestamp column to a new column, e.g., when extracting the year part, given the timestamp column of birth-date, a new numeric column named year will be created. Then, the date and time are combined, and the numerical fields such as year, month, and day are combined into timestamp columns, e.g., when the composite column birth_ year, birth_ Month, and birth_ Date, a birthday date with timestamp type will be created.

Infer the type of data characteristics, and infer the type of selected column data. Note: This component only re-infers the type of string data columns, and then converts array columns to vector, double arrays to vector, and column types in data-frame. One or more columns can be converted at the same time. Each selected column

Table 19.4 Data feature extraction operator analysis

Item	Input qualifier	Description
Multi-column feature	Multiple columns selector	Select multidimensional feature columns to be standardized
Normalized/standardized type	Choice	Set normalization and standardization types
Zoom	Numeric	Scale amount and floating point number
Advanced parameters	Advanced parameter options	Advanced parameters will overwrite the selected parameters
Report type	Report type	The optional values are metadata report and extended report

will be converted to the same (selected) target type. It supports conversion to string, Boolean, timestamp, double, float, long, int, and other types.

As is shown in Table 19.4, the maximum absolute value standardization results are listed. Divide by the maximum absolute value of each feature to scale each feature to the specified range. It does not move/center data, so it does not break any sparsity. Use robust statistics to scale data with outliers (outliers). The processing method is to delete the median and scale the data according to the percentile range (the default value is 1: inter-quartile spacing), which is the range between the first quartile (25%) and the third quartile (75%).

19.5 Application Results

This paper has carried out three aspects of research, namely the topic-oriented feature data management technology, the automatic realization of feature engineering processing flow, and the technical realization of feature engineering based on distributed architecture. It has realized the research and development of the prototype of the data feature engineering automatic processing system, and has carried out pilot applications in Jiangsu consumer system and production management system based on the prototype system.

In the aspect of "subject-oriented feature data management technology", a data extraction interface suitable for multiple types is proposed, which can be the widest data entry; in addition, a multimodal scheme using image tag pair data to infer the set of synonymous tags is also proposed. First, an unsupervised method is used to obtain the initial set of synonymous tags, and then a small part of the data is manually checked and corrected. Finally, a supervised method is used to improve the result of synonymous tag set and manually correct it. This process can greatly reduce the workload of manual annotation.

In the aspect of "automatic realization of feature engineering processing flow", a set of rich feature operators are refined to realize the automation of processing

flow, including general operators and special feature operators of power industry; An automatic feature engineering method for power time series data set is proposed, which can reduce the workload of power data processing personnel and reduce the prior knowledge requirements for feature construction to a certain extent; a feature extraction algorithm for multimodal data is also proposed to avoid the huge resource consumption caused by exhausting all feature transformations and save more than 50% of the data extraction time. Use drag and drop to connect different operators into business workflows, and process and analyze data. They cooperate with each other and combine the algorithm and computing layer closely, which greatly improves the computing efficiency of distributed work scenarios. In the future, the research method proposed in this article will be promoted and applied to power data quality improvement projects and power customer data service businesses.

Acknowledgements This work was supported by State Grid Corporation of China's Science and Technology Project (5400-202258431A-2-0-ZN) which is "Research on deep data fusion and resource sharing technology of new distribution network".

References

1. Hu, D., Huang, Y.: Research on feature extraction method of electric power data based on integrated learning. Think Tank Times **000**(024), 261–261267 (2020)
2. Yang, J., Deng, M., Ma, Y., et al.: PRPD data feature extraction method based on deep learning. Electr. Measur. Instr. **57**9(728)(03), 104–109+120 (2020)
3. Yang, Y., Li, H.: Automatic extraction method of power line based on airborne LiDAR point cloud data. Progr. Laser Optoelectr. **57**(9), 6 (2020)
4. Li, Z., Liu, H., Bi, T., et al.: Robust power system disturbance identification based on data drive. Chin. J. Electr. Eng. **41**(21), 14 (2021)
5. Olson, R.S., Bartley, N., Urbanowicz, R.J., et al.: Evaluation of a tree-based pipeline optimization tool for automating data science. Proc. Genetic Evol. Comput. Conf. **2016**, 485–492 (2016)
6. Udayan, K., Fatemeh, N., Horst, S., Elias, K., Deepak, T.: Automating feature engineering. Transformation **10**(10), 10 (2016)
7. Cheng, Y., Jiang, M., Ma, Y.: Design of real-time monitoring of power customer flow data based on metadata. Inf Technol **44**(4), 5 (2020)
8. Wang, L.: Load forecasting of regional power grid based on deep learning neural network. Shandong University (2020)
9. Zhang, Q., Song, Q., Zhang, J., et al: A new method of power economic feature extraction based on extended panel big data. Power Grid Clean Energy (2021)
10. Sun, Y., Zhang, F., Chen, S., et al.: Anomaly detection method of power data based on multi domain feature extraction. J. Electr. Power Syst. Autom. **006**, 034 (2022)

Chapter 20
Design and Optimization of Business Decision Support System Based on Deep Learning

Yiyun Li

Abstract Decision support system (DSS) is based on the data of daily business processing system, using mathematical or intelligent methods to analyze the data and comprehensively predict the future business trends. The deep learning model based on neural network architecture can effectively process and analyze big data. This paper constructs a business DSS based on deep learning and optimizes it. In the middle platform, AI and deep learning technology are used to dynamically model and predict commercial retail. A convolutional neural network (CNN) model based on network performance evaluation is proposed, which effectively improves the dynamic performance of CNN and improves its prediction ability. The system test results show that the integration of CNN prediction function into traditional DSS not only maintains the characteristics of traditional DSS, but also has the characteristics of neural network, makes full use of normative knowledge and empirical knowledge, and improves the intelligent level of traditional DSS. Help enterprises achieve their business goals more steadily, improve work efficiency, and reduce labor costs.

20.1 Introduction

With the increasing complexity of modern society, there are more and more uncontrollable factors in decision analysis, which increases the difficulty of decision-making. Corresponding decision-making has also developed from individual decision-making to think tank, collectivization, and group decision-making, and a large number of decision-making consulting institutions have emerged. With the continuous development of commercial marketing, the functions and business of commercial information system are becoming more and more complex, the tasks undertaken by the system background are increasing, and the system load is becoming heavier and heavier, which affects the operation efficiency, system maintenance, and

Y. Li (✉)
School of Economics and Management, Shanghai Institute of Technology, Shanghai 200235, China
e-mail: 1536477721@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_20

function expansion of the marketing system [1, 2]. Decision support system (DSS) is based on the data of daily business processing system, using mathematical or intelligent methods to analyze the data and comprehensively predict the future business trends. At present, foreign DSS technology has entered the application stage, especially in the investment and finance, sales and production, forecasting and planning, and budget analysis of large, medium and small enterprises, and has achieved remarkable results [3, 4].

Using DSS, we can decide whether to buy or sell this product by comparing historical data at the first time. Therefore, it is necessary to design and develop targeted DSS to support correct, timely, and effective decision-making in combination with the practical operation experience of middle and senior managers for many years [5]. Around business intelligence, relevant research and development have been carried out at home and abroad, mainly in data mining, data warehouse, and other technical aspects. Literature [6] puts forward scientific investment decision through the analysis of business information mining. Literature [7] holds that AI is applied in the retail industry, providing personalized recommendation service for customers and improving customer satisfaction and consumption willingness. In the early days, each DSS usually had to extract data from multiple data platforms, which formed a spider web structure over time, which caused extreme confusion for enterprise data management and maintenance. At the same time, DSS also encountered problems such as data fragmentation and data inconsistency, and the difference between transaction processing system and analytical processing system led to the emergence of data warehouse technology [8].

The business flow plan is a target plan made by various indicators in the purchase, sale, deployment, and storage of goods by large commercial enterprises, and the time span is generally one year. Business flow planning decision-making is a typical semi-structural decision-making problem, which requires both qualitative analysis and quantitative calculation of the influence degree of each factor [9]. The deep learning model based on neural network architecture can effectively process and analyze big data. Compared with traditional methods, it has obvious advantages in terms of image, voice and text recognition, reasoning, analysis, and judgment [10]. This paper constructs a business DSS based on deep learning and optimizes it. In the middle platform, AI and deep learning technology are used to dynamically model and predict commercial retail. A convolutional neural network (CNN) model based on network performance evaluation is proposed, which effectively improves the dynamic performance of CNN, optimizes the model structure, and improves the prediction ability.

20.2 Research Method

20.2.1 Structural Design of Commercial DSS

DSS is a man–machine interaction system which uses modern information technology as a means, comprehensively uses computer technology, management science, AI technology, and other scientific knowledge, and provides help for managers to make correct decisions by providing background materials, helping to determine the problem, modifying and perfecting the model, and enumerating possible ways. Traditional DSS often encounters two problems in decision-making: first, the decision-making information is too large, so it is difficult to find the internal relationship between the data from the huge data, and there is no clear mathematical analytical formula for calculation, which is a complex nonlinear problem; second, decision-making information is incomplete and some information is unclear, so decision-making technologies such as expert systems and traditional statistical methods can do nothing about it.

How to make qualitative analysis and quantitative calculation of the factors that affect the business flow plan, so as to help decision-makers grasp the range of market changes as much as possible, and thus predict the total sales level in the next year. From the practical point of view, the system should simulate the decision-making process as accurately as possible, and the established model should also describe the actual decision-making process as much as possible, in order to support the decision-makers more effectively. In addition, the design of the system should develop in the direction of DSS generator and fully reflect the characteristics based on knowledge.

DSS is composed of data warehouse and decision-making reasoning. Data warehouse provides a unified data source for it, and decision-making reasoning collects all relevant data and information, which is processed and arranged to provide information for the decision-making management of enterprises and provide a basis for decision-makers [11]. It enables analysis and decision-makers or managers to quickly, consistently, and interactively access the information transformed from massive raw data, which can be truly understood by users and truly reflect the dimensional characteristics of enterprises, so as to gain a deeper understanding of the data.

The bottleneck of this system performance is mainly reflected in two aspects, one is the bottleneck of customer access ability, which is the concurrent ability of the system, and the other is whether the server can handle the data in time when the data server records data in millions. As the market can provide more and more cheap high-performance processors, memory, and hard drives, or you can improve your performance by purchasing higher-performance hardware devices. The software architecture of commercial DSS is shown in Fig. 20.1.

The operation process of DSS is that the user inputs the problem to be solved through the man–machine interface, and the man–machine interface transmits the input information to the problem solving system to collect data, find the model to solve the problem, get the result through calculation and analysis, and feed back the

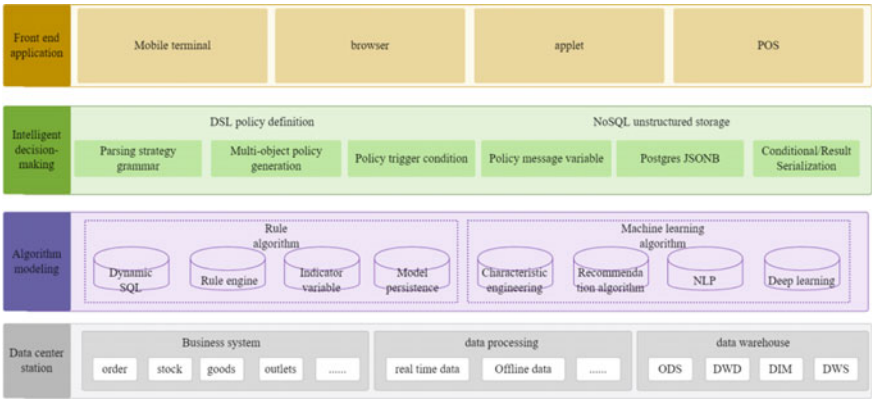


Fig. 20.1 Software architecture of commercial DSS

decision information to the user. If there is an accident, it will be solved through dialog with the user through the man–machine interface. The system developer establishes and maintains the model through the model base system, and the database system is responsible for collecting and providing data.

In the middle platform system, it is necessary to realize the management of commodity information, prediction analysis based on deep learning, and auxiliary function interface design. Use MySQL database technology to manage commodity information data, transaction order data, shopping mall employee information, customer information, and data used for sales forecast.

20.2.2 System Optimization

At present, the dispersion and fragmentation of information and the irrationality of the old system have caused great troubles to the decision-makers of computer sales enterprises in China, wasted many resources, and lost many opportunities, making computer sales enterprises in an unfavorable position in the increasingly fierce market competition. Due to the nonstandard data, errors often occur in batch processing, which takes a lot of time to maintain. The development efficiency is low. There is a lack of tools in data processing, analysis, and presentation, and traditional programming methods are often adopted, and it is difficult to achieve modular development. This approach does not fully meet the various needs of visitors, basically use text, no graphical interface, can not give the user intuitive, perceptual awareness and understanding.

Enterprises with different scope and content of business flow plan may have different requirements, but generally it should include the quantity and amount of goods purchased, sold, transferred and stored, gross profit and profit, and various

expenses. From the practical point of view, the system should simulate the decision-making process as accurately as possible, and the established model should also describe the actual decision-making process as much as possible, in order to support the decision-makers more effectively. The reasoning of knowledge adopts two kinds of reasoning strategies: forward reasoning and backward reasoning. The reasoning of professional knowledge adopts forward reasoning and the reasoning of system knowledge adopts backward reasoning, which is convenient and rapid. However, the positive reasoning strategy for professional knowledge is close to the actual situation.

Business decision-making is the most complex and challenging activity in the business field, which promotes the emergence and development of business-aided decision support technology. It can be predicted that the continuous progress of deep learning technology will have a profound and significant impact on the field of business intelligence-aided decision-making [12]. Deep learning is developed from ANN (Artificial Neural Network). Deep learning is supported by data and computing power, by building a neural network with multiple hidden layers to extract features from external input data and obtain the required information. The development of deep learning has achieved impressive results to this day, thanks to the in-depth development of deep learning models such as deep belief networks (DBN) and convolutional neural network (CNN).

Inspired by the mechanism of visual nerve, CNN is a multilayer perceptron designed to recognize two-dimensional or three-dimensional signals. This network structure is highly invariant to translation, scaling, tilt, and other deformations. The basic structure of CNN includes two special neuron layers, one is convolution layer, and the input of each neuron is connected with the local part of the previous layer, and the characteristics of the local part are extracted; the second is pool layer, which is used to calculate local sensitivity and secondary feature extraction. This twice feature extraction structure reduces the feature resolution and the number of parameters to be optimized. The parameters of CNN feature extraction layer are learned from training data, so it avoids manual feature extraction and learns from training data. Secondly, the neurons in the same feature graph share the weights, which reduces the network parameters, which is also a great advantage of convolutional networks over fully connected networks.

Commercial DSS is a complex nonlinear dynamic system, and the business logic and signal correlation within the business are very complicated. In order to improve the comprehensive efficiency of business marketing and establish an effective business DSS model, this paper designs a network model CNN based on neuron performance evaluation and screening strategy, which mainly includes two parts: dynamic evaluation of neurons and convolution performance screening, so as to realize the comprehensive evaluation of CNN's neuron performance.

Dynamic evaluation of neurons;

$$Q = \frac{\sum_{i=1}^{n} \left(w_i - \frac{\sum_{j=1}^{n} w_j}{n} \right)^2}{n} \tag{20.1}$$

n is the input vector and w is the synaptic weight of neurons.

In the convolution layer, the feature map of the previous layer interacts with the convolution kernel to form the output feature map j of the convolution layer, and each output feature map j may contain convolution with multiple input feature maps. Generally speaking, we have:

$$z_j^{(l)} = \left(\sum_{i \in M_j} a_i^{(l-1)} \otimes k_{ij}^{(l)} \right) + b_j^{(l)} \tag{20.2}$$

$$a_j^{(l)} = f\left(z_j^{(l)}\right) \tag{20.3}$$

Here, M_j represents a selection set of input feature maps. You can select all the input features, some of them, or a specific combination by learning parameters. w, b represent the corresponding weights and biases, respectively. $a_i^{(l)}$ is the activation value of the i neuron in the l layer, and $b_j^{(l)}$ is the bias term of the j neuron in the l layer. $z_j^{(l)}$ is the weighted sum of the input of the j neuron in the l layer.

The overall architecture of the network is shown in Fig. 20.2, and its main structure is composed of multilayer CNN and self-attention unit. In order to adapt to the problems of different scales, multilayer structures with different complexity are designed. Each layer is composed of CNN, but the network structure of each layer is different (network layer and number of neurons), and its complexity decreases in turn. The simple layer is responsible for the problem of small-scale data, and the complex layer is responsible for the problem of large-scale data.

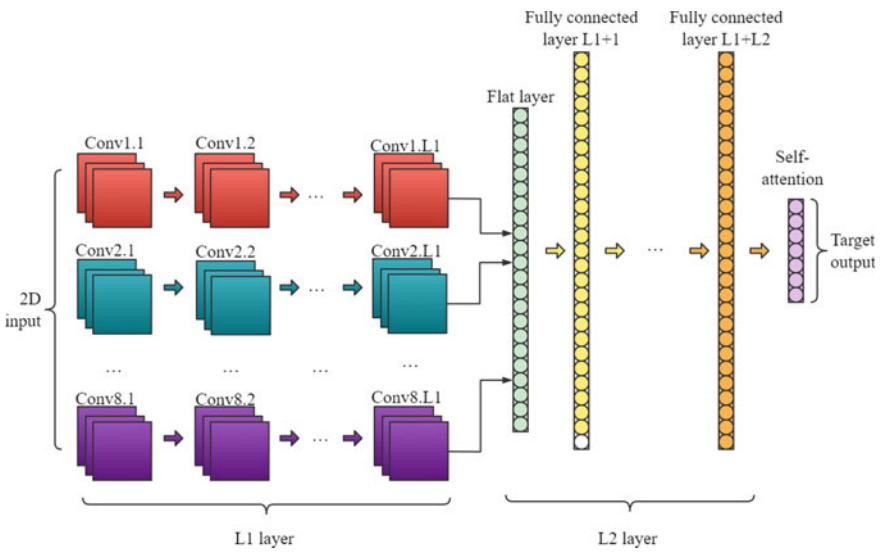


Fig. 20.2 Network model

Self-attention mechanism is introduced into the model to correlate and process the outputs of multiple layers, as given in Formula (20.4). It generates a matching weight Attention for each layer to measure the importance of each layer.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{20.4}$$

$Attention(Q, K, V)$ represents the importance of each layer, Q, K, V, d_k represents the matching weight, and QK^T represents the association and processing.

Because of the large scale of enterprises and many unpredictable factors, enterprises used to make business flow plans based on the inertia of the market, which caused a great deviation from the actual situation and affected the operation of enterprises. Call the system environment through the coordinator to deal with the problems raised by users. Among them, pattern knowledge is the abstraction and regularization of specific decision-making patterns [13].

The reasoning of knowledge adopts two kinds of reasoning strategies: forward reasoning and backward reasoning. The reasoning of professional knowledge adopts forward reasoning and the reasoning of system knowledge adopts backward reasoning, which is convenient and rapid. However, the positive reasoning strategy for professional knowledge is close to the actual situation.

20.3 System Test

This system is now applied to a large domestic enterprise and runs well. The data management subsystem in the system completely realizes the automatic collection function of distributed and heterogeneous data; the application service, acquisition task, and configuration parameters of the remote start-stop acquisition terminal are realized. The system automatically gives an alarm when it encounters an abnormality, and the delay from the fault to the alarm is less than 1 min; for the hardware environment, the CPU utilization rate is not more than 60% during real-time monitoring, playing an increasingly important role.

The interaction between terminal POS and this system mainly tests whether the commodity information query interface provided by this system is normal. In this part of the test, the system is deployed on Tencent Cloud server, and finally, the data interaction test is carried out with POS cashier terminal deployed on Android platform. The test case table of data interaction test is given in Table 20.1.

Table 20.1 Terminal POS interactive test case

Test specification	Operating procedure	Expected result	Testing result
Statistical analysis test	Click the statistical analysis link	Display data analysis	Pass

Discover market rules from historical data, predict future business development trends, predict and monitor risks, assist decision-makers to discover new profit growth points, optimize enterprise resources and management norms, help enterprises achieve their business objectives more steadily, and improve work efficiency and reduce labor costs.

In order to illustrate the feasibility and effectiveness of the method, this paper takes the situation of social consumer goods in China from January to December 2022 as an example to predict and verify the retail situation of social consumer goods. It is necessary to choose the correct evaluation index to make an accurate forecast. According to the characteristics of consumer goods market, four factors, such as consumer index, national retail price index, industrial ex-factory price index, and money supply, are selected as the main factors affecting the total retail sales of social consumer goods, and these factors are digitally quantified and normalized into values in the range of [0, 1].

After the network training, the data of another six months are selected as input data for prediction. Table 20.2 presents the prediction results, and Fig. 20.3 shows the relative errors.

The traditional DSS model cannot learn by itself, but can only reason according to the existing data, which requires high accuracy of data, especially when a new model appears, the information provided to decision-makers will be scarce. After a period of training, CNN can identify patterns without human intervention, solve problems in specific fields, and provide comprehensive information to decision-makers when the data is incomplete.

CNN provides an effective solution for the operation mechanism of some complex objects. It can learn the mapping relationship of a large number of patterns and avoid the artificial factors in determining the membership function in fuzzy evaluation. Based on the model, the test is carried out with data, and the output result is ideal.

In this paper, the prediction function of CNN is integrated into the traditional DSS, which not only keeps the characteristics of traditional DSS, but also has the characteristics of neural network, makes full use of normative knowledge and empirical knowledge, improves the intelligent level of traditional DSS, and achieves good results in practical application.

Table 20.2 Prediction result

Month	Actual (100 million yuan)	Forecast (100 million yuan)
January	7652.491	7421.212
February	7605.99	7550.945
March	7445.402	7415.432
April	7385.818	7148.35
May	6938.219	7104.133
June	7256.25	7168.184

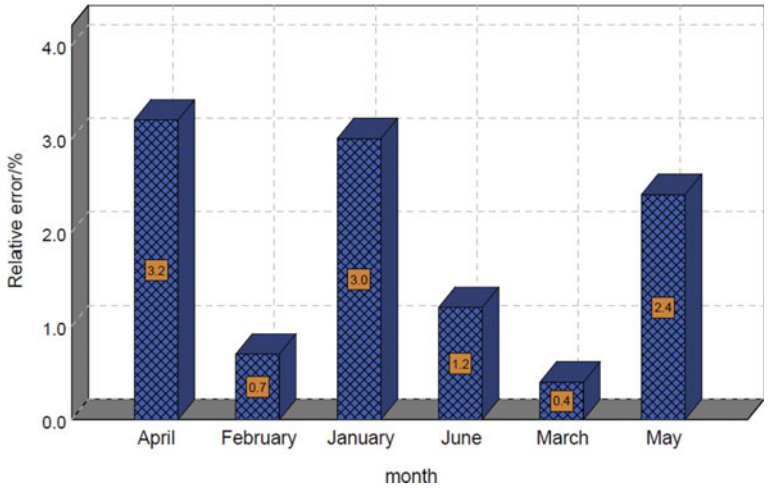


Fig. 20.3 Relative error

20.4 Conclusion

This paper constructs a business DSS based on deep learning and optimizes it. In the middle platform, AI and deep learning technology are used to dynamically model and predict commercial retail. A CNN model based on network performance evaluation is proposed, which effectively improves the dynamic performance of CNN, optimizes the model structure, and improves the prediction ability. The system test results show that the system established in this paper can discover the market rules from historical data, predict the future development trend of business, predict and monitor risks, assist decision-makers to find new profit growth points, optimize enterprise resources and management norms, help enterprises achieve their business goals more steadily, and improve work efficiency and reduce labor costs. Integrating CNN prediction function into traditional DSS not only keeps the characteristics of traditional DSS, but also has the characteristics of neural network, makes full use of normative knowledge and empirical knowledge, and improves the intelligent level of traditional DSS. Through the above analysis, it is not difficult to find that there are still many unsatisfactory contents for this system, which need further improvement and perfection. DSS is only a solution for an enterprise. Of course, it is not universal, but its data management subsystem is powerful and relatively independent in design, which has great market potential and can be developed into products.

References

1. Guo, Z.X., Ngai, E., Yang, C., Liang, X.D.: An RFID-based intelligent decision support system architecture for production monitoring and scheduling in a distributed manufacturing environment. Int. J. Prod. Econ. **159**(78), 16–28 (2015)
2. Dweiri, F., Kumar, S., Khan, S.A., Jain, V.: Designing an integrated AHP based decision support system for supplier selection in automotive industry. Expert Syst. Appl. **62**(15), 273–283 (2016)
3. Zhou, Z.G., Liu, F., Li, L.L., Jiao, L.C., Zhou, Z.J., Yang, J.B., Wang, Z.L.: A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer. Knowl.-Based Syst. **85**(10), 62–70 (2015)
4. Deptula, A., Osinski, P., Radziwanowska, U.: Decision support system for identifying technical condition of combustion engine. Arch. Acoust. **41**(3), 449–460 (2016)
5. Miller, K.J.: Towards a distributed mobile agent decision support system for optimal patient drug prescription. Inf. Syst. Front. **19**(1), 1–20 (2015)
6. Wang, Y., Montas, H.J., Brubaker, K.L., Leisnham, T., Shirmohammadi, A., Chanse, V., Rockler, K.: A Diagnostic decision support system for BMP Selection in small urban watershed. Water Resour. Manage. **31**(5), 1649–1664 (2017)
7. Rinaldi, F., Jonsson, R., Sallnäs, O., Trubins, R.: Behavioral modelling in a decision support system. Forests **6**(2), 311–327 (2015)
8. Keum, J., Coulibaly, P.: Information theory-based decision support system for integrated design of multivariable hydrometric networks. Water Resour. Res. **53**(7), 1–21 (2017)
9. Kopka, P., Mazur, A., Potempski, S., Wojciechowicz, H.: Application of the RODOS decision support system for nuclear emergencies to the analysis of possible consequences of severe accident in distant receptors. Ann. Nucl. Energy **167**, 108837 (2022)
10. Kar, A.K.: A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network. J. Comput. Sci. **6**(21), 23–33 (2015)
11. Lei, N., Moon, S.K.: A decision support system for market-driven product positioning and design. Decis. Support Syst. **69**(13), 82–91 (2015)
12. Camacho-Collados, M., Liberatore, F.: A decision support system for predictive police patrolling. Decis. Support Syst. **75**(8), 25–37 (2015)
13. Wanderer, T., Herle, S.: Creating a spatial multi-criteria decision support system for energy related integrated environmental impact assessment. Environ. Impact Assess. Rev. **52**(15), 2–8 (2015)

Chapter 21
Performance Evaluation of Container Identification Detection Algorithm

Zhangzhao Liang, Wenfeng Pan, Xinru Li, Jie You, Zhihao Long, Wenba Li, Zijun Tan, Jianhong Zhou, and Ying Xu

Abstract The identification of container surfaces carries a large amount of crucial information regarding production and logistics. Research on the detection and identification of containers is lacking in both academia and industry, and the efficiency is low due to the need for manual completion of related tasks. In order to tackle this problem, we have created a large-scale text detection dataset for container surface identification called IdentificationText. This dataset consists of 12,000 high-resolution images, providing bounding boxes annotations for text detection tasks. We have discussed downstream applications of the IdentificationText dataset as well as our annotation techniques used in the dataset's creation. The text in this dataset exhibits challenges such as deformations, multi-direction, and multi-scale. We conducted extensive experiments to evaluate the effectiveness and difficulty of this dataset using advanced text detection methods. In our experiments, we found that repeated textures and vertical text at multiple scales would cause missed detections, which was an extremely serious problem. The experimental results indicate that it is challenging for current text detection methods to locating text on container surfaces. Achieving higher accuracy in detecting text on containers requires more in-depth research. The experimental results serve as the benchmark performance for the IdentificationText dataset, providing reference for future researchers.

21.1 Introduction

Container is one of the most important carriers of modern transportation. The identification on the container carries information about logistics transportation and manufacturing. Container production, logistics, and other container-related industries typically need to verify or extract information from these identifications for tasks such as quality inspection, scheduling, and management. Recently, the powerful feature extraction ability and excellent performance of deep learning have been demonstrated

Z. Liang · W. Pan · X. Li · J. You · Z. Long · W. Li · Z. Tan · J. Zhou · Y. Xu (✉)
Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, Guangdong, China
e-mail: xuying117@163.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_21

in [1]. Therefore, deep learning-based methods have gradually gained favor among industry practitioners, especially in the fields related to containers. The application of deep learning technology in container-related tasks can overcome the limitations of manual tasks, such as safety and physical issues, by automating them. This reduces costs and improves productivity.

However, as deep learning is a data-driven technology, there is still a lack of datasets on container identification in academia and industry, which seriously hinders the application of deep learning in container-related fields. Therefore, developing a container identification dataset can effectively promote the development of deep learning in container-related fields and accelerate the process of industrial intelligence. To this end, we have developed a large-scale container identification text detection dataset named IdentificationText, which consists of over 9000 training images and 3000 test images. The images in this dataset were captured using various devices in different scenes, with resolutions of 1920 × 1080 or higher, thus demonstrating diverse characteristics and data generalization capabilities. Figure 21.1 shows sample images of IdentificationText dataset. We conducted comprehensive experiments on the proposed dataset. And we provided experimental results of the current advanced methods on the IdentificationText dataset, which are regarded as a benchmark. This benchmark can serve as relevant performance references for future researchers. The main contributions of this article are as follows:

Fig. 21.1 Sample images from the IdentificationText dataset

1. We have constructed a real-world container identification text dataset named IdentificationText, which consists of 12,000 high-resolution images for container identification text detection tasks.
2. To accurately measure the difficulty level of the IdentificationText dataset, we selected text detection methods from the past five years to evaluate their performance on the IdentificationText dataset and used it as a benchmark.

21.2 Related Work

The scene text detection is a developing branch of OCR tasks. The features of scene text include complex background, blur, irregular shape, and text oriented in various directions. Karatzas et al. introduced ICDAR2013 [2] and ICDAR2015 [3] as competitions for document analysis that provided word-level annotations. The images in these datasets depict natural scenes. The annotation information in these datasets can be used for both text detection and recognition. Though these datasets are widely used in OCR research, they cannot be easily generalized to various scenarios due to their small scale. Gupta et al. [4] proposed a large-scale synthetic dataset of text images to provide abundant data for neural network training. This dataset consists of more than 0.85 million natural images onto which text content is superimposed. In recent years, a large number of challenging datasets have been proposed to address more complex OCR problems, such as those involving handwriting or skewed text. Total-text [5] and CTW1500 [6] are proposed to be used for the curved text recognition task, with polygon annotations provided for images that contain at least one clearly visible trademark or signboard with curved text. Additionally, Singh et al. [7] constructed an arbitrarily-shaped scene text dataset named TextOCR, which consists of over 28 K images where the texts appear in various orientations. For multi-lingual OCR tasks, the ICDAR2017-MLT [8] and ICDAR2019-MLT [9] datasets are suitable natural scene text datasets with up to 9 or 10 different scripts, such as Chinese, English, or Arabic. The rectangle annotation is used for text localization, and the word-level annotation is used for recognizing text. The impressive feature extraction ability of deep neural networks in recent years has facilitated the industrial application of scene text detection and recognition technology. However, there is still a lack of open-source research or practice in the manufacturing of containers. Some companies in the transportation industry create container number datasets and use OCR technology solely for container number recognition. However, these datasets are limited in scope, with annotations only including container numbers and ignoring other marks on the surface of containers. Therefore, to promote deep learning in container-related fields, we have proposed IdentificationText, which is the largest-scale container identification dataset that we are aware of Table 21.1 compares the IdentificationText dataset with other scene text datasets in terms of dataset scale.

Table 21.1 IdentificationText versus other datasets

Dataset	Images	
	Train	Test
ICDAR2013 [2]	229	233
ICDAR2015 [3]	1000	500
SynthText [4]	800 K	–
Total-text [5]	1255	300
CTW1500 [6]	1000	500
TextOCR [7]	24,902	3232
ICDAR2017-MLT [8]	9000	9000
ICDAR2019-MLT [9]	10,000	10,000
IdentificationText (ours)	9359	3084

21.3 IdentificationText Dataset

In this section, we first provide a detailed description of the IdentificationText dataset and analyze both its data characteristics and challenges. Furthermore, we introduce the annotation method utilized on the dataset.

21.3.1 Dataset Description

The performance of deep learning is significantly impacted by dataset scale and annotation quality. When data are insufficient, a model is prone to overfit. Poor annotation quality will seriously affect the quality of models' feature learning. There is currently no large-scale dataset with annotations for detecting text on containers. Thus, we proposed a large-scale container text detection dataset and provided annotations for detecting text on containers. All images and annotations have undergone multiple manual cleaning and auditing to ensure their quality. Specifically, this dataset consists of 12 K container images with text box annotations. We have annotated texts of different scales on the containers, which mainly comprise numbers, english, punctuation marks, as well as a small number of chinese characters and mathematical notations. In order to ensure strong generalization of the dataset, images with various devices, resolutions, and angles were captured and collected. In terms of data characteristics and task challenges, the container's uneven surface causes distortion and occlusion of the text. Examples of the text in the proposed dataset are shown in Fig. 21.2. Additionally, the repetitive textures on the surface of containers cause interference effects on text detection, and like many text detection datasets, container text also has characteristics of multiscaling. Table 21.2 analyzes the scales distribution of text instances in the IdentificationText dataset.

Fig. 21.2 Deformation, indistinct text, and multidirectional text in the images of the Identification-Text dataset

Table 21.2 Scales distribution of text instances in IdentificationText dataset

	Train	Test	Total
Small	17,352	5473	22,825
Mid	60,134	21,139	81,273
Large	26,876	10,677	37,553

21.3.2 Annotation Method

Semi-automatic annotation, assisted by a pre-trained model that has been trained on a similar task, is currently a popular method of annotation. Nevertheless, when it comes to container text, many of the open-source pre-trained models currently used in the academic community have proven to be ineffective; in addition, there is a scarcity of training data that pertain to container text. To address the growing amount of container text data during the annotation process, we have implemented an incremental learning approach [10]. This strategy reduces the expense of manual annotation while maximizing efficiency. Figure 21.3 outlines the annotation process in detail. Initially, pre-trained open-source weights are utilized on the SynthText dataset for automated annotation. Afterwards, the annotation results undergo manual review and modification. Subsequently, the reviewed data are merged in proportion with the SynthText dataset, creating a mixed subset for fine-tuning the pre-trained model.

21.3.3 Downstream Application of IdentificationText Dataset

Our proposed dataset is a genuine representation of an open-world scenario, comprising a vast variety of containers depicting different models, usage levels, and

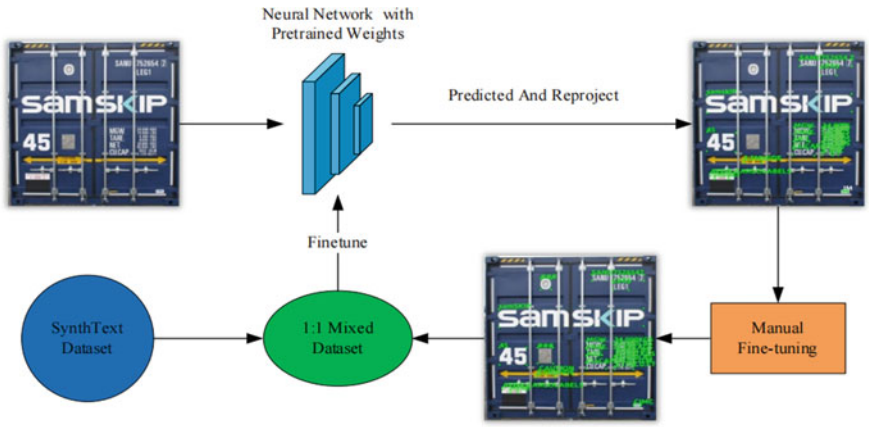


Fig. 21.3 Flowchart of semi-annotation

scenarios. In our selection process, we have carefully chosen an extensive selection of container images based on equipment used for shooting, environmental lighting, degree of damage, container type, and image distortion. As a result, modifying the annotation information allows the images in this dataset to be applied to other deep learning-related tasks, such as container classification and damage detection.

21.4 Benchmark

In order to establish benchmark performance for the IdentificationText dataset, we conducted extensive experiments on currently available advanced text detection algorithms. Our experimental results provide a performance reference for future researchers who wish to compare their results to these benchmarks.

21.4.1 Experimental Setup

To ensure fair and accurate comparisons between experiments, all tests were conducted on an Intel Core i5-12600K CPU, NVIDIA RTX 3060 12 GB GPU, using Python 3.8, PyTorch 1.13, CUDA 11.6 on Ubuntu 18.04. We followed the experimental setup of [11–15], except for replacing the backbone of the model with ResNet-50 to ensure the same feature extraction ability of the model. ResNet-50 serves as the backbone for all selected models and initializes image classification tasks using pre-trained weights in the ImageNet dataset. We employed the Adam optimizer and PolyLR learning rate update strategy with an initial learning rate of 0.0001 and a batch size of 4 during the training process.

Table 21.3 Experiment results of text detection

Method	Hmean	Precision	Recall
TextSnake [11]	0.7858	0.8557	0.7264
PANet [12]	0.6699	0.7868	0.5833
PSENet [13]	0.7560	0.8082	0.7102
DBNet [14]	0.8028	0.8572	0.7548
DBNet++ [15]	0.8417	0.8764	0.8096

21.4.2 Evaluation Metric

As was common in most OCR datasets [2–9], we evaluated text detection performance using Precision, Recall, and Hmean metrics. If a text was unrecognizable, we flagged it as a difficult sample, which contributed to the training process. However, during testing, it was automatically excluded to prevent adverse impacts on the final performance evaluation.

21.4.3 Result and Analysis

21.4.3.1 Text Detection Result

Deep learning has proven to be highly effective in many visual tasks, facilitating the development of numerous deep learning-based approaches for tackling challenges in the OCR field. To establish a benchmark performance for the IdentificationText dataset, we selected several methods from recent years and conducted numerous experiments. The chosen methods performed well and exhibited similar levels of performance. The experimental results of the text detection are presented in Table 21.3. Despite recent progress in text detection on the IdentificationText dataset, significant challenges persist due to its high complexity. Specifically, we conducted a visualization analysis of the experimental results to evaluate the performance of the selected methods.

21.4.3.2 Text Detection Analysis

In the text detection experiment, DBNet++ showed the highest comprehensive detection performance among all methods with scores of 0.8417, 0.8764, and 0.8096 for Hmean, Precision, and Recall, respectively. The novel differentiable binarization method proposed by DBNet++ [15] addresses the shortcomings of traditional binarization techniques, which rely on fixed thresholds and lack differentiability. Integrating the binarization process into the training procedure allows the model to learn the optimal binarization threshold automatically. Furthermore, the proposed

adaptive scale fusion module enhances the model's ability to adapt to multi-scale text. Despite this improvement, the model exhibits a noticeable difference between Precision and Recall of 0.07, indicating a relatively sizeable number of missed detections. Our visual analysis revealed that most missed detections occur in images with both vertical and multi-scale text. The results of the DBNet++ model are presented in Fig. 21.4. This issue has been observed in other models as well as in DBNet++. Therefore, the development of detectors capable of adapting to vertical text and more efficient multi-scale feature extraction and fusion structures are crucial in overcoming the challenges of detecting container text.

21.5 Conclusion

In this paper, we presented the IdentificationText dataset for detecting text on container surfaces. This dataset comprises 12 k high-resolution images, each annotated with bounding boxes for text detecting purposes. As far as we know, the IdentificationText dataset is presently the most substantial collection of container images, characterized by both diversity and versatility. Consequently, in addition to primarily serving as a resource for container identification text detection tasks, the dataset is also suitable for other container-related applications. The dataset production process included an annotation process that utilized incremental learning and semi-automatic annotation methods. These methods resulted in a reduction in the OCR dataset's annotation cost. We conducted multiple experiments on detecting identification text on containers utilizing the latest state-of-the-art methods. The results indicate that the problem of multi-scale and vertical text loss remains challenging in text detection tasks. Therefore, in future research, more refined multi-scale information fusion mechanisms can be added to neural networks or corresponding improvements can be made for vertical texts to alleviate model miss detection caused by multi-scale and vertical texts. Our experimental results serve as the benchmark performance of this dataset and provide a reference for future researchers.

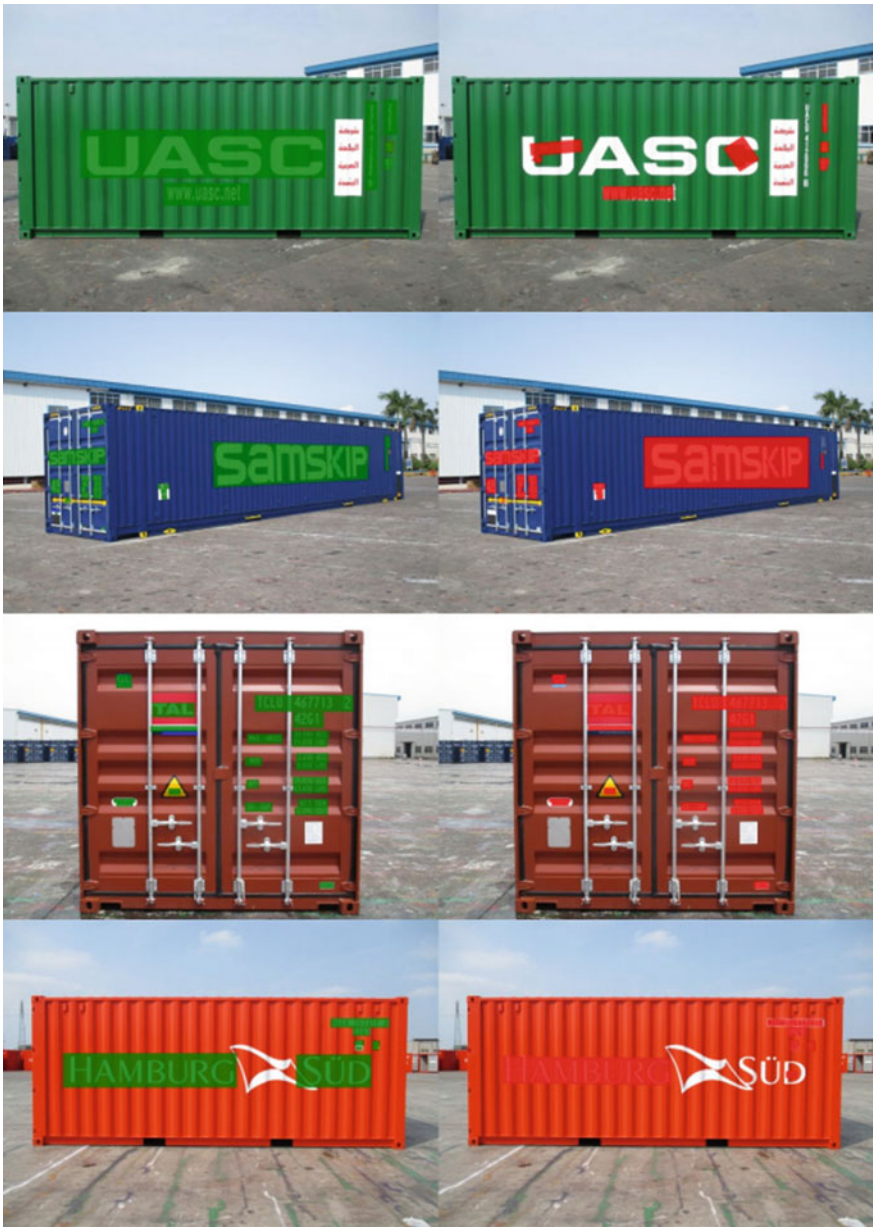


Fig. 21.4 DBNet++ text detection results. (Left) Ground truth highlight with green mask. (Right) Prediction highlight with red mask

Acknowledgements This work was supported by Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515011576), Guangdong, Hong Kong, Macao and the Greater Bay Area International Science and Technology Innovation Cooperation Project (No. 2021A050530080, No. 2021A0505060011), Key Research Projects for the Universities of Guangdong Provincial Education Department (No. 2020ZDZX3031, No. 2022ZDZX1032), Jiangmen Basic and Applied Basic Research Key Project (2021030103230006670), Jiangmen Science and Technology Plan Project (2220002000246), and Key Laboratory of Public Big Data in Guizhou Province (No. 2019BDKFJJ015), Development of a Container Intelligent Panoramic Trademark Quality Inspection System Based on Machine Vision(HX22105).

References

1. Bengio, Y., Lecun, Y., Hinton, G.: Deep learning for AI. Commun. ACM **64**(7), 58–65 (2021)
2. Karatzas, D., et al.: ICDAR 2013 Robust reading competition. In: 2013 12th International Conference on Document Analysis and Recognition, pp. 1484–1493, Washington, DC, USA (2013)
3. Karatzas, D., et al.: ICDAR 2015 robust reading competition. In: 13th International Conference on Document Analysis and Recognition, pp. 1156–1160, Tunis, Tunisia (2015)
4. Gupta, A., Vedaldi, A., Zisserman, A.: Synthetic data for text localisation in natural images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2315–2324 (2016)
5. Ch'ng, C.K., Chan, C.S., Liu, C.L.: Total-text: towards orientation robustness in scene text detection. Int. J. Doc. Anal. Recogn. **23**, 31–52 (2020)
6. Liu, Y., Jin, L., Zhang, S., et al.: Curved scene text detection via transverse and longitudinal sequence connection. Pattern Recogn. **90**, 337–345 (2019)
7. Singh, A., Pang, G., Toh, M., et al.: Textocr: towards large-scale end-to-end reasoning for arbitrary-shaped scene text. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8802–8812 (2021)
8. Nayef, N., Yin, F., Bizid, I., et al.: ICDAR 2017 robust reading challenge on multi-lingual scene text detection and script identification-RRC-MLT. In: 14th International Conference on Document Analysis and Recognition, pp. 1454–1459 (2017)
9. Nayef, N., Patel, Y., Busta, M., et al.: ICDAR 2019 robust reading challenge on multi-lingual scene text detection and recognition-RRC-MLT-2019. In: 2019 International Conference on Document Analysis and Recognition, pp. 1582–1587(2019)
10. Rebuffi, S.A., Kolesnikov, A., Sperl, G., et al.: ICARL: Incremental classifier and representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2001–2010 (2017)
11. Long, S., Ruan, J., Zhang, W., et al.: Textsnake: a flexible representation for detecting text of arbitrary shapes. In: Proceedings of the European Conference on Computer Vision, pp. 20–36(2018)
12. Wang, W., Xie, E., Song, X., et al.: Efficient and accurate arbitrary-shaped text detection with pixel aggregation network. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 8440–8449 (2019)
13. Wang, W., Xie, E., Li, X., et al.: Shape robust text detection with progressive scale expansion network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9336–9345 (2019)
14. Liao, M., Wan, Z., Yao, C., et al.: Real-time scene text detection with differentiable binarization. Proc. AAAI Conf. Artif. Intell. **34**(07), 11474–11481 (2020)
15. Liao, M., Zou, Z., Wan, Z., et al.: Real-time scene text detection with differentiable binarization and adaptive scale fusion. IEEE Trans. Pattern Anal. Mach. Intell. **45**(1), 919–931 (2022)

Chapter 22
Application and Prospect of Deep Learning and Machine Learning Technology

Qiaoni Zhao and Tong Liu

Abstract This paper systematically expounds the relationship between the current hot concepts such as artificial intelligence, machine learning and deep learning, and makes it clear that the technical category covered by them is a descending hierarchical relationship. In addition, deep learning and machine learning technologies are compared and analysed from five aspects: data volume, data feature analysis, hardware dependency, model training time and learning algorithm types. Finally, the application fields and prospects of deep learning and machine learning are summarized.

22.1 Introduction

Since the ChatGPT chatbot was launched in 2022, artificial intelligence has received widespread attention around the world. Artificial intelligence is the computer to simulate people's thinking processes and intelligent behaviour, which is an important technology emerging in the twenty-first century [1].

Artificial intelligence, from birth to date, has 67 years of development history, its development process is not smooth, but full of twists and turns, has experienced the winter trough period, but also experienced a brilliant and prosperous period; its development process can be generally divided into three stages such as the embryonic stage, exploration stage and high-speed development stage, as shown in Fig. 22.1; the embryonic stage is the early artificial intelligence [2–4].

(1) Embryonic stage: (1950–1980s): The "artificial intelligence" was first coined in 1956 by cognitive scientist John McCarthy, who pointed out that artificial intelligence is about making machines behave as if they were intelligent as humans do. That is, any learning behaviour or intelligence characteristic behaviour of human being can be accurately described or simulated by machines [5].

Q. Zhao · T. Liu (✉)
Hunan Railway Professional Technology College, ZhuZhou 412001, China
e-mail: 569434706@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_22

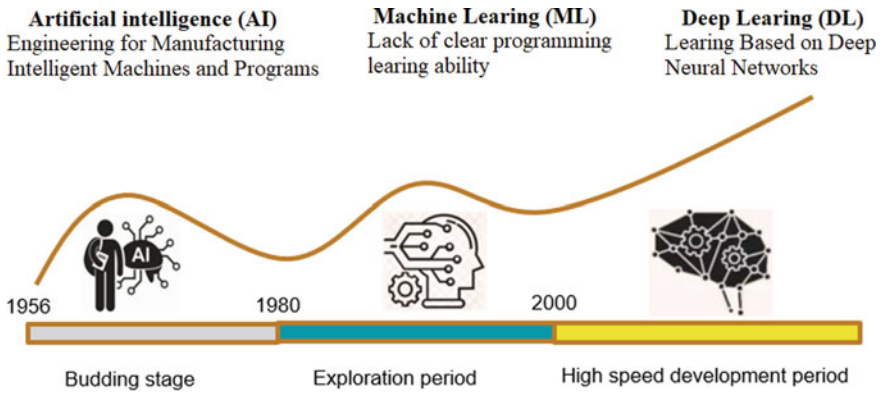


Fig. 22.1 The development of artificial intelligence

Since the first proposed the term of "artificial intelligence", artificial intelligence has ushered in the history of the first small peak, researchers research enthusiasm is high, after more than a decade to achieve many breakthrough results, such as the birth of industrial robots, chatbots, machine theorem proof, checkers program and other major achievements. However, since most of the early artificial intelligence completed specific tasks through fixed instructions, if the research problem became complex, artificial intelligence programs were overwhelmed, limited by the limitations of the algorithm theory and computer computing power performance at that time, artificial intelligence ushered in the first winter, and the research fell into the trough bottleneck [6, 7].

(2) Exploration period (1980–1990s): Although some people denied the development and value of artificial intelligence at that time, researchers did not stop the pace of progress. In the process of research, researchers realized the importance of knowledge base for solving complex tasks, so they built a variety of expert knowledge base systems and achieved fruitful results in specific fields [8].

At this moment, machine learning began to rise, and a variety of expert systems have emerged. Expert systems promote artificial intelligence from theoretical research to practical application, and achieve success in medical, meteorological, geology and other fields [9].

However, with the expansion of the "artificial intelligence", the shortcomings of expert systems have gradually emerged: Such as narrow application field, single reasoning method, lack of common sense knowledge and other shortcomings, these exposed problems make the development of artificial intelligence into a stagnant state, although there are also neural network algorithms at this stage, but due to the hardware performance limitations of the computer at that time, there is no good landing effect, artificial intelligence ushered in the second winter period [10].

(3) Rapid development period: (2000 ~ present): Thanks to improvements in information technology hardware and software, it provides the basic conditions for the development of artificial intelligence. At this stage, the theoretical algorithms of artificial intelligence are also constantly being precipitated, and the algorithms represented by statistical machine learning have achieved better application results in many fields such as the Internet and industry. To solve many of the principles of human intelligent behaviour, such as language understanding and image understanding, researchers have turned their research focus to the stage of computer learning in "data". In 2006, Professor Hinton proposed the concept of "deep learning"; in 2012, Hinton led his research team to participate in the ImageNet image recognition competition and won the championship by a wide margin, causing a stir in academia and industry, and drawing people's attention to deep learning.

In recent years, artificial intelligence algorithms represented by deep learning have made a great improvement in the categories of image processing, text detection, identity authentication, computer vision, voice recognition and natural language processing. On the one hand, the performance of computers has been greatly improved. The GPUs provide the computing power support and improve the availability of algorithms. New artificial intelligence chips and cloud computing technology provide the basic platform for large-scale neural network computing. On the other hand, the development of the information technology has accumulated massive data resources. That is, the combination of algorithms, computing power and data will push artificial intelligence into the boom period again [11–14].

22.2 Relationship Between Deep Learning and Machine Learning

Artificial intelligence (AI), any technology that generally instructs computers to mimic human intelligence; machine learning: referred to as ML, which refers to the ability of machines to use statistical techniques to gradually improve their ability to complete works using experience; deep learning (DL) is a research field that has developed very rapidly and has achieved great success in many subfields of artificial intelligence. The deep learning process is deep because the structure of an artificial neural network consists of multiple input, output and hidden layers. Each layer contains units that convert input data into information for the next layer to use for specific predictive tasks, and thanks to this structure, machines can learn through their own data processing [15].

Therefore, deep learning is a subset and branch of machine learning and is the focus area of machine learning; machine learning is a subset and branch of the field of artificial intelligence, as shown in Fig. 22.2.

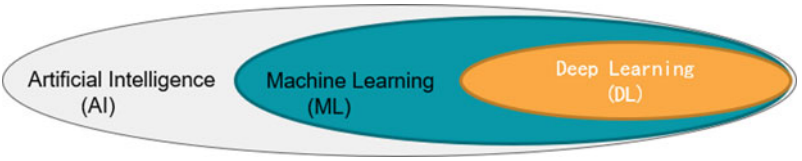


Fig. 22.2 Relationship between AI, ML and DL

22.3 Comparative Analysis of Machine Learning and Deep Learning

At present, machine learning and deep learning are two important branches in the field of AI, and their application prospects are very broad. Although the two are subordinate relationships, they are also very different. As shown in Table 22.1, the differences between the two are expounded from five aspects: data volume, data feature processing process, hardware requirements, model training time and learning methods [16–18].

22.3.1 Comparison of the Required Data Quantity

The performance of deep learning and machine learning for different amounts of data is different, as shown in Fig. 22.3. Deep learning performs better when there is a lot of data, and machine learning performs better when there is a relatively small amount of data.

Table 22.1 Comparative analysis of machine learning and deep learning

Content	Machine learning	Deep Learning
Data volume size	The data amount is small and simple	The data are large and complex
Data feature processing	Human intervention is required to provide accurate characteristics	Automatically obtain eigenvalues from large data sets and solve problems end-to-end to complete the learning process
Hardware requirement	Computer skills are not required	High-end GPU machines that can do a lot of matrix multiplication
Model training time	Anywhere from seconds to hours	Maybe half a month to complete the training
Type of learning algorithm	Linear regression, logistic regression, decision tree, etc	Convolutional neural network, recurrent neural network, etc

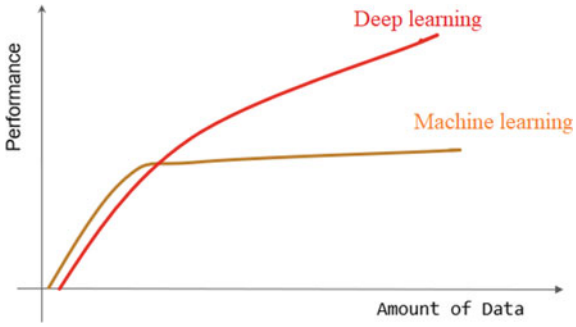


Fig. 22.3 Differences in the amount of data required for machine learning and deep learning

Deep learning requires to train machine models which needs large amounts of data, and as the amount of data increases, so does the generalization performance of its models. When the data are small, the deep learning algorithm cannot perform the task well, and it may lead to overfitting problems, which will affect the results of deep learning. This is because deep learning algorithms can fully understand the features of the model by learning from many data sets of different instances, from which they learn the features they need to find and generate outputs with probability vectors. The more data you have in deep learning, the more likely it is that your results will be correct.

Machine learning algorithm is suitable for use in the case of small amount of data, suitable for simple data and problems, and has better performance than deep learning while having small amount of data. With the increase of the amount of data, the performance of machine learning method tends to be flat. This is because traditional machine learning requires a process called feature extraction, where the programmer must explicitly tell a particular training set that certain features must be looked for, and when any of the features are missing, the machine learning model cannot recognize the object [19].

22.3.2 Comparison of Data Processing Processes

Machine learning and deep learning process data differently, as shown in Fig. 22.4. In general, traditional machine learning often needs to use manual experience to train computer systems and manually adjust system parameters, such as manual feature extraction + classification (algorithm) to adjust system parameters, extract rules from input data, and predict and classify new data, so the number of parameters is very limited.

Deep learning uses multi-layer neural networks and algorithms of massive data to deal with complex data and tasks. These neural networks imitate the working principle of human brain, save the trouble of manual data processing, automatically obtain thousands of parameters from big data, automatically learn effective features

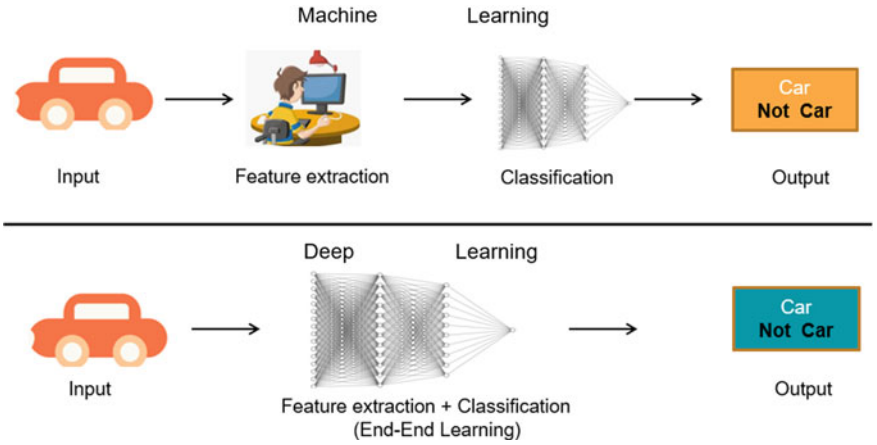


Fig. 22.4 Difference in processing data

from data, automatically classify data sets, and find correlations among them. It achieves end-to-end learning and more accurate data processing and prediction.

22.3.3 Hardware Dependency Comparison

In contrast, deep learning relies more on hardware, while machine learning has much lower hardware requirements.

Deep learning requires strong computer computing power, so the hardware configuration of the computer is naturally very high. This is because deep learning needs to deal with massive data, parallel computation and a large number of matrix multiplication calculations, which requires the use of high-end GPU graphics processors to help decompose complex tasks and perform multiple operations at the same time, and GPU needs to deal with a large number of data calculations to improve the running efficiency of the machine and save data processing time. It can be said that GPU is the core heart of deep learning applications, so deep learning needs to be configured with a dedicated GPU server.

Because of the small amount of data processed by traditional machine learning algorithms and the simple processing tasks, they can work on low-end machines relatively, and do not have high requirements for the hardware configuration of computers [20].

22.3.4 Comparison of Time Required for Algorithm Model Training

The training algorithm model of machine learning generally takes a short time, while the model training algorithm of deep learning needs more time, such as some complex deep learning even takes about 15 days to complete the training, because the deep learning model is more complex, training and adjusting a large number of parameters of the complex model takes a lot of time.

22.3.5 Comparison of Learning Algorithm Types

22.3.5.1 Machine Learning Algorithm

Machine learning algorithm is a class of algorithms used to learn rules from data. Common machine learning algorithms include linear regression, logistic regression, decision tree, support vector machine, etc.

(1) Linear regression

Linear model is the model that makes predictions through the linear combination of sample features and is a regression analysis that models the relationship between independent variables and dependent variables. Given an N-dimensional sample $\chi = [\chi_1; \chi_2; \chi_3; \ldots \chi_D;]^T$, a linear model is a vector:

$$f(x; w, b) = w^T x + b \tag{22.1}$$

(2) Logistic regression

Logistic regression is a commonly used linear model for binary classification problems. Used to predict the value of a categorical variable, it establishes a nonlinear relationship between an input variable and an output variable. The essence of logistic regression is to add sigmoid function because of linear regression function, and compress the value predicted by linear regression into a probability. Figure 22.5 shows an example of using linear regression and logistic regression to solve the binary classification problem of one-dimensional data.

(3) Decision tree

Decision tree contains three types of nodes: decision points (rectangular representation), opportunity nodes (circle representation) and result nodes (triangle representation), as shown in Fig. 22.6a.

For example: Suppose we have a project to decide on now. The management needs to decide whether to start construction next month. The corresponding decision tree is shown in Fig. 22.6b.

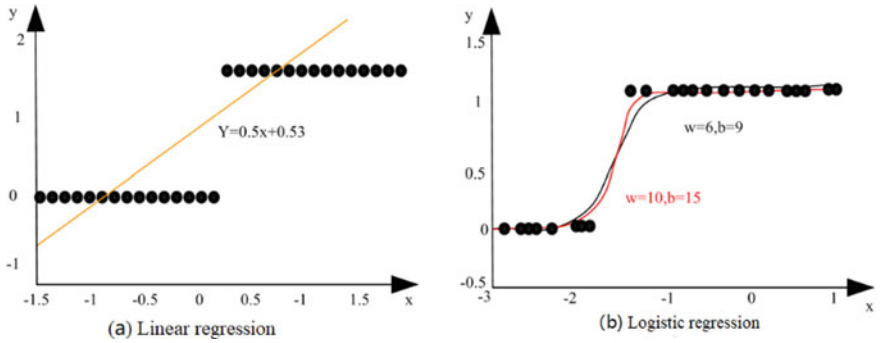


Fig. 22.5 Example of binary classification problem with one-dimensional data

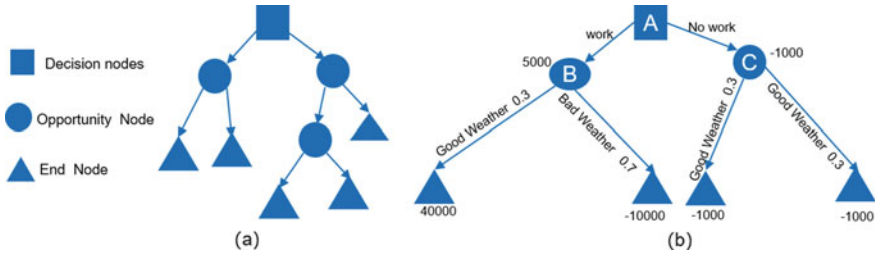


Fig. 22.6 Decision tree model

22.3.5.2 Deep Learning Algorithm Model

Deep learning can be used to process complex nonlinear relationships and be able to train on large-scale data. Common deep learning models include convolutional neural network (CNN), recurrent neural network (RNN) and so on.

(1) Convolutional Neural Network (CNN): a kind of neural network that mimics biology, mathematical or computational models of structure and function. It is mainly used to process data types such as images and videos, and extract features by sliding convolution kernel over input data. CNN is generally a feed-forward neural network concluding convolutional layer (feature extraction), pooling layer (under sampling) and fully connected layer (classification). As shown in Fig. 22.7, CNN is the schematic diagram of bird recognition process.

(2) Recurrent Neural Network (RNN)

RNN is composed of artificial neurons and one or more feedback loops, as shown in Fig. 22.8a. The network structure after expanding the loop contained in the hidden layer is shown in Fig. 22.8b.

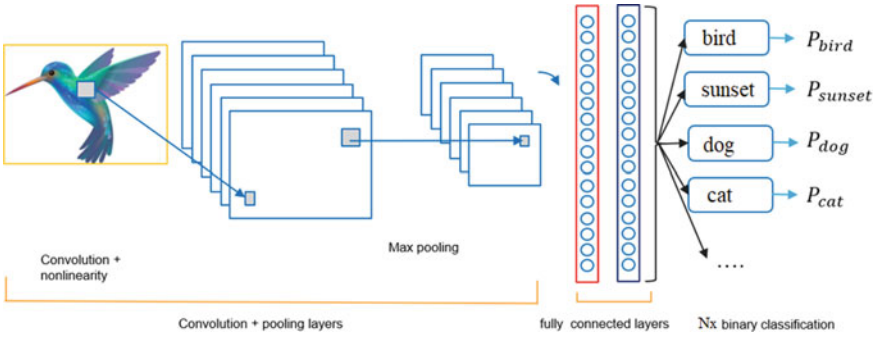


Fig. 22.7 CNN schematic of the process of identifying birds

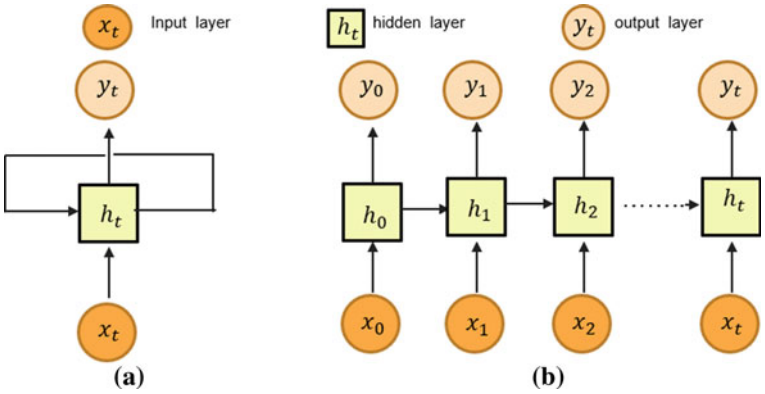


Fig. 22.8 RNN Network structure

22.4 Prospects

In the future, machine learning and deep learning will continue to be applied and developed in various fields. For example, machine learning and deep learning will achieve a higher level of autonomous learning and reasoning, so that computers can better understand and cope with complex real-world problems, and there will be more automated and intelligent products and services in the future, to achieve more comprehensive and in-depth artificial intelligence applications, to provide more convenient services for human life, and to bring a better future for mankind [21–23].

Acknowledgements This work was supported by Hunan Provincial Natural Science Foundation of China (2022JJ50096).

References

1. Wang, H., Zheng, S.J., Xu, R.N.: Modular robot self-reconfiguration algorithm based on swarm agent deep reinforcement learning. Comput. Sci. **2019**(06), 266–273 (2019)
2. Yan, H.: Application research of machine learning in library and information field. Publ. Stand. **2023**(12), 164–166 (2023)
3. Lv, X.Y., Zhao, S.Q., Su, Z.H.: Application of internet of things and machine learning in agriculture. Intell. Agricult. Guide **3**(12), 10–13 (2023)
4. Qi, H.O., Xu, X., Xue, Z.H.: Application of machine learning algorithm in operating room management. Nurs. Rehabil. **22**(06), 92–96 (2023)
5. Malik, J.: What led computer vision to deep learning. Commun. ACM **60**(6), 123–129 (2017)
6. Dahl, G.E., Dong, Y., Li, D.: Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. Audio Speech Lang. Process. IEEE Trans. **20**(1), 30–42 (2012)
7. Tian, W.H.: Research on curriculum design method of teaching resource library based on deep learning technology. In: 6th International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2023), vol. 3, pp. 234–239 (2023)
8. Yuan, L., Niu, T., Li, Z.P.: Study on teaching strategies of scratch programming in elementary schools based on deep learning theory. In: 5th International Workshop on Education Reform and Social Sciences, vol. 10, pp. 112–118 (2022)
9. Guo, X.K.: Analysis on the latest research hotspots of computer deep learning optimization algorithms. In: 4th International Conference on Computer Engineering, Information Science and Internet Technology, vol. 9, pp. 148–153 (2022)
10. Liu, L.Q., Wang, H.F., Liu, Y.Z., Zhang, M.: Task scheduling model of edge computing for AI flow computing in Internet of Things. In: 3rd International Symposium on Information Science and Engineering Technology, vol. 7, pp. 87–95 (2022)
11. Wang, J.Y., Li, X.Q.: Based on machine learning classification correlation study and species classification of glass artifact under the influence of weathering. In: 2023 International Conference on Mathematical Modeling, Algorithm and Computer Simulation (MMACS 2023) vol. 2, pp. 131–138 (2023)
12. Chen, X.R.: Stock price prediction using machine learning strategies. In: 2022 International Conference on Company Management, Accounting and Marketing, vol. 12, pp. 68–73 (2022)
13. Zhang, Z.: Research on sentiment analysis model of Dou ban book short reviews based on machine learning. In: 5th International Workshop on Education Reform and Social Sciences, vol. 10, pp. 321–327 (2022)
14. Zhou, Z.: Breast cancer diagnosis with machine learning. In: 4th International Conference on Computer Engineering, Information Science and Internet Technology, vol. 09, 157–164 (2022)
15. Su, H., Yang, X., Xiang, L.: Mechanical fault diagnosis under varying working conditions based on deep contrast transfer learning. J. Vibr. Eng. **36**(03), 845–853 (2019)
16. Jiang, R.Y., Wei, Y.Z., Wang, H.J.: Differential neural differentiator solution based on deep learning. Comput. Eng. Des. **44**(06), 1629–1634 (2019)
17. Yu, J.B., Yao, J.M., Xie, R.T.: Uhf tag recognition system based on deep learning. Comput. Sci. 50(S1), 666–671 (2019)
18. Liu, C.: An intelligent tunnel crack detection system based on deep learning algorithm. China Sci. Technol. Inf. **2023**(12), 96–98 (2023)
19. Liu, Y.F., Duan, T.Z., Gong. W.: Construction and application of proxy model for forward modeling of stratigraphic sedimentation based on deep learning. Acta Sedimentological Sonica **41**(03), 791–803 (2019)
20. Li, R.Y., Ye, Z.Q.: Fund return forecast based on Machine learning. Stat. Decis. **39**(11), 156–161 (2019)
21. Li Y., Liang Z.H., Chang Y.P.: Review and prospect of research on user's algorithm aversion to artificial intelligence products. J. Manage. **19**(11), 1725–1732 (2022)

22. Xu P., Xu X.Y.: Logical and analytical framework of enterprise management change in the era of artificial intelligence. Manage. World **46**(06), 583–590 (2020)
23. Guo F.: Reflections on the future development of strong artificial intelligence. Digital Technol. Appl. **41**(02), 73–75 (2023)

Chapter 23
Simulation of Intelligent Image Processing Model Based on Machine Learning Algorithm

Yanfei Zou

Abstract Medical image segmentation is an important link in medical image processing and analysis, and it is also a very arduous task. In the process of image processing and image analysis, image segmentation is the most basic step, and the accuracy of segmentation directly affects the accuracy of subsequent work, which is the premise of medical image analysis, understanding, description and three-dimensional reconstruction. The topic selected in this paper is the simulation of intelligent image processing model based on machine learning algorithm. A novel MRI (magnetic resonance imaging) image segmentation algorithm is obtained by combining machine learning with fuzzy theory. The spatial filter provides strong noise filtering performance. By guiding the clustering process of FCM (Fuzzy C-means) through the filtered results, we can expect to get stronger anti-noise ability than the standard FCM. The simulation results show that the improved FCM algorithm proposed in this paper has high diagnostic accuracy, high segmentation accuracy, stable algorithm and strong robustness. The improved FCM algorithm improves the anti-noise ability by fusing spatial filter, and this method has achieved good segmentation and anti-noise effect.

23.1 Introduction

Intelligent image processing technology keeps up with the forefront of science and technology, and its algorithm is widely used in practical production, which brings a lot of convenience and economic value, especially the image recognition technology based on computer vision. With the support of artificial intelligence algorithm, image recognition technology has been gradually applied to all walks of life. The combination of human image recognition and computer image recognition fully embodies the quality service brought by modern information technology and plays an extremely

Y. Zou (✉)
Department of Computer, Xian Yang Normal University, Xianyang 712000, China
e-mail: zouyanfei2021@163.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_23

important role in agriculture, medical care, transportation and other industries. This brand-new development model is gradually improving social value.

With the rapid development of computer technology, Internet technology and big data technology, computer-aided diagnosis and treatment has become more and more popular. Because medical image processing and analysis are restricted by many factors, such as image gray, texture, color and people's interest; medical image segmentation technology is a basic and classic problem [1, 2]. Literature [3] lists some current automatic segmentation algorithms, but the segmentation effect of most automatic segmentation algorithms is still far from practical application, and it is still in the laboratory stage. Literature [4] uses the classification method of Bayesian model to segment MRI (magnetic resonance imaging) images. Compared with other clustering algorithms, the spectral clustering algorithm based on normalized Laplacian matrix proposed in reference [5] has many obvious advantages, and has achieved good segmentation results in practice.

With the development of medical technology, all kinds of images related to medical diagnosis occupy a very important position in modern disease auxiliary diagnosis. When analyzing and reading grayscale medical images, the contrast, edge characteristics and signal-to-noise ratio of the images are very important for the correctness of diagnosis [6, 7]. Machine learning is an artificial intelligence science and the core of artificial intelligence. The purpose of its research is to use computers to simulate human learning behavior through learning, so as to acquire new knowledge and skills and continuously improve performance to achieve self-improvement. The topic selected in this paper is the simulation of intelligent image processing model based on machine learning algorithm. Combining machine learning with fuzzy theory, a novel MRI image segmentation algorithm is obtained, which solves the limitations of the existing segmentation algorithm and is much better than the traditional segmentation algorithm in segmentation accuracy and time.

23.2 Research Method

23.2.1 Medical Image Segmentation Method

There are still many problems to be solved in the research of computer-aided diagnosis, especially how to find the relationship between previously unknown symptoms and diseases and symptoms in a large number of case images, which has not been solved by manual diagnosis and computer-aided diagnosis at present. Medical image segmentation is an important link in medical image processing and analysis, and it is also a very arduous task. The so-called medical image segmentation is the process of dividing the medical image into several disjoint "connected" regions according to some similar characteristics of the medical image [8].

In the process of image processing and image analysis, image segmentation is the most basic step, and the accuracy of segmentation directly affects the accuracy of

subsequent work, which is the premise of medical image analysis, understanding, description and three-dimensional reconstruction. Threshold-based segmentation method is a traditional segmentation method, which is simple to implement, requires little calculation and has relatively stable performance. The image of the target object is extracted from the background image, and the image is segmented by using the edge, which is suitable for the image in which the target object and the background occupy different grayscale ranges [9]. At present, in the field of medical research, the commonly used evaluation criterion is to compare the segmentation results of the computer with the real results, so as to evaluate the performance of this algorithm. The real image here is the result of manual segmentation by clinicians and experts according to their professional knowledge and experience.

Canny operator is a commonly used optimal edge detection operator [10]. Compared with the general edge detection operator Canny operator has several obvious advantages: Canny operator realizes the optimal edge detection, has strong anti-noise ability, and the edge detection result has the width of a single pixel. Its disadvantage is that there are many processing steps, so the time and space complexity are high. The general flow of edge detection of Canny operator is shown in Fig. 23.1.

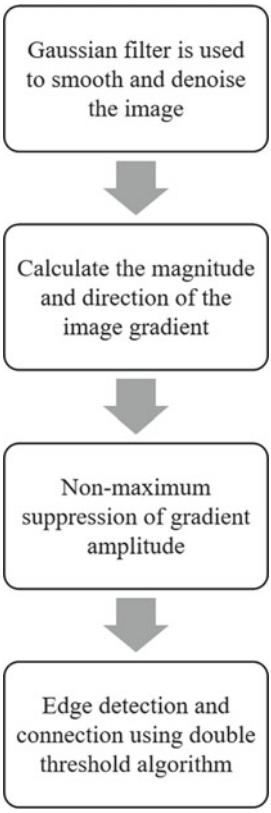


Fig. 23.1 Canny operator edge detection process

The purpose of region segmentation is to connect pixels with some similar properties and finally form a segmented region. Edge segmentation is a method to find the edge of a region based on the discontinuity of gray level, and region segmentation determines the belonging region of each pixel based on the criterion of consistent pixel attribute characteristics (such as brightness). The key to this process is to choose the appropriate growth criteria first. For the selection of growth criteria, features such as brightness value, texture, shape and gradient information can generally be selected.

Mathematical morphology is a nonlinear image processing and analysis method, and the basic operations of morphology include expansion, erosion, opening and closing [11]. If f is the original image and b is the structural element diagram, the gray expansion operation of image f using the structural element diagram b is denoted as $f \oplus b$, and D_b is the localization of b, and the definition formula is as follows:

$$(f \oplus b)(x, y) = \max\left\{ f\left(x - x', y - y'\right) + b\left(x', y'\right) \middle| \left(x', y'\right) \in D_b \right\} \tag{23.1}$$

Morphological open operation can deal with isolated small points, eliminate burrs and small bridges (i.e., small points connecting two areas), while the overall position and shape remain unchanged. Definition formula of binary image's open operation is as follows:

$$(f \circ b) = (f \oplus b) \oplus b \tag{23.2}$$

In the field of medical research, many uncertainties can be dealt with by fuzzy theory. In view of the strong anti-noise ability of fuzzy theory itself, it is the best choice to use fuzzy technology to segment medical images. FCM (Fuzzy C-means) clustering is one of the most widely used methods. It is an unsupervised clustering method, which is based on unsupervised theoretical system and supervised theoretical system. In order to meet the needs of practical problems, many researchers have devoted themselves to the research field of fuzzy clustering and have improved this traditional algorithm and achieved good results.

23.2.2 Improved FCM Intelligent Image Processing

From the perspective of machine learning, clustering is an unsupervised learning process. Unlike classification, unsupervised learning does not require predefined classes or training samples with class labels, while clustering learning algorithm automatically determines labels. Clustering is observational learning, not exemplary learning [12, 13]. Up to now, many researchers have put forward many clustering algorithms, such as density-based clustering algorithm, partition-based clustering algorithm, hierarchical clustering algorithm, grid-based clustering algorithm and graph-based clustering algorithm.

For a given data set, firstly, it is necessary to judge whether there is a clustering structure in the data set. If it is determined that there is a clustering structure in the data set, it is necessary to find the clustering structure through an algorithm. Secondly, after using the algorithm to get the clustering structure, it is necessary to test the obtained clustering structure, verify whether the clustering result is reasonable through research, and finally verify the validity of the clustering result. From a single point of view, we can't solve all the clustering validity problems. To verify the validity of fuzzy clustering results, that is, to verify the rationality of the number of clusters, we need to find a function that is more in line with the actual classification, and FCM clustering solves this problem. The algorithm can effectively guide clustering, so as to get a more practical clustering result.

FCM clustering algorithm optimizes the objective function of clustering analysis, which can be expressed by mathematical expression as follows:

$$\min J(X, \mu, v) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} d_{ij}^{2} \tag{23.3}$$

d_{ij} is used to measure the similarity between samples, and Euclidean distance is used here, that is to say, d_{ij} represents the Euclidean distance from the jth element X_j in the sample set to the cluster center of class i.

Medical image segmentation is a challenging task, and fuzzy clustering algorithm is especially suitable for solving the fuzzy medical image segmentation problem [14]. Scholars have used the classical FCM algorithm to solve the problem of medical image segmentation and achieved good results. However, the classical FCM algorithm only considers the gray value of pixels, and the actually processed MRI images are often inevitably disturbed by noise to some extent.

Because the standard FCM method only uses the gray information of the image, its tolerance to noise is poor. One way to improve the anti-noise ability of FCM method is to integrate the information of spatial filter. The spatial filter provides strong noise filtering performance. By guiding the clustering process of FCM through the filtered results, we can expect to get stronger anti-noise ability than the standard FCM. The improved FCM intelligent image processing flow is shown in Fig. 23.2.

Spatial filter has the advantages of good denoising effect and less computation and is widely used in image preprocessing. Using the denoising performance of spatial filter and the classification ability of FCM, an improved FCM fuzzy clustering algorithm with good anti-noise ability can be proposed.

The block subgraph is regarded as a single sample, and each sample subgraph contains several data points s_i. To find the intra-class similarity matrix A^{np*np} that constitutes the new sample, the similarity relationship between data points in the sample can be established by Gaussian kernel function, and the formula is as follows:

$$a_{i,j} = \exp\left(-\frac{\mathrm{MD}}{2\sigma}\right) \tag{23.4}$$

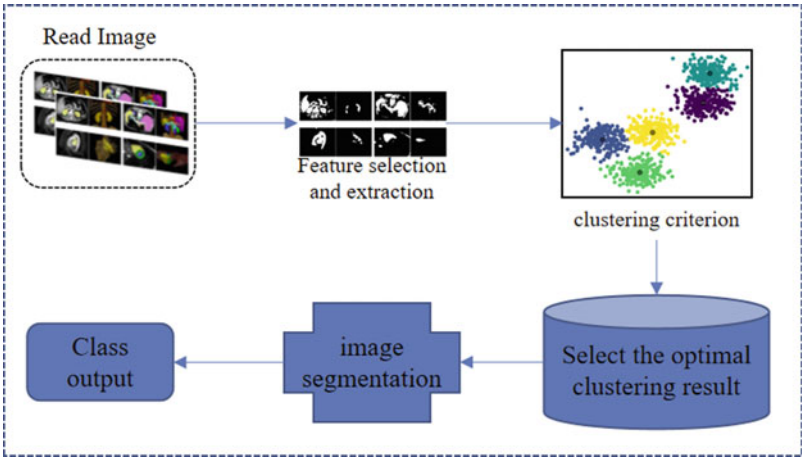


Fig. 23.2 Improved FCM intelligent image processing flow

Among them, MD represents the geometric distance between two data points, and σ is the scale parameter.

The advantage of fuzzy theory is that many complicated or uncertain nonlinear problems can be described and solved by simple rules, and it has high adaptability to pattern recognition and image processing. Fuzzy theory has been successfully applied in many different fields, such as nonlinear control, data classification, function modeling, pattern recognition and video compression [15].

In this paper, we exclude Euclid measurement method and use the elastic function method of fuzzy segmentation in fuzzy theory to measure the performance of code vectors. In this case, the Gaussian fuzzy membership function is used to develop this soft decision machine. For the j-th fuzzy rule, HC can be expressed as follows:

$$\mathrm{HC}_{(j,i)}\left(x_i, c_{ji}, \gamma\right) = \exp\left[\frac{\left(x_i, c_{ji}\right)^2}{\gamma^2}\right] \tag{23.5}$$

The attribute function value of fuzzy set is used as a soft estimator to check the similarity between video and code vector. The data with the largest attribute value are the video pattern closest to the code vector. The contour of the attribute function with different shapes and central positions can be used to approximate the appropriate fuzzy segmentation, and the codebook of vector quantization coding can be designed by fuzzy inference.

Spatial filters are often used to remove noise from images. Commonly used spatial filters include Gaussian filter and median filter, which are suitable for Gaussian noise and salt and pepper noise, respectively. The standard FCM method is improved, and the objective function of the improved FCM method with spatial filter is proposed as follows:

$$J_m = \sum_{i=1}^{c} \sum_{(x,y)\in I} \mu_i^m(x, y)\|I(x, y) - v_i\|^2 + \alpha \sum_{i=1}^{c} \sum_{(x,y)\in I} \mu_i^m(x, y)\|G(x, y) - v_i\|^2 + \beta \sum_{i=1}^{c} \sum_{(x,y)\in I} \mu_i^m(x, y)\|M(x, y) - v_i\|^2 \quad (23.6)$$

$M(x, y)$ is the data set in the unfiltered image (x, y). c is the number of cluster centers, m is the modular paste weighted index of the algorithm, $V = \{v_i\}$ is the centripetal quantity in the cluster, and $\mu = \left\{\mu_i^m(x, y)\right\}$ is the matrix of fuzzy membership degree of pixel (x, y) to the ith cluster center. $G(x, y)$ is the filtering result of Gaussian filter, and $M(x, y)$ is the filtering result of median filter. α, β is the penalty factor.

23.3 Simulation Analysis

For the feasibility and effectiveness of the algorithm, the simulation experiment is carried out with MATLAB R2018b software. The artificial images and MRI images of brain tumors provided by a tumor hospital were used for processing, respectively. The algorithm is implemented on a computer with advanced configuration. The main hardware configuration is as follows: Intel dual-core processor, 16G memory, 17-inch LCD, etc. Software: Win10 operating system.

Three quantitative indicators, namely Jaccard index, sensitivity and specificity, are adopted to define the evaluation criteria. The numerical range of Jaccard is between [0, 1]. When there is no target area of interest in the output picture, the value of Jaccard is 0, and the closer the value of Jaccard is to 1, the higher the correct rate, that is, the better the effect. Sensitivity, that is, the true positive rate, means the percentage of patients who are actually ill and are correctly diagnosed according to some diagnostic criteria. Specificity, that is, true negative rate, means the percentage of patients who are actually disease-free but are actually diagnosed disease-free according to some diagnostic criteria.

Figure 23.3 shows the diagnosis statistical results of the proposed algorithm.

It can be seen that the improved FCM algorithm proposed in this paper has a high diagnostic accuracy and shows stable experimental results in different experimental groups. The algorithm in this paper not only shows high segmentation accuracy, but also is stable and robust. The superiority of the improved FCM algorithm is verified.

In order to quantitatively analyze the results of image segmentation, segmentation coefficients can be used to evaluate the segmentation results. In the case of ideal deterministic segmentation, the value of membership matrix is two discrete values [0, 1]. At this time, the value of segmentation coefficient is 1, so it is impossible to have such ideal data in actual situation. However, the closer the segmentation coefficient is to 1, the better the segmentation result is.

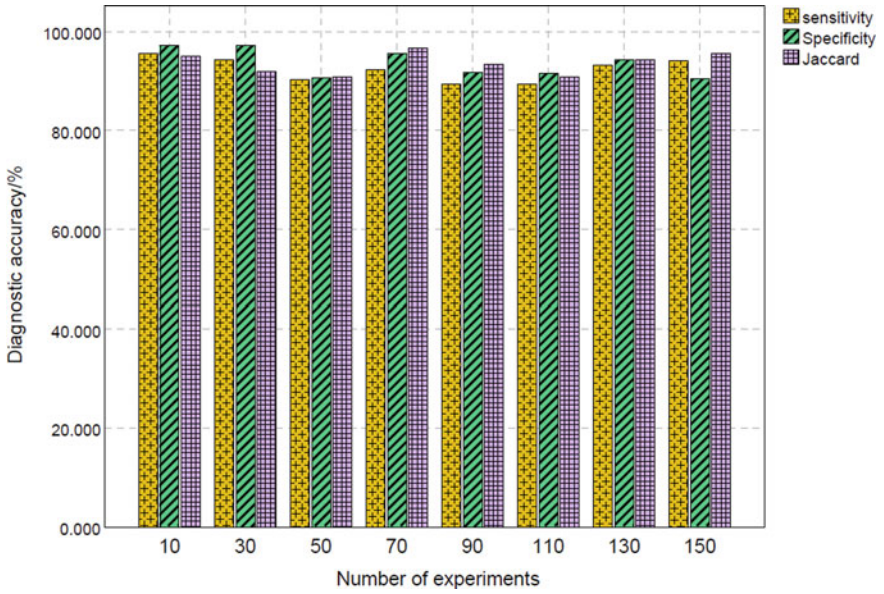


Fig. 23.3 Diagnostic statistical results

Figure 23.4 below lists the evaluation of the segmentation effect of artificial images and MRI images.

The above results show that the standard FCM has poor ability to process noisy images, and the segmentation result is not ideal under the influence of noise, and the convergence process becomes longer. The improved FCM algorithm improves the anti-noise ability by fusing spatial filters, which is close to the effect of standard FCM method in processing noise-free images, and the convergence process is only slightly increased. Experimental data show the effectiveness of the improved FCM algorithm. Experimental results and evaluation functions show that this method has achieved good segmentation and anti-noise effects.

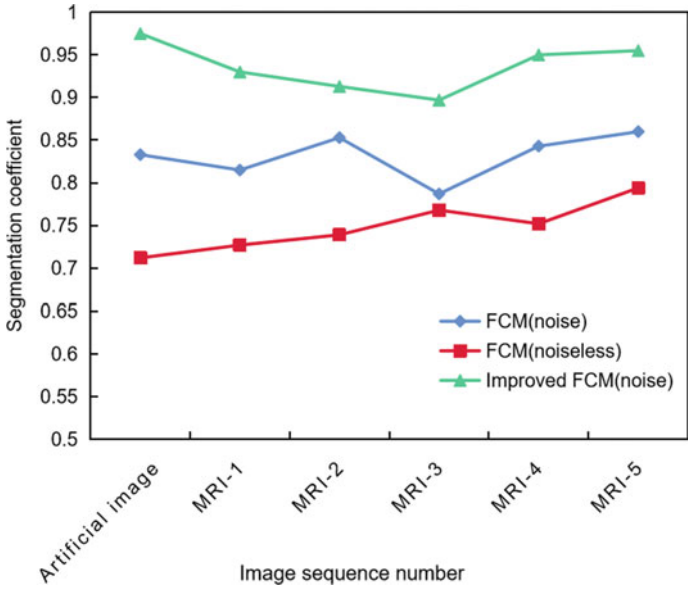


Fig. 23.4 Image segmentation effect

23.4 Conclusion

The topic selected in this paper is the simulation of intelligent image processing model based on machine learning algorithm. Combining machine learning with fuzzy theory, a novel MRI image segmentation algorithm is obtained, which solves the limitations of the existing segmentation algorithm and is much better than the traditional segmentation algorithm in segmentation accuracy and time. The simulation results show that the improved FCM algorithm proposed in this paper has a high diagnostic accuracy and shows stable experimental results in different experimental groups. The algorithm in this paper not only shows high segmentation accuracy, but also is stable and robust. The improved FCM algorithm improves the anti-noise ability by fusing spatial filters, which is close to the effect of standard FCM method in processing noise-free images, and the convergence process is only slightly increased. This method has achieved good segmentation and anti-noise effect.

Acknowledgements Foundation Item: Education and Teaching Reform Research Foundation of Xianyang Normal University (2019Y013).

References

1. Zhang, M., Wang, J., Pechauer, A.D., et al.: Advanced image processing for optical coherence tomographic angiography of macular diseases. Biomed. Opt. Express **6**(12), 4661–4675 (2015)
2. Osher, S., Shi, Z., Wei, Z.: Low dimensional manifold model for image processing. SIAM J. Imag. Sci. **10**(4), 1669–1690 (2017)
3. Couprie, M., Bezerra, F.N., Bertrand, G.: Topological operators for grayscale image processing. J. Electr. Imag. **10**(4), 1003–1015 (2015)
4. Jain, S., Jagtap, V., Pise, N.: Computer aided melanoma skin cancer detection using image processing. Proced. Comput. Sci. **48**, 735–740 (2015)
5. Erden, F., Velipasalar, S., Alkar, A.Z., et al.: Sensors in assisted living: a survey of signal and image processing methods. IEEE Signal Process. Magazine **33**(2), 36–44 (2016)
6. Mullapudi, R.T., Adams, A., Sharlet, D., et al.: Automatically scheduling halide image processing pipelines. ACM Trans. Graph. **35**(4), 83 (2016)
7. Hegarty, J., Daly, R., Devito, Z., et al.: Rigel: flexible multi-rate image processing hardware. ACM Trans. Graph. **35**(4), 85 (2016)
8. Pretorius, A.J., Zhou, Y., Ruddle, R.A.: Visual parameter optimisation for biomedical image processing. BMC Bioinform. **16**(11), S9 (2015)
9. Vaksman, G., Zibulevsky, M., Elad, M.: Patch-ordering as a regularization for inverse problems in image processing. SIAM J. Imag. Sci. **9**(1), 287–319 (2016)
10. Kalafi, E.Y., Tan, W.B., Town, C.: Automated identification of monogeneans using digital image processing and K-nearest neighbour approaches. BMC Bioinform. **17**(19), 259–266 (2016)
11. Chen, Y., Xu, L.J., Gong, Z.L., et al.: Application of image processing and analysis techniques in molecular imaging. J. Shanghai Jiaotong Univ. Med. Sci. **35**(4), 605–610 (2015)
12. Cui, X.Y., Gui, Z.G., Zhang, Q., et al.: Learning-based artifact removal via image decomposition for low-dose CT image processing. IEEE Trans. Nucl. Sci. **63**(3), 1 (2016)
13. Abry, P., Roux, S.G., Wendt, H., et al.: Multiscale anisotropic texture analysis and classification of photographic prints: art scholarship meets image processing algorithms. IEEE Signal Process. Mag. **32**(4), 18–27 (2015)
14. Nejati, H., Pomponiu, V., Do, T.T., et al.: Smartphone and mobile image processing for assisted living: health-monitoring apps powered by advanced mobile imaging algorithms. IEEE Signal Process. Mag. **33**(4), 30–48 (2016)
15. Lin, C.F., Hsieh, K.H.: Straight-line motion control for autonomous land vehicles using ID image processing techniques. J Field Robot **15**(10), 537–549 (2015)

Part III
Health Systems and Security

Chapter 24
Design and Implementation of a Health Monitoring Management Platform Based on IoT and DL

Yineng Xiao

Abstract At present, there is still a big gap in the field of medical and health services in China. China has a large population, and per capita medical resources are relatively scarce; on the other hand, there are some problems in China's medical system, such as backward medical facilities and unbalanced distribution of resources, which make it impossible to effectively use medical resources. The establishment of an intelligent health management system with patients as the core is the key to the establishment of an intelligent health service system. In the process of establishing an intelligent health management platform, we should start with improving the health management of patients and their satisfaction, and take their economic burden into account economically. At present, our government has issued a series of policies to support the development of intelligent medicine. The intelligent health service platform is based on the Internet of Things (IoT) and deep learning (DL) technology and uses China's advanced medical resources to improve the medical quality and satisfaction of patients.

24.1 Introduction

In China, chronic diseases have become one of the main factors affecting the health of residents, especially common chronic diseases such as hypertension, diabetes and hyperlipidemia, which have become one of the major killers threatening the health of Chinese residents. The Health Monitoring Management Platform (HMMP) developed in this paper uses mobile terminal and IoT technology to achieve data transmission between sensor nodes and cloud computing platform, and uses big data mining methods to provide more and more perfect health knowledge. HMMP can

Y. Xiao (✉)
The Global Intellectual Property Institute, Nanjing University, Suzhou 215163, China
e-mail: xiaoyineng@pku.edu.cn

Advanced Institute of Information Technology, Peking University, Hangzhou 311200, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_24

run on a computer with a Windows 10 OS, a 4-core CPU, 4G memory or above, and Internet access.

Due to the increasing population aging, how to effectively provide long-term care services with limited resources to meet the needs of the elderly is emerging. To solve this problem, nursing homes are making efforts to adopt intelligent health of the IoT and artificial intelligence to improve the efficiency and sustainability of medical care. The personal health information management system proposed by Pravin Pawar supports the acquisition, storage, organization, integration and privacy sensitive retrieval of consumer health information [1]. Farooq Ali adopts induction and basic theoretical methods, because it is necessary to describe the stakeholder network in detail, especially relationship management [2]. Valerie Tang proposed a two-level response health analysis model to provide appropriate medical services in nursing homes under the medical IoT environment [3]. This paper combines cloud computing with big data analysis technology to achieve a personalized comprehensive health evaluation system.

The HMMP is a multi-level health monitoring and early warning management system developed with advanced technologies such as cloud computing, cloud storage, big data and artificial intelligence, with the IoT technology, big data processing and DL as the core. The platform collects data and processes data through mobile terminals and the IoT, realizing real-time data statistics, analysis, intelligent early warning management, services and other functions. On this basis, the platform adopts dynamic planning technology to conduct real-time monitoring of users' equipment operation and provides users with healthy life services through big data mining and analysis.

24.2 Functional Conception and Structural Design of HMMP

24.2.1 Health Monitoring

Health monitoring is to provide accurate and effective individualized medical and health management services for patients in areas with imperfect medical conditions. Through real-time monitoring and dynamic management of patients' daily activities, psychological status, sleep quality, eating habits, etc., timely and effective personalized health management can be provided for patients [4, 5]. Through the analysis of the patient's historical behavior and psychological state, we can assess the patient's current condition and physical condition and living state in the future. It can also conduct a comprehensive evaluation based on the patient's previous treatment effect, medication situation, side effects, etc., and then determine whether it is necessary to change the treatment plan and adjust the dosage. A real-time monitoring network based on Internet technology can help improve the health management effect through real-time monitoring of patients' physiological and psychological conditions [6, 7].

According to the analysis results of health monitoring data, it can be determined whether the patient is in a sub-health state or in the pre-disease stage; you can also provide personalized advice on their current physical conditions, such as increasing exercise time or adjusting their eating habits. Through the analysis of the patient's psychological state, doctors can also help them find out the potential psychological problems of the patient in time and give targeted treatment guidance [8].

24.2.2 Management Services

In order to facilitate the use of patients, the system uses voice recognition, image recognition and other technologies based on artificial intelligence and designs a patient management system. The voice intelligent assistant can help patients with disease consultation and management [9, 10]. For example, when the patient's condition is unstable, the doctor can ask him about the treatment plan, medication advice and other information through the voice assistant. For patients who are seriously ill and are not convenient to use voice assistant, doctors will also use their smartphones to carry out telemedicine for patients [11, 12]. The system also has an automatic early warning function, which can automatically remind medical staff to handle emergencies. Health data can be stored and transmitted in the cloud to view and manage health data anytime and anywhere. Patient health data records are sent to doctors or nurses in the form of short messages on the basis of cloud storage.

AI is based on DL processes big data and uses its feature learning and feature extraction functions to quickly identify, locate and track abnormal data and results. The information filled in by the user in the management system can be stored in a spreadsheet generated by the background system and provided to doctors or nurses in the background. Patients can download spreadsheets from the management platform and print them for doctors to consult. Because it is a voice recognition and image recognition function based on DL technology, the system can carry out diagnosis, treatment and rehabilitation guidance for a specific disease.

24.2.3 Functional Design

(1) Health monitoring

For the monitoring data of various organs and tissues in the human body, the health status can be adjusted in time according to the development of the disease, and real-time treatment plans can be provided for patients.

(2) Auxiliary diagnosis

To assist in the analysis of the patient's health, we can determine whether the patient has a disease by analyzing the monitoring data. If there is a disease, we can take treatment measures.

(3) Health assessment and intervention

To provide a comprehensive and objective health assessment report in the medical environment, the health assessment model can be used to judge whether a person has a serious disease and whether he needs timely treatment to maintain the stability of his condition.

(4) Telemedical monitoring

For the elderly, medical monitoring needs remote data collection and transmission. Due to the limitation of their physical functions, it is difficult to take care of their daily life 24 h a day.

(5) Remote management center

The management center can monitor the patient's physical condition and health status in real time, and adjust the treatment plan in time.

24.2.4 Platform Construction

The information collected by the IoT is used to judge the patient's status, and the information is transmitted to the platform's system server. Then, manual analysis is used to determine whether there are patients who need to be monitored. Before the patient enters the hospital or after the patient leaves the hospital, the doctor can start to monitor and analyze the data to judge their health status. According to the needs of different patients, the platform has designed different versions. After the user leaves the hospital, the platform will send the information to the medical institutions and upload their health data to the platform. When users enter the hospital again, the platform will create new patient information records in the database, and use artificial intelligence algorithms to analyze and judge whether there are patients who need health monitoring.

After the maturity of big data and cloud computing technology, the platform will use AI technology to analyze health data and determine whether there are patients who need health monitoring and their specific conditions. If a patient needs further treatment or is transferred to another hospital, the intelligent health management service platform will send information to the patient. In the whole data analysis process, a large number of people are required to participate in judgment and operation. When the doctor judges whether the patient needs to be transferred or treated for other diseases, the platform will provide the personal information and relevant data entered by the patient or his family members as the judgment basis. When the doctor decides whether further treatment is needed, the patient will send information to the doctor through the platform to apply for transfer. If the user is a doctor or his family member, the platform will also provide him with real-time monitoring data reports.

24.3 HMMP Design Implementation and Function Inspection

24.3.1 HMMP

Based on convenience, real time, ease of use, scalability and security principles, this paper designs and completes the code, carrier service framework model, functional module design, system framework design and other aspects of the regional health service platform based on the IoT.

In order to ensure the quality of the system, the following principles will be followed in the design, development, deployment and operation of the management plan.

Security issues. Through the system carrier related to security settings, the goal of security is finally achieved. Ensure that unauthorized users will not invade the system, will not cause data loss, and will not be illegally obtained, tampered with, and will not confirm the identity of users and recipients in the transmission process.

Accuracy. Ensure accurate understanding of business needs through detailed systematic investigation and analysis, the accurate detection of the system and the daily management of standard items ensure the processing accuracy of the system. At the same time, during the platform's design and implementation, various inspections and audit supervision are provided to ensure the processing accuracy of the system.

Expansion. The development platform of the system should be able to realize simple expansion, real-time monitoring of the platform's working conditions, and automatic load balancing to ensure the parallel processing of many terminals. There is no need to make too many changes to the architecture, and the design carrier can improve the application level, increase the demand and scale.

24.3.2 Health Management Indicators

Body mass index (BMI) is a good measurement standard. By measuring the height and weight of current users, BMI can be quickly obtained. BMI data is an important basis for the current obesity level of users, and accordingly, relevant health news is recommended to users. Physical health mainly includes fat rate, water rate, muscle rate, bone lean rate, basic metabolism, etc. Fat ratio refers to the ratio of total weight to fat. Fat is the main fatty acid and can also be stored. Essential fatty acids are the fats that the human body needs for survival and reproduction.

Fat storage refers to the accumulation in adipose tissue. Generally speaking, if the body fat of a man is more than 25% and that of a woman is more than 33%, it is considered obese. Because the water content in the body accounts for a lot, the determination of water content can also reflect a person's weight and weight. Muscle rate refers to the proportion of muscles inside the body, that is, the weight of the body and the percentage of the body weight. The average female is about 27% under normal

circumstances. Bone rate, also known as bone density, is an important indicator to measure the degree of osteoporosis. Basic metabolism refers to the minimum energy required by the body to maintain life in the digestion stage under the condition of natural temperature and inactivity.

Personalized recommendation refers to personalized recommendation of users based on their shopping habits and preferences. With the rapid development of e-commerce, the number and variety of products are also increasing rapidly, and consumers need to spend a lot of time and energy to find the products they need. Such browsing of unimportant information and products will inevitably be swallowed up by customers engulfed by information overload. Aiming at the above problems, a personalized recommendation system is proposed. In the health management system, there are many "commodities," such as recipes, dishes, exercise plans, etc. Therefore, personalized recommendations can quickly find out the information that is suitable for you, thus solving the problem of finding massive information for users.

BMI is the most important user feature in the health management system. It divides the degree of obesity into five categories: thin, thin, normal, obese and extremely obese. According to the user's gender, age and other characteristics, upload them to the background cloud database and feed them back to mobile phones and other terminal devices.

24.3.3 Platform Function Module Design

The health service portal module, health management system and health mobile terminal (according to the division method of mobile IoT) shown in Fig. 24.1 are several functional modules of the health service carrier.

The remote health management subsystem and the office intranet health management network with health management tracking, management service/information, health consulting center, appointment management and other functions are its core business modules, as shown in Fig. 24.2. The medical service network is connected with the medical service network, realizing the data exchange between the internal and external networks of 271.

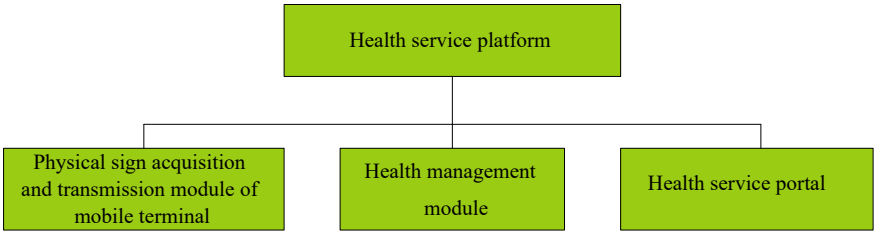


Fig. 24.1 Schematic diagram of design carrier function module

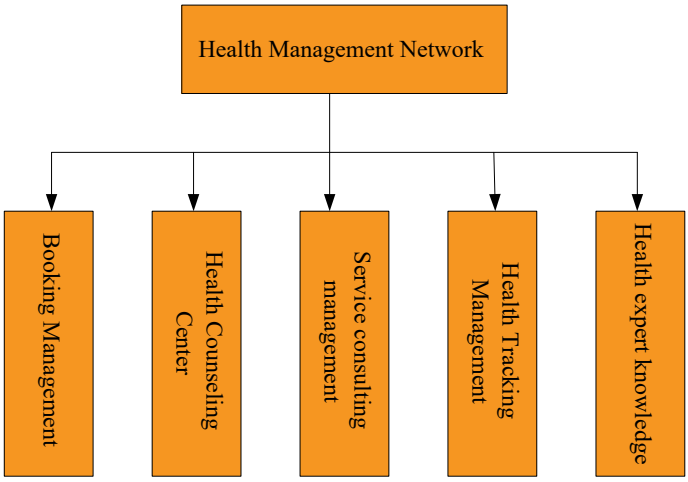


Fig. 24.2 Functional module diagram of health management network

24.3.4 Research on DL Algorithm

As time goes by, more and more image recognition technologies are used in our daily life. As an effective image recognition method, deep learning has been widely used in image recognition. This paper first describes the image recognition process in detail, and describes it in detail. Then, the artificial neural network (ANN) is introduced in detail, and the neurons that constitute ANN and the commonly used ANN are introduced. Finally, this paper gives a brief introduction to deep learning and elaborates on the commonly used deep learning algorithms.

On this basis, ANN uses the processing function to process the input signal and uses the activation function to calculate the input signal, so as to obtain the output of the neuron. Generally, the processing function usually adopts the summation method. There are many activation functions, and the following one is commonly used.

As a new technology based on traditional neural network, DL has been widely used in image recognition, speech recognition, automatic driving, etc. Deep learning is to imitate the connection between neurons in the human brain, and on this basis, simulate the processing of visual, auditory and other input information in the brain. The depth of deep learning refers to that there are many nonlinear changes at different levels in deep learning. Deep learning can transform the extracted low-level features into high-level features, so as to achieve hierarchical feature extraction.

Sigmoid function: Sigmoid function is a monotone incremental function, its value range is [0, 1], so the function value will not diverge, and the solution is simple. However, its disadvantage is that the closer the function value is to 1, the smaller its gradient will be.

A function of Sigmoid is as follows:

$$W(a) = \frac{1}{1 + c^{-a}} \tag{24.1}$$

The derivative of the sigmoid function is as follows:

$$W'(a) = \frac{c^{-a}}{(1 + c^{-a})^2} = W(a)(1 - W(a)) \tag{24.2}$$

Assuming that there is one neuron in layer j and one neuron in layer $j-1$, layer j can be represented by a vector x_1, while in layer j, the offset of neurons can also be represented by a vector called bias vector. The weight values of layer j and layer $j-1$ can be represented by a matrix called weight matrix. The weight of the network is composed of the weight space of the network and the offset space. Then, the first output of the neural network is the following:

$$x_j = \phi\left\{\left[\sum\nolimits_r s_j x_{j-1} + y_j\right]\right\} \tag{24.3}$$

The forward propagation of the neural network is to change the input signal step by step and nonlinearly at the input layer, and then transfer the output signal to the neuron at the next layer to obtain the output of the neural network.

In deep learning, Softmax layer is usually used as the output layer of neural network. The Softmax layer will output an r-dimension vector to represent the probability of each category to which the input samples belong. Assume that the input of Softmax layer is c, the output is $C = g(c)$, and c is n-dimensional column vector. The mathematical expression of Softmax classifier is as follows:

$$C = g\left(c^i\right) = \frac{\exp\left(c^i\right)}{\sum_{i=1}^{r} \exp\left(c^i\right)} \tag{24.4}$$

So:

$$q(b = i|a) = g\left(c^i\right) = \frac{\exp\left(c^i\right)}{\sum_{i=1}^{r} \exp\left(c^i\right)} \tag{24.5}$$

From here, we can see that the Softmax layer normalizes its input, and the total probability of each sample is 1.

The loss function is used to measure the difference between the actual marked value and the predicted value in the input samples predicted by the neural network. It is recorded as $O(\phi)$, and ϕ represents the current space of the current neural network. When training the neural network, the value of the loss function must be reduced as much as possible. In depth learning, many different loss functions are used. Softmax loss function is used in this paper. Its mathematical formula is as follows:

$$O(\phi) = -\log\left[g\left(c^n\right)\right] = -\log\left[\frac{\exp(c^n)}{\sum_{i=1}^{r}\exp\left(c^i\right)}\right] \tag{24.6}$$

24.4 Experimental Results and Discussion of HMMP

Testing is an important link in the software life cycle. It is a stable period for product release and submission to the end user. Some functions of the system are to integrate existing systems. The performance optimization and stability of the system will have a great impact on existing systems. On the basis of actual system work, by comparing the requirements of the system, we can check whether a complete software can establish an accurate connection with the system, and obtain that the rules in the software development protocol are inconsistent or inaccurate with the design of the system file.

See Table 24.1 for main test examples. In order to verify the performance of the system, it is necessary to make a detailed description of it, so as to determine all its functions. When testing individual sub-functions, classify the selectable items tested and the data entered, and test each type of data separately. The test design adopts equivalence class division, random test and error guessing test, and takes G002 family as an example. See Table 24.1 for test emphasis.

After continuous system repair and optimization, the whole system has been improved and tried out for 4 months. Now the system has perfect functions, friendly interface, stable operation, and users are satisfied.

Table 24.1 System test trial list

Case number	CASE name	Test type	Requirements traceability
A001	Rationality of the user interface	Usability test	Ease of use requirements
A002	Menu interface, shortcut key interface	Usability test	Ease of use requirements
A003	User authorization mechanism	Security test	Security requirements
G001	System login	Functional test	Functional requirements
G002	Modification of resident files	Functional test	Functional requirements
G003	Health monitoring data input	Functional test	Functional requirements
G004	Health data query	Functional test	Functional requirements

24.5 Conclusion

The application of IoT and DL technology in health management holds great promise. It not only enhances disease detection, prevention and treatment but also empowers patients to improve their quality of life and fosters family well-being. The Health Management and Monitoring Platform (HMMP), based on IoT and DL, is a comprehensive information system that leverages comprehensive analysis of multiple data and methods to obtain valuable information sources. Through this system, patients can access their health status, medical expenses and other relevant information via mobile phones, computers or other terminal devices. Likewise, doctors can make professional diagnoses using the platform. The system incorporates a range of user-defined diagnostic indicators and scoring standards, including parameters like heart rate, blood pressure, blood oxygen level, as well as auxiliary indicators such as electrocardiogram and B-ultrasound. While this paper primarily focuses on the security of the platform in terms of safeguard measures, it does not address the various aspects of ensuring its normal operation in specific operational scenarios. Further research is required to explore these aspects thoroughly.

References

1. Pawar, P., Parolia, N., Shinde, S., Edoh, T.O., Singh, M.: eHealthChain: a blockchain-based personal health information management system. Ann. des Télécommun. **77**(1–2), 33–45 (2022)
2. Ali, F., Haapasalo, H., Tampio, K.P., Haapasalo, H.: Analysing the challenges in stakeholder relationship management in the healthcare process: a social network perspective. Int. J. Netw. Virtual Organ. **26**(1/2), 125–156 (2022)
3. Tang, V., Lam, H.Y., Wu, C.H., Ho, G.T.S.: A two-echelon responsive health analytic model for triggering care plan revision in geriatric care management. J. Organ. End User Comput. **34**(4), 1–29 (2022)
4. Mehdizadeh, N., Farzaneh, N.: An evidence theory based approach in detecting malicious controller in the multi-controller software-defined IoT network. Ad Hoc Sens. Wirel. Netw. **51**(4), 235–260 (2022)
5. Radha, D., Kavitha, M.G.: DL enabled privacy preserving techniques for intrusion detection systems in the industrial IoT. Ad Hoc Sens. Wirel. Netw. **52**(3–4), 223–247 (2022)
6. Papanagnou, C.I.: Measuring and eliminating the bullwhip in closed loop supply chains using control theory and IoT. Ann. Oper. Res. **310**(1), 153–170 (2022)
7. Alattas, K., Wu, Q.: A framework to evaluate the barriers for adopting the internet of medical things using the extended generalized TODIM method under the hesitant fuzzy environment. Appl. Intell. **52**(12), 13345–13363 (2022)
8. Wu, J.M.T., Li, Z., Srivastava, G., Yun, U., Lin, J.C.W.: Analytics of high average-utility patterns in the industrial IoT. Appl. Intell. **52**(6), 6450–6463 (2022)
9. Corno, F., Russis, L.D., Roffarello, A.M.: How do end-users program the IoT? Behav. Inf. Technol. **41**(9), 1865–1887 (2022)
10. Kim, J., Park, E.: Understanding social resistance to determine the future of IoT (IoT) services. Behav. Inf. Technol. **41**(3), 547–557 (2022)

11. Almagrabi, A.O., Ali, R., Alghazzawi, D., AlBarakati, A., Khurshaid, T.: A reinforcement learning-based framework for crowdsourcing in massive health care IoT. Big Data **10**(2), 161–170 (2022)
12. Chang, V., Kacsuk, P., Wills, G., Behringer, R.: Call for special issue papers: big data and the internet-of-things in complex information systems: selections from IoTBDS 2022 and COMPLEXIS 2022: deadline for manuscript submission: September 30, 2022. Big Data **10**(2), 93–94 (2022)

Chapter 25
Design of Hospital Equipment Information Management System Based on Computer Vision Technology

Benhai Yu and Gaofeng Xia

Abstract The information management of medical equipment should reflect two movement states of medical equipment, namely material movement state and value movement state, so that managers, decision makers, and users can clearly grasp the movement process of the equipment. In this paper, the hospital equipment information management system based on browser/server (B/S) is realized by using Java as the software development language and adopting the service-oriented architecture design. And the computer vision technology is used to realize the fault detection and alarm of hardware equipment in the computer room. The system has the functions of inputting, inquiring, and modifying the basic information of medical equipment. Based on computer vision technology, this paper adopts the faster region-based convolutional neural networks (R-CNN) algorithm in the two-step detection algorithm to study the infrared image detection of hospital equipment. The abnormal area of hospital equipment and the equipment detection network under infrared images based on faster R-CNN are established. The system test shows that the average response time of the system is < 4 s when < 100 people log in to the system at the same time, which shows that the system runs well and has stable performance, and can meet the requirements of simultaneous online operation of multiple people. The infrared detection method of hospital equipment based on faster R-CNN has a high detection accuracy in the detection of abnormal areas in infrared images of hospital equipment.

25.1 Introduction

With the continuous improvement of modern science and technology in China, medical equipment technology has also been continuously broken through, playing an increasingly important role in medical services. At the same time, the scientific

B. Yu · G. Xia (✉)
School of Economics and Management, Shanghai Institute of Technology, Shanghai 200235, China
e-mail: xgfwust@163.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_25

and technological content of medical equipment is getting higher and higher, and the corresponding equipment management is becoming more and more complicated. Medical equipment is an instrument or equipment used by hospitals to diagnose, prevent, monitor, treat, or alleviate diseases. It has become the most basic production factor for medical institutions to carry out diagnosis and treatment, scientific research and teaching, and is also the basic condition for improving the level of medical science and technology [1, 2]. As the assets of the hospital, medical equipment assists the clinical completion of medical diagnosis and treatment, and unified information management helps to ensure the safety and effectiveness of clinical use [3, 4].

Although with the popularization of information technology, the hospital information management system has been widely promoted, but the concept of the corresponding equipment management information system has not been well promoted. Due to the technical blockade of equipment providers, the number of engineers and technicians in many hospitals in the current period is too small, and the technology is not comprehensive, resulting in the situation that many medical equipment can only be repaired outside once it fails [5]. Because hospital engineers do not know much about the fault itself, the maintenance cost is usually determined by the repairer of the equipment, and the cost is imaginable. Therefore, it is very necessary to formulate a set of practical equipment use supervision system, whether it is to improve the sense of responsibility of supervisors or to urge equipment users to develop the habit of caring for equipment [6].

Medical equipment is the material guarantee of the medical level of modern hospitals and the modern means of diagnosis, treatment, and rehabilitation [7]. With the progress of modern science and technology, medical equipment is becoming more and more complicated and modernized, so it is imperative to implement information management. At the same time, medical equipment information management is also an organic part of hospital information system and an important means for hospitals to improve efficiency and ensure medical level. In this paper, the hospital equipment information management system based on browser/server (B/S) is realized by using Java as the software development language and adopting the service-oriented architecture design. And the computer vision technology is used to realize the fault detection and alarm of hardware equipment in the computer room.

25.2 Research Method

25.2.1 System Structure Design

The information management of medical equipment should reflect two movement states of medical equipment, namely material movement state and value movement state, so that managers, decision makers, and users can clearly grasp the movement process of the equipment. At the same time, medical equipment information

management should cover all aspects of medical equipment management, so as to achieve comprehensive and dynamic management. The implementation of information management will take time. To achieve this goal perfectly, we should not only establish a good hospital information system for the information management of medical equipment, but also require relevant managers to have professional technical ability. In the development and normal operation of all databases that have appeared so far, the "client–server" system mode plays a leading role [8].

The main tasks of medical equipment management include equipment procurement, equipment installation and use, equipment maintenance, and equipment scrapping. Each of them is composed of many small tasks, such as equipment procurement involves the collection of equipment procurement information, the quotation of equipment manufacturers, the performance of equipment, and the signing of contracts in the process of equipment procurement, so effectively extracting the main business is the key to development.

In the system design industry specification, there are national standards that put forward specific requirements for the overall design of the system, which also requires that the overall design must follow the relevant principles, and the overall design of the hospital equipment information management system follows the following principles:

1. Cost control: In the process of system design and development, we should follow the principle of cost control, use the most cost-effective design scheme, and purchase the most cost-effective software and hardware configuration.
2. Man–machine friendliness: The design of the system should be simple to use and operate, and the man–machine interface should be harmonious.
3. Efficient performance: After the system is designed and put into use, it does not consume too much system resources to ensure that the system can run smoothly on a general configuration computer [9].

The traditional client–server (C/S) mode needs to be debugged after the client installs the software before it can be put into use. Later, relying on the development of the Internet, it gradually developed the server and browser mode, that is, the B/S structure operation mode [10, 11]. Compared with the traditional C/S mode, the workload of the client is greatly reduced after adopting the B/S mode, which has more considerable economic benefits, and realizes giving different permissions to multiple groups of users in different places. Even under the premise of different users, data security is still guaranteed, and there is no need for secondary development in this process.

As the most commonly used design language in network system design at present, Java should be chosen as the design language for developing hospital equipment information management system. Java language has the advantages of simple and easy use, supporting object-oriented design, high reliability and security, and supporting multithreading.

The purpose of the study is to improve the management efficiency and level of medical equipment, manage medical equipment in a unified and information-based way, and further ensure the safety and effectiveness of medical equipment [12].

Therefore, we designed a hospital equipment information management system based on B/S structure. The system has the functions of inputting, inquiring, and modifying the basic information of medical equipment. The overall architecture design of the system is shown in Fig. 25.1.

Hospital equipment information management system is a commercial application software with hospital assets and maintenance management as the core, which meets the requirements of simple data manipulation and strong transaction processing ability, and the security and integrity of data are also guaranteed. All data are saved on the server, and data can be manipulated and processed through corresponding applications. This system chooses Windows XP as the operating system and is compatible with Windows 10, Windows Vista, and other operating systems. The network adopts Ethernet technology, and the transmission protocol uses TCP/IP and Internet.

Data is mainly divided into configuration data and operation data, in which configuration data is basically data with little change, which is used as the operation parameters of each component of the system; operation data refers to the data generated during the operation of the system, which covers four forms of information data requested for access: Product configuration, file, database, and cache, in which the data of product configuration, file and database are backed up on the ground regardless of whether the system is running or shutting down, while the cached data belongs to memory data and only exists at runtime.

The main functions of medical equipment information management system include the following aspects: equipment application management, equipment plan management, equipment contract management, equipment acceptance management, equipment fixed assets management, equipment use management, equipment measurement management, equipment maintenance management, and system setting. Its functional structure is shown in Fig. 25.2.

25.2.2 Implementation of Key Technologies of the System

Computer vision is a science that endows machines with the ability to see and think. It refers to the application of various imaging systems such as aerial camera and computer to identify, track and detect targets instead of human eyes, and attempts to establish an artificial intelligence system that can obtain “information” from images or multidimensional data. At present, in the field of intelligent warehousing, materials management is mostly based on RFID technology. In the process of warehouse management, warehouse procurement, quality verification of warehousing and warehousing, ledger entry, and other work are all completed by manual identification RFID tag pasting and entry, and there are many types and models of warehouse materials and products.

Aiming at the complexity and low efficiency of the existing hospital equipment information management system, a visual inspection system based on faster-CNN is proposed. The computer vision detection algorithm based on deep convective

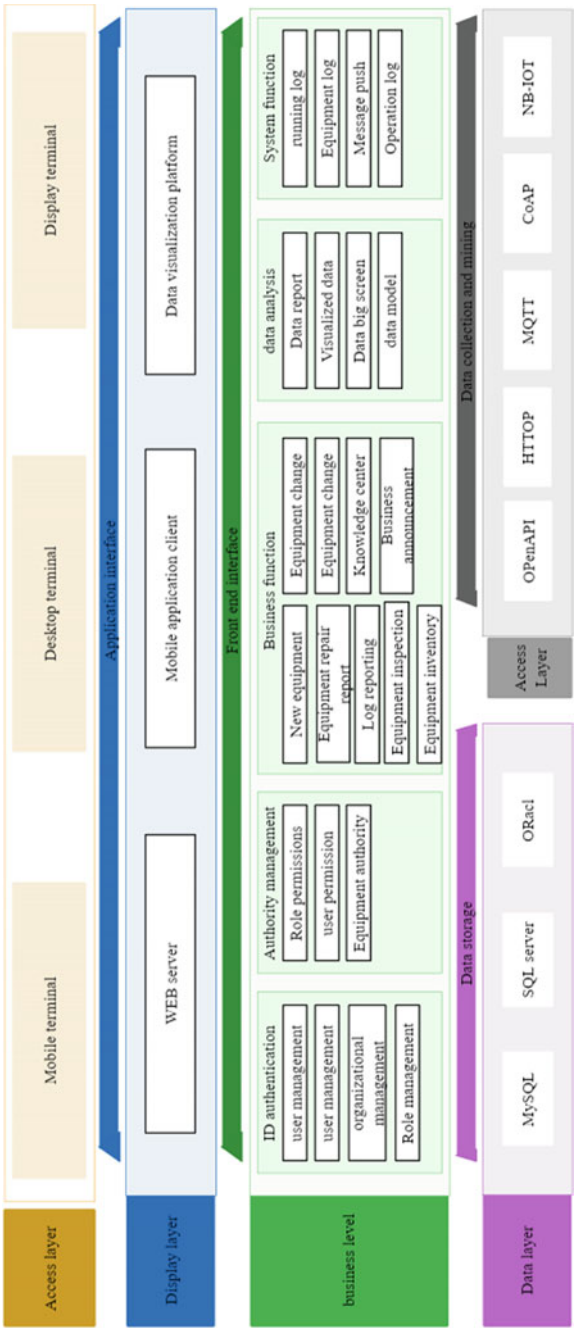


Fig. 25.1 Overall system architecture

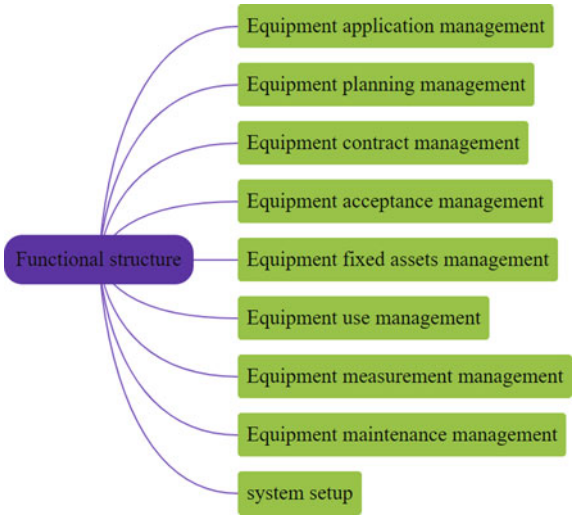


Fig. 25.2 Functional structure

network (DCN) makes it possible to automatically detect hospital equipment faults according to the infrared detection image characteristics of hospital equipment.

When testing hospital equipment, if it is defect-free, the heat dissipation will generally spread evenly on the surface of the equipment; once there is a defect in the internal or surface connector of the equipment, the heat dissipation will be uneven or abnormal, and the surface will be overheated or undercooled, which will be reflected on the infrared thermal imager which can adaptively adjust the imaging pixels according to the ambient temperature. Because the environment in the equipment room, the position of the robot and the angle of the camera will have a certain impact on the acquisition of video images, which increases the difficulty of image processing and reduces the detection efficiency and accuracy. In order to solve the above problems, this paper develops an intelligent inspection system based on the contour and color recognition of equipment indicator lights.

In recent years, the computer vision detection algorithm based on DCN has made breakthrough progress again and again. Compared with many manual feature extraction methods in traditional computer vision methods, it can extract the detailed image features of the detected object better. Compared with traditional feature extraction methods, DCN method has absolute advantages in the sub-field of computer vision target detection and has become the mainstream method. In the application of infrared detection of hospital equipment, according to the existing image feature judgment method in infrared detection of hospital equipment, the mainstream target detection algorithm of computer vision based on deep convolution neural network will be able to automatically locate and identify the thermal abnormal area of hospital equipment and typical hospital equipment categories.

In this paper, the faster R-CNN algorithm, which is one of the two-step detection algorithms, is used to study the infrared image detection of hospital equipment. Set

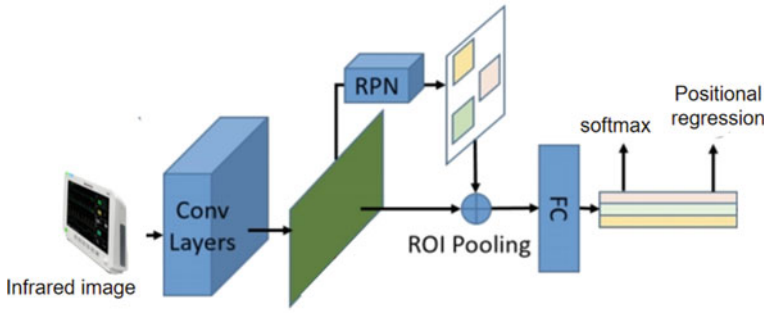


Fig. 25.3 Frame diagram of faster R-CNN detection network

up a hospital equipment abnormal area and equipment detection network based on infrared images based on faster R-CNN, and its main structure is shown in Fig. 25.3.

Firstly, the trained infrared data is transmitted to ResNet50 shared feature extraction network for feature extraction. Then, the first branch continues to introduce region proposal network (RPN) and carries out foreground and background classification and position regression training to obtain regional suggestions and output them to the ROI average pool layer, and the characteristic layer of the second branch is also introduced to the ROI average pool layer; finally, the output of ROI average pool layer is fully connected, and the final detection classification and target regression output are obtained by convolution. The incoming target function is propagated forward and backward, and the end-to-end training is carried out to get the final result.

Suppose that the input image is convolved to get the feature map, and the original input training image is reduced by 18 times to get the feature layer, and the number of anchor frames is 6137. The candidate frames are continuously translated and scaled to get the prediction frame, that is, frame regression, and the objective function is trained to make the learned translation and scaling parameters approach the labeling frame continuously.

$$\text{Loss} = \sum_{i}^{N} \left(t_*^i - W_*^i * h\left(P^i\right)\right)^2 \tag{25.1}$$

$h\left(P^i\right)$ is the feature vector of the candidate frame on the feature map, W_*^i is the transformation parameter from the learned candidate frame to the prediction frame, t_*^i is the offset between the labeled frame and the candidate box computing, and the final objective function is to make the prediction frame and the labeled frame as close as possible to the objective function.

RPN position regression loss is fine-tuned for anchor frame position training, target classification loss is trained for ROI category, and target position regression loss is fine-tuned for final target position training. Training loss function:

$$L(\{p_i\},\{u_i\}) = \frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}\left(p_i, p_*^i\right) + \lambda\frac{1}{N_{\text{reg}}}\sum_i p_*^i L_{\text{reg}}\left(t_i, t_*^i\right) \tag{25.2}$$

where p^i is the probability that the candidate box is the target, and p_*^i is the prediction label. t_*^i is the deviation of the label box from the candidate box, L_{cls} is the classification loss function, L_{reg} is the regression loss function, N_{cls} is the foreground second classification, and N_{reg} is the foreground second classification.

25.3 System Test

Relying on the mainstream system use environment, the external test is carried out by using the black box test method, and the main functions of the system are operated and the results are verified, so as to find the loopholes in the system and repair them. Through the response speed, reliability, security, concurrency and other system performance tests, to ensure the smooth operation of the follow-up system.

Black-box testing is based on the software requirement specification, and the tester takes the software user as the starting point and only tests whether the system function can respond correctly, without considering the internal logic of the system. Only perform relevant operations, complete the input and compare the actual output result with the expected result to judge the correctness of the system function.

According to the experimental requirements of this system performance test, the multi-person response performance of the system is measured by Loadrunner software, and the results are shown in Fig. 25.4.

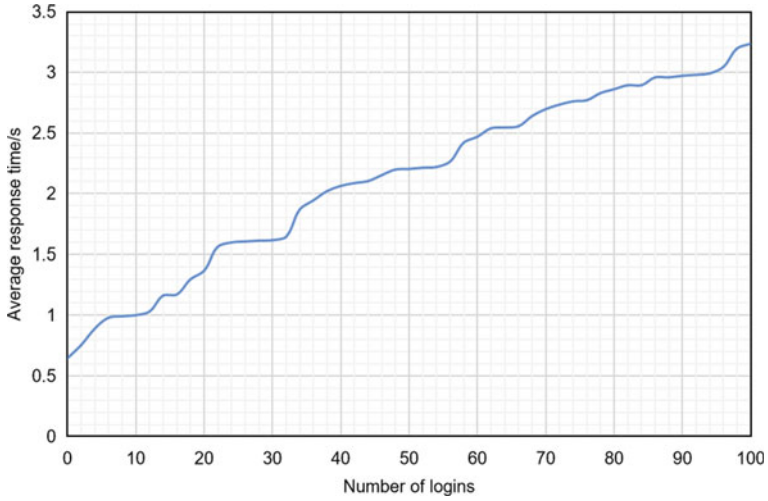


Fig. 25.4 Performance test results

As can be seen from the figure, with the increase of the login number, the average response time of the system also increases slightly. When < 100 people log in to the system at the same time, the average response time of the system is < 4 s, which shows that the system runs well and has stable performance, and can meet the requirements of simultaneous online operation by many people. The system automatically judges the performance of the system through the test tool, completes the test, obtains the good response characteristics and throughput of the system, and finds that the system performance meets the user's requirements.

Read the verification data in batches, verify it with the training model, and verify it in verification set, test set, and mixed set in turn. Among them, the mixed set includes 85 randomly selected verification set pictures, 85 test set pictures, and 45 infrared images of hospital equipment without abnormality. The verification results of the equipment and anomaly detection model based on faster R-CNN are shown in Fig. 25.5.

When the number of model classification increases, the classification accuracy drops obviously. Because the test and verification data are limited, and a large part of the collected infrared detection images of typical hospital equipment are only a part of the field of view of a typical hospital equipment, most of the tested images only detect abnormal areas.

The results show that the infrared detection method of hospital equipment based on faster R-CNN has a high detection accuracy in the detection of abnormal areas of hospital equipment infrared images and also has a good effect in the detection of abnormal areas and equipment categories of hospital equipment infrared images.

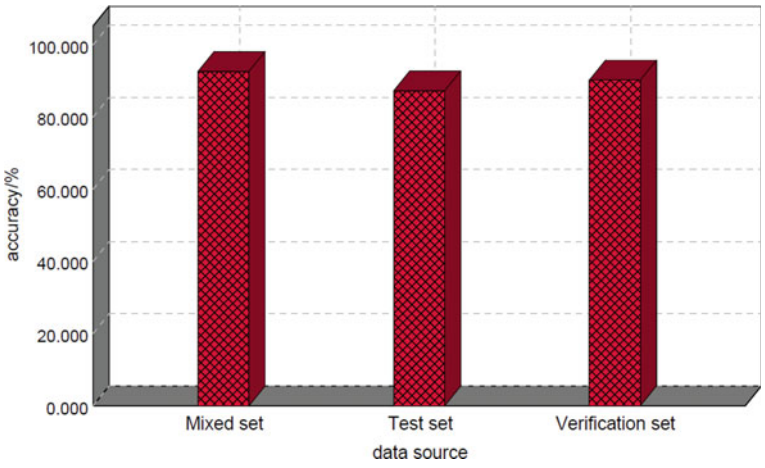


Fig. 25.5 Test data

25.4 Conclusion

In this paper, the hospital equipment information management system based on B/S is realized by using Java as the software development language and adopting the service-oriented architecture design. And the computer vision technology is used to realize the fault detection and alarm of hardware equipment in the equipment room. The system test shows that with the increase of login number, the average response time of the system also increases slightly. When < 100 people log in to the system at the same time, the average response time of the system is < 4 s, which shows that the system runs well and has stable performance, and can meet the requirements of simultaneous online operation by many people. The infrared detection method of hospital equipment based on faster R-CNN has a high detection accuracy in the detection of abnormal areas of hospital equipment infrared images and also has a good effect in the detection of abnormal areas and equipment categories of hospital equipment infrared images. The system lacks intelligent processing functions such as data mining. If the corresponding data mining algorithm can be introduced into the system, it can better provide reasonable data analysis and decision support for equipment managers.

Acknowledgements Research on health insurance performance evaluation model and coordination mechanism toward multi-subject interest balance in big data environment (71974131).

References

1. Zhou, J.H., Ren, J., Wu, W.S., et al.: Design and implementation of medical equipment integrated management system based on ad hoc network and Internet of Things. China Med. Equip. **12**(1), 5 (2015)
2. Liu, X.S., Li, Y.X., Wang, Y.J., et al.: Design and implementation of information management system for nurses in tertiary general hospitals. China Med. Equip. **33**(6), 6 (2018)
3. Xin, Y.P., Liang, J.: Design and construction of hospital bidding information system. China Med. Equip. **31**(03), 124–126 (2016)
4. Zhang, L.J., Jin, M., Zhao, D.X., et al.: Design and implementation of logistics management system in our hospital. China Med. Equip. **34**(12), 5 (2019)
5. Zhang, Y.J., Zuo, C.S., Chen, M.K., et al.: Design and application of a medical equipment maintenance management system. China Med. Equip. **33**(5), 4 (2018)
6. Hao, R., Zheng, J., Li, N.: Design and application of hospital archives electronic information management system based on internet plus platform. China Med. Equip. **18**(2), 4 (2021)
7. Zhao, J., Shang, S.L., Zhang, H., et al.: Hospital medical consumables information management system and key technology design. Mod. Sci. Instrum. **39**(2), 17–22 (2021)
8. Liu, B.Z., Liu, Z.G., Wang, X.W.: Design and implementation of mobile operating room information management system based on electronic medical records. J. Biomed. Eng. **32**(3), 7 (2015)
9. Wang, Y.D., Chen, G., Niu, W.X.: Hospital equipment information management and system architecture design. China Med. Equip. **13**(3), 4 (2016)
10. Shen, X.B., Zhou, X.J., Zhou, Q.L.: Design and implementation of hospital bed reservation closed-loop management information system. China Med. Equip. **031**(007), 90–101 (2016)

11. Yang, Y.Q., Zou, P.L., Wang, Y., Tu, X.C., Xiao, W.C., Wang, D.X.: Design and application of hospital equipment life cycle management system. China Health Qual. Manag. **24**(2), 124–126 (2017)
12. Zhang, X., Hu, X.B., Huo, Y.L.: Design of automatic blood collection management system. China Med. Equip. **036**(003), 119–123 (2021)

Chapter 26
Design and Implementation of an Internet of Things-Based Real-Time Five-Layer Security Surveillance System

Kamlesh Narwani, Fahad Liaquat, Asif Ali Laghari, Awais Khan Jumani, Junaid Jamshed, and Muhammad Ibrar

Abstract The Internet of Things (IoT) is one the major part of the fourth Industrial Revolution; it's the process to enable things to gather data and get controlled remotely from anywhere in the world through the Internet. IoT is changing the world so that impossibilities are converting into possibilities, e.g., self-driving cars, automated machine payments without human involvement, and remote surgeries. It's changing almost every domain. The phenomenon of IoT has been observed to have a pervasive impact across various domains. Despite its potential benefits in enabling remote control of devices, security has emerged as a persistent challenge for both businesses and households. This issue has been observed to be particularly pronounced in the context of goldsmith shops, where the implementation of IoT systems has been found to be associated with relatively lower levels of security.

Kamlesh Narwani, Fahad Liaquat, Awais Khan Jumani, Junaid Jamshed—These authors contributed equally to this paper

K. Narwani
School of Electronic Information and Communication, Huazhong University of Science Technology, Wuhan, China

Computer Science Department, DCK Campus, DHA SUFFA University, Karachi, Sindh, Pakistan

F. Liaquat · A. K. Jumani
Computer Science, ILMA University, Karachi, Sindh, Pakistan

A. A. Laghari (✉) · M. Ibrar
Software College, Shenyang Normal University, Shenyang, China
e-mail: asiflaghari@synu.edu.cn

A. K. Jumani
School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong, China

J. Jamshed
Shanghai University, Baoshan, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_26

Even large corporations with substantial financial resources have been observed to face significant security challenges in the context of IoT. In this research, a security system can only allow authenticated users and inform the owners if any unknown intruders try to enter the place. The system contains state-of-the-art authentication methods involving face recognition, speech recognition, optical character recognition (OPT), radio frequencies identification (RFID), and finger, but the user can pass any two authentications to get verified. The reason for enabling five methods is to provide ease in unexpected situations without compromising security. This system can secure any place, e.g., homes, shops, shopping malls (at night time), restricted official buildings, the university's teacher's area, etc.

26.1 Introduction

In today's era, security is a part of daily life concern. There are many circumstances in which security threats can be faced. These threats could be of any type like street crimes, cybersecurity, network security, etc., where street crimes like theft and robbery against a shop are the big threats around us [1]. The "shop security system" is an IoT-based project designed to secure a shop and provide a variety of authentication processes in combination, so the user does not feel the hustle and can do it very quickly. Users can find many security systems for their shops with various security systems. The IoT-based smart system can easily protect shops and homes with user needs compilations. Usually, we do not see that much variety in existing solutions for authentication in their system. They focus on a single authentication process or maybe two. Nowadays, home security or shop security system needs more security layers. It is noticed that such types of security things are very important for our shops and home, so people can easily protect their shops and home with smart technology.

To overcome this security threat, a complete security solution based on IoT is proposed, which can secure a shop from theft behind the owner of that shop. The IoT is an emerging technical, social, and economic topic. Consumer products, industrial and utility components, sensors, and other everyday objects combined with Internet connectivity promise to transform our lives and work, where projections for the impact of IoT on the Internet and economy are impressive, with some anticipating as many as 100 billion connected IoT devices and a global economic impact of more than $11 trillion by 2025 [2, 3].

This research consists of five authentication methods to secure any place from unwanted intruders and generate alerts to the owner. The project's backbone is the microcontroller, which is always powered on to listen to the sensors to act. The process starts with the passive infrared (PIR) sensor, which notifies the microcontroller as it detects any movement in its 13-m range [4, 5]. The microcontroller sends a command to the camera to take a picture, runs a fast scan to see if there is a match,

and then displays the message on LCD for the user [6]. If face verification succeeds, users must pass only one more verification to unlock the digital doors inside; if face verification fails, users must pass two authentications to unlock the digital locks [7]. The voice module gets into the listening mode as face verification is completed, and it receives any voice stream from the user, which it sends to the microcontroller. Then the microcontroller will respond with a message. The speech verification passed or failed; in either situation, a message will be displayed on the liquid crystal display (LCD) for the user [8, 9]. The keypad has a button to send one time password (OTP) to a registered phone number using a GSM module and a numpad to enter the received code [10, 11]. RFID scanner and a fingerprint sensor are also exposed to the user to get verified [12], and suppose someone tries to break the system to enter the place forcefully, in that case, a force-sensitive resistor gets activated. It sends a message to the microcontroller, and the microcontroller will turn on the sirens and send a message to the owner to notify them about this activity using the global system for mobile communication (GSM) module [13, 14].

The rest of the paper is organized as follows. Section 26.2 describes the related work to this security system. Section based on the proposed architecture, Sect. 26.4 describes the methodology we used to implement the system, along with screenshots of the actual system. Section 26.5 describes the results after testing the system by the owner and other individuals, along with screenshots. Finally, Sect. 26.6 describes the conclusion with some indication for future work.

26.2 Related Work

Various studies have been proposed to enhance shop security systems' security and convenience. Verma et al. [15], proposed the system, mainly designed to implement a security system containing a door locking system using RFID which can activate, authenticate, and validate the user and unlock the door in real time for secure access. The system creates a log containing checks-in and checks-out of each user and the basic information. Hemalatha and Gandhimathi [16], designed a system to provide users a secure environment with integrated technology (using mechanical and electronic devices). They developed an electronic code lock system using an 8051 microcontroller that inputs an RFID tag and matches the input through an RFID tag reader. They used a keyboard to enter an OTP pin for authentication. In [17], the researcher tried to build and implement a smart door security system that provides easy and more secure options for the owner to unlock his door; the proposed lock can detect the user's recorded fingerprint and unlock the door. It can also unlock the door through knocks when it matches the predefined pattern. In [18], the author proposed a vehicle security system using an Arduino UNO integrated with a fingerprint sensor module to give access to its owner by using their thumb impressions. It can be done by ensuring that only the persons approved by the owner can access the vehicle. The

owner will be notified using hidden techniques that the vehicle has been accessed utilizing some physical damage. Mirikar et al. [19], proposed a system where authors implemented a three-level authentication process (i.e., RFID, fingerprint, and OTP) to secure an automatic teller machines (ATM's) traditional cash withdrawal process. When the user passes all these security levels, they can withdraw the cash; otherwise, they are not. Ha [20], proposed an IoT-based system to control a door lock using a face recognition algorithm and password. Novosel et al. [21], proposed a system where they implemented a face recognition technology to identify its users on a little computer (Raspberry Pi). Ali et al. [22], proposed a system enabling its user to control the home appliances using voice commands, which will be sent to the microcontroller only when it matches the stored dataset. In [23], we proposed a similar system for building a home automation system using the Bluetooth module. In [24], the system was proposed by a group of two researchers and required two authentications, including OTP on the registered number and fingerprint verification to unlock the door.

26.3 Proposed Architecture

A security system is incorporated with electronic control of shop activities like opening and closing the door, keeping a track record of any malicious activity outside the shop, and informing the owner about that activity in real time. These features are assisted by using different sensors, controls, and other communication modules, as shown in Figs. 26.1 and 26.2.

This paper seeks to strengthen the security system, e.g., we have developed an IoT-based smart security system model, which bypasses two or three authentications, and you will be able to open the shutter locks automatically. The method may also protect us from thieves by recording every suspicious activity. We have implemented five authentications on the system. Initially, the facial recognition process will be initiated through the utilization of a webcam. The precision of facial recognition technology is of utmost importance when the webcam identifies the user's face, as it enables the system to accurately discern the user's facial features under optimal lighting conditions. Conversely, suboptimal lighting conditions may impede the system's ability to accurately recognize the user's face. Following the implementation of facial recognition, four additional authentication methods have been incorporated into the system, namely speech recognition, fingerprint scanning, RFID technology, and one time password verification. Upon successfully bypassing two other authentication measures, the system will automatically grant access to the locks. If the user's face is not recognized, the system will allow him to authenticate himself using other implementation methods. In speech recognition, the user must record his voice or command to pass this authentication method. The recorded voice command is sent to the database to match it with pre-registered voice-trained data, and when the voice matches, the system will check how many authentications passed and how many are left to open the locks. In the RFID process, the user must place their RFID card

Fig. 26.1 Outside demo of the shop security system

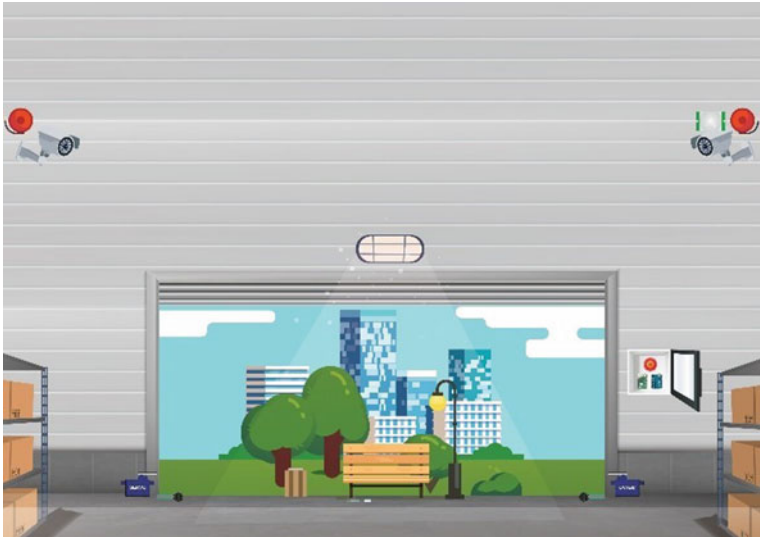


Fig. 26.2 Inside demo of the shop security system

to scan the tag; if it matches, the system checks for how many authentications are left to open the locks. In the OTP process, the user must have to enter the received OTP pin on the system in the required time frame (i.e., in 1 min or less), and when the system validates the pin, the system will check for the left authentications and

proceed itself accordingly. In the fingerprint process, the user must place their finger on the fingerprint module to authenticate him. Suppose the user's fingerprint matches the pre-registered prints, in that case, the system will check for the authentications whether all the required authentications passed or not, and if all the authentications passed successfully, then the locks will open automatically; otherwise, it will not unlock itself and sends a message to the owner about suspicious activity. The purpose of using our Arduino Mega is to develop this system at the lowest cost so that it is easy for every user to buy this system.

The whole system is implemented on the Arduino IDE using Arduino Mega. The proposed architecture of the system is shown in Fig. 26.3. The first process in this system is power ON/OFF. As the user switches the power button ON, the system will start and complete setup in 5 min. After powering ON, the current will pass through the relay module. The relay module operates the electrical switch and can control low voltage, 5 V provided by the Arduino pins. The Arduino Mega 2560 will be open after the voltage is set according to what happens in the Arduino program; it will set all the sensors connected to the sensor with Arduino. The first step in the Arduino will be PIR sensor. Its range is between 6 and 7 m (20 feet), and the detection angle is 110 $\times$ 70 degrees. If a person crosses the distance of the PIR sensor so at the same time PIR sensor will go to Arduino, and Arduino will open the webcam, and at the same time, a message will be received by the owner via GSM module and as long as someone is in the distance range until then PIR sensor. Arduino will keep all sensors ON. If the thief touches the door, the force sensitive will tell the Arduino, then the Arduino will play a siren, and in real time owner will get the message from GSM module. If the thief comes to the webcam to touch, the webcam will capture the picture of the thief, save it to the memory using secure digital card (SD Card) module, and at the same time, store it in Dropbox via Wi-Fi module. The webcam has resistor that controls the voltages.

The face must be pre-registered in the database if a person wants to open the door. If the face is verified, an LCD will get the message "face is verified." After that, LCDs a verification message "to open the door, you have to verify another authentication, which you can choose from fingerprint, speech-recognition, OTP verification, and RFID." The face must be pre-registered in the database if a person wants to open the door. If the face is not verified, then an LCD will get the message "face is not verified." After that, LCD will give two verification messages "to open the door, which you can choose from fingerprint, speech recognition, OTP verification, and RFID." The fingerprint has to be registered first in the Arduino IDE to verify the fingerprints. The speech recognition must first be trained in the Arduino IDE to verify the speech. If a user wants to verify from OTP, the user will have an OTP password message via GSM module; then, the user will input PIN via a keypad. The RFID has to be registered first in the Arduino IDE to verify the user's card. If verified, Arduino will go to servo motors, and the door will open with the help of servo motors.

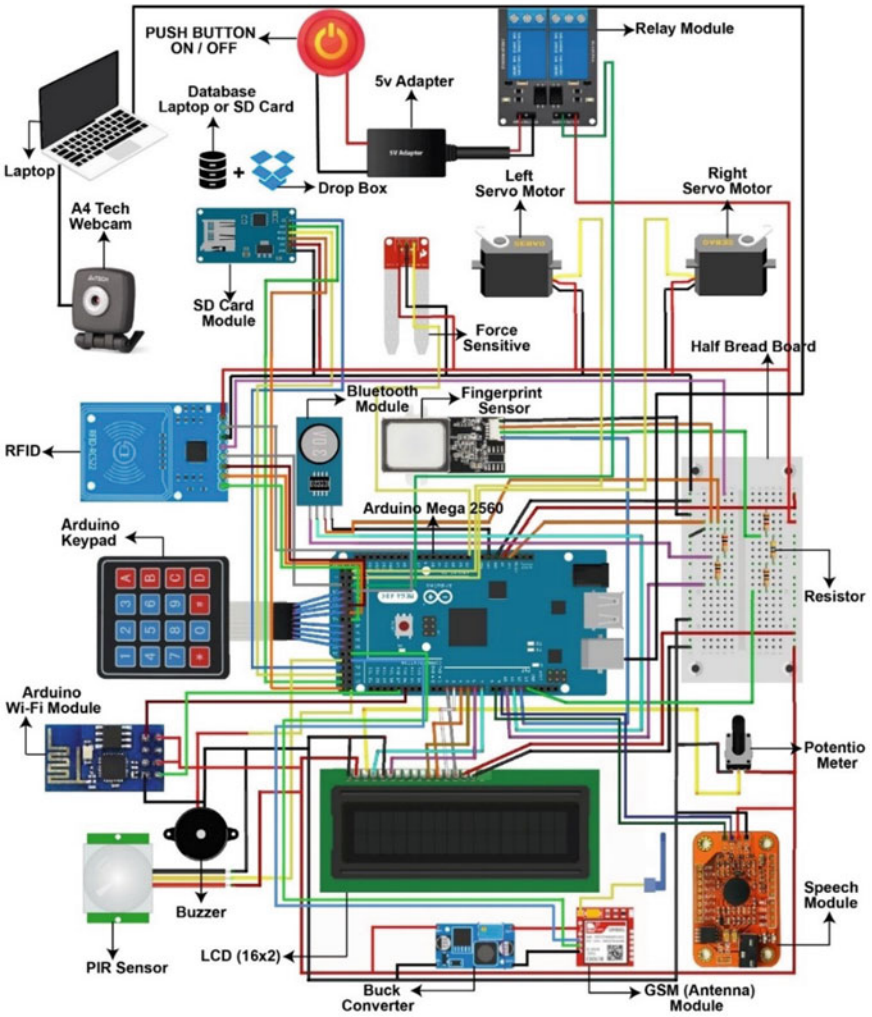


Fig. 26.3 Proposed model of smart security system

26.4 Hardware Implementation

26.4.1 Significance of Arduino Mega

The Arduino Mega 2560 is a microcontroller board based on the ATmega2560. It has 16 analog inputs and 54 digital inputs/outputs. The purpose of this Arduino Mega board is to contain everything needed to support the microcontroller, connect it to a computer with a USB cable, and you can run OpenCV with the help of a computer

and all the sensors by storing your program in the sensors from the Arduino board to get started.

26.5 System Architecture

In Fig. 26.4, there are different modules in our system. In the first module, PIR must be activated. The second module, PIR, will start the camera. After activating, the user will scan the face. After being trained, it will detect the user after being recognized. In the third module, the system will ask for another authentication. After verifying the card, the shop's shutter will be unlocked automatically. If the face is not recognized, we will move toward fingerprint in the fourth module.

After finger verification, we will have to do two more authentications. In the fifth module, you must enter the password after verifying the finger and scanning the card. If it is correct, the door will open. In the sixth module, if the password is incorrect. So you can open the door by recognizing the voice; if the voice is incorrect, then the system will go back to the first module. The seventh module is the GSM module, which will notify the owner of whether the authentication is verified and will sound the siren in case of incorrect authentication. The eighth module is force sensitive. If someone forcibly rips from the shop, they will notify the system owner and replace the siren.

26.5.1 System Experimental

We have experimented with different modules in the system, and a step-by-step process for storing user data is given. We will check the system's accuracy and then display the result of the experiment.

26.5.2 Face Recognition

We experimented with how to store, train user face data, and recognize them. All this is explained step-by-step. First, we have created a dataset folder with the user's face stored in each folder. The total time assumed to collect images and capture the user's face was 10 s as shown in Fig. 26.5. When the owner captures the user's face, he will capture each user's 21 face samples at the same time given in Fig. 26.6 and then generate it in a folder because there is a lot of accuracy in it. All images were captured in .JPEG format in RGB color, 70–73 B for each image. After that, the user's faces are stored in the folder; then we collect their data, which is stored in the database. After storing each user's information or face images in the database, we trained all the data. After training, knowing which user is recognized and unknown

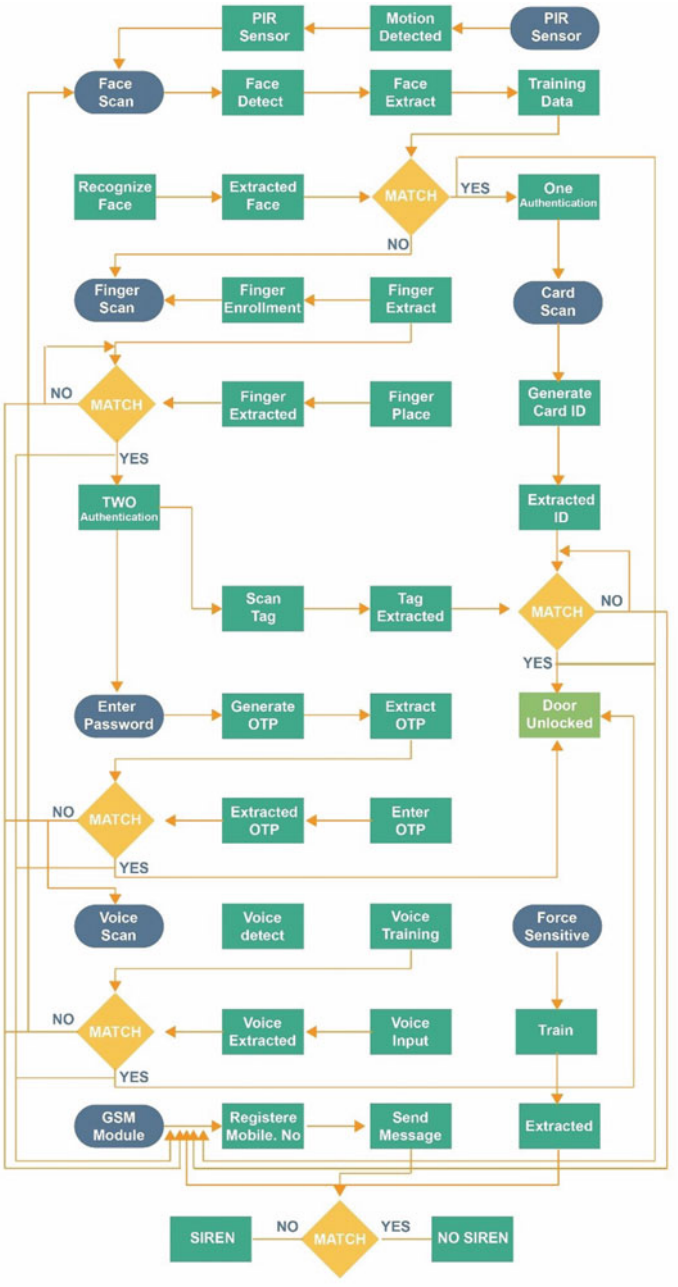


Fig. 26.4 System architecture of security system

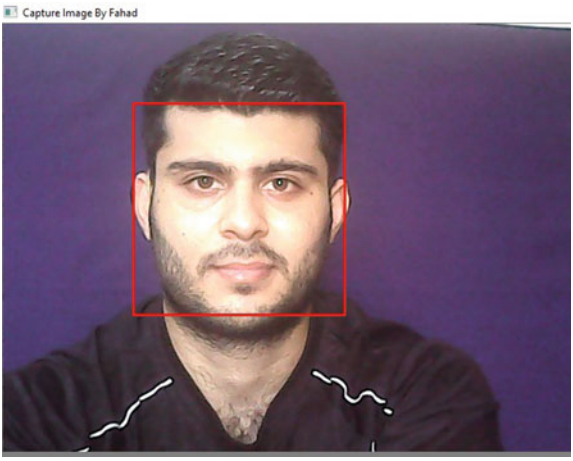


Fig. 26.5 Identification of user face

in the criminal records list will be possible. Finally, after the data train, we see in Fig. 26.7 that the user is recognized, while in Fig. 26.8, one user is not recognized.

26.5.3 Voice Recognition

We first downloaded the V3 library and installed it in Arduino IDE. In Fig. 26.9, after that, we went to the serial monitor of voice recognition, where the voice is trained, and once loaded, we can check our voice to see if there is a record in the voice module in Fig. 26.10.

26.5.4 Fingerprint

In Fig. 26.11, enter the user's ID number to scan the user's finger, then enter it. Then the finger module tells us to place the finger. When we place it, it asks us to place the finger again. After that, the user's fingerprint is stored in the module's memory.

26.5.5 Radio Frequency Identification

In Fig. 26.12, to know the tag's ID, we will move the tag closer to the RFID, then after being detected, we will get the ID number of the tag in the Arduino IDE serial monitor.

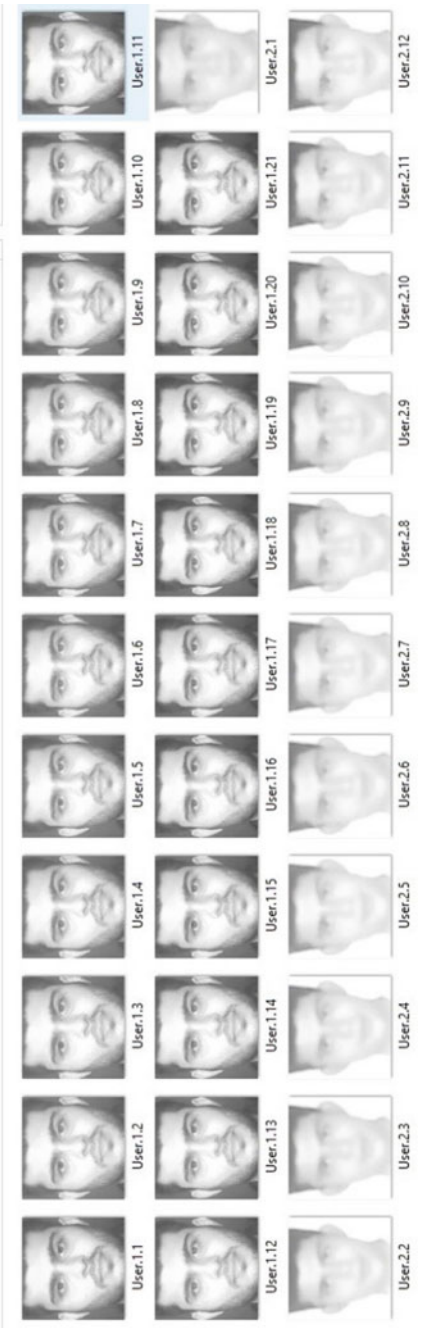


Fig. 26.6 Screenshot of user face in dataset folder

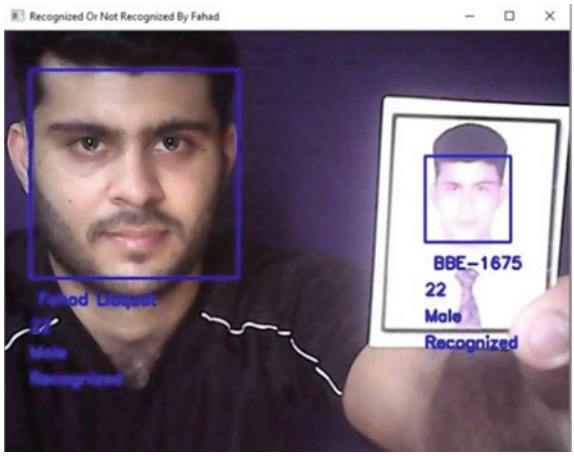


Fig. 26.7 Screenshot of recognized user

Fig. 26.8 Screenshot of not recognized user

```
------------------------------
sigtrain 0 Door_Open
------------------------------
Record: 0       Speak now
Record: 0       Speak again
Record: 0       Success
Success: 1
Record 0        Trained
SIG: Door_Open
```

Fig. 26.9 Screenshot of user voice trained

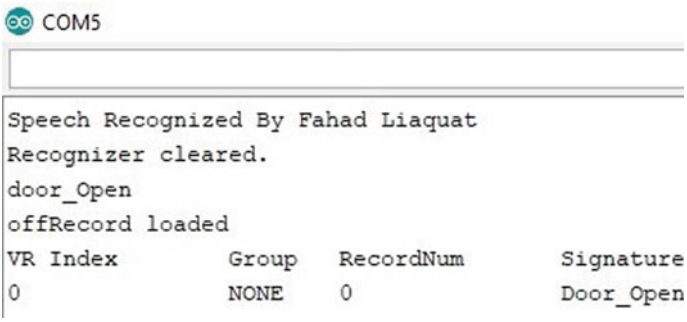


Fig. 26.10 Screenshot of user voice recognized

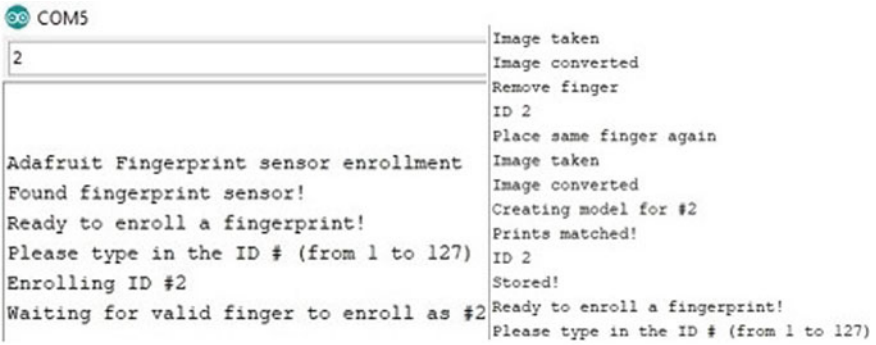


Fig. 26.11 Screenshot user's finger enrolled

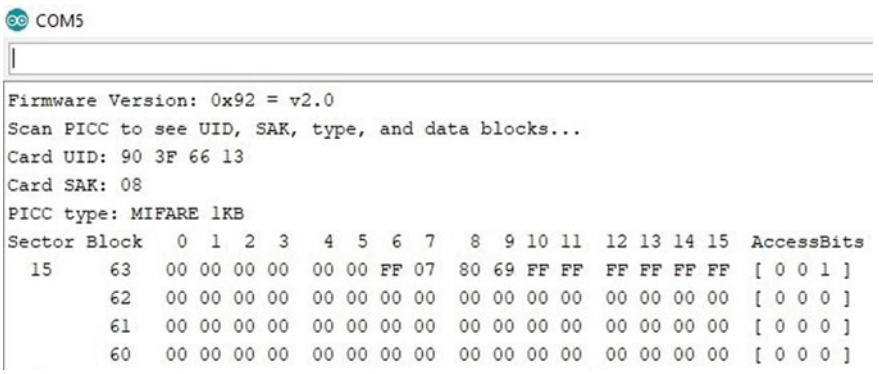


Fig. 26.12 Screenshot of scan tag process

26.6 Implementation and Results

The test results are divided into three phases. One is when a user is recognized; the second case is when a user is not identified, and the last point is when a user enters the wrong password, door lock and unlock mechanism. We have connected the Arduino board to the laptop to run the security system. Once connected, the system welcomes us to the first LCD screen.

26.6.1 Identification of User

The system will ask us to do a face scan. After receiving the message "scan your face" from the system, you will scan the face from the laptop using a webcam. The system will show in Fig. 26.13, recognized you if your data is enrolled in the database. The system will give you the message recognized on LCD. Once recognized, the system will ask you for one more authentication, then the user will scan the tag on the system and gives the tag-matched message. After the tag match, the system will give a door unlock message, and the shop will open the shutter after doing two authentications. After completing the authentication process, the system will send a message to the shop's owner.

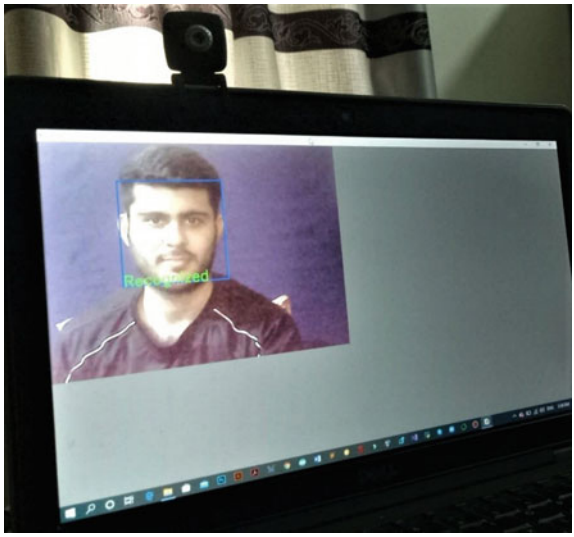


Fig. 26.13 Screenshot of user face recognized

26.6.2 User not Identified

The system will give a "face unknown" message if the user's face is not recognized. When someone tries with the wrong authentication, the system will send a message to the owner's mobile number. After doing the wrong authentication, the system will give the option of a fingerprint. The system will ask us to do a fingerprint scan. If the user's fingerprint is scanned appropriately and matches the registered print, the system will ask you to perform two more authentications for further validation in which the user will pass the RFID and password.

26.6.3 Wrong Entry of Password

After doing the wrong authentication, the system will give the option of voice recognition, asking us to record a speech of 20 s. If the voice is recognized, then the system will display the "voice recognized" message on LCD. If a voice is not recognized in 20 s, the system will restart.

26.7 Conclusion and Future Scope

The system provides various state-of-the-art authentication methods to make a place secure without causing trouble for the user. The system is fast, secure, reliable, and easy to use. As anyone comes close to the entrance, it detects the person and automatically starts to find a match against the person standing by using face recognition techniques. The user only has to pass any two authentications to unlock the automated doors. And the reason for adding five methods is to provide ease, as its impossible even for the authenticated person to get verified by limited options. If the face matches with any record in the database, the user has to pass only 1 more authentication to proceed; otherwise still has to complete two more. Other methods include RFID, speech recognition, OTP, and fingerprint. The system records the captured images as a log of the people who entered the place. It also provided alerts to the owner on the registered number by a text SMS when anyone tries to unlock it forcefully. There is also a trap for the stealers and thieves; we have traditional locks outside and automated ones inside, so if a person tries to break the traditional locks, the place is still secure by inside automated locks. Also, an alarm will be triggered to panic him. As this project consists of various authentication methods, our team believes there is a huge potential to extend this project variously or use the implemented mechanisms in other projects to facilitate human beings. We have mentioned some future directions below:

- By adding CCTVs in the system, this security system will be able to be used for security + surveillance.

- We can extend the system by using multiple microprocessors for each entrance.
- The security system can be used to secure the entrance of any place, e.g., office, shop, mall, building, ground, or any other place that needs to be secured.
- This system can be integrated with the attendance system by using face recognition and that solution will benefit industries, schools, and all the places where many people come.
- We can connect the system with the nearby police station, so in case of causality police can be alert and do the needful at the time.
- We can add one more authentication in the existing echo system which is knowledge-based authentication (KBA).
- Another authentication method can be identity documents, e.g., driving license, CNIC #, or bank cards.
- We can connect the airport system with this project to reduce the efforts, e.g., if the passenger is in the airport on not for whom flight is waiting for the last 10 min.

References

1. Yoo, C.: The emerging internet of things: opportunities and challenges for privacy and security. In: Governing Cyberspace During a Crisis in Trust Essay. CIGI, Waterloo, ON. www.cigionline.org/articles/emerging-internet-things (2019)
2. Soleh, S.S.S.M., Som, M.M., Abd Wahab, M.H., Mustapha, A., Othman, N.A., Saringat, M.Z.: Arduino-based wireless motion detecting system. In: Proceedings of the 2018 IEEE Conference on Open Systems (ICOS), pp. 71–75. IEEE (2018)
3. Nainan, S., Ramesh, A., Gohil, V., Chaudhary, J.: Speech controlled automobile with three-level biometric security system. In: Proceedings of the 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), pp. 1–6. IEEE (2017)
4. Phoka, T., Phetsrikran, T., Massagram, W.: Dynamic keypad security system with key order scrambling technique and OTP authentication. In: Proceedings of the 2018 22nd International Computer Science and Engineering Conference (ICSEC), pp. 1–4. IEEE (2018)
5. Nath, S., Banerjee, P., Biswas, R.N., Mitra, S.K., Naskar, M.K.: Arduino based door unlocking system with real time control. In: Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), pp. 358–362. IEEE (2016)
6. Siddiqui, M.F., Siddique, W.A., Ahmedh, M., Jumani, A.K.: Face detection and recognition system for enhancing security measures using artificial intelligence system. Indian J. Sci. Technol. **13**(09), 1057–1064 (2020)
7. Stojkoska, B.L.R., Trivodaliev, K.V.: A review of internet of things for smart home: challenges and solutions. J. Clean. Prod. **140**, 1454–1464 (2017)
8. Mawgoud, A.A., Taha, M.H.N., Khalifa, N.E.M.: Security threats of social internet of things in the higher education environment. In: Toward Social Internet of Things (SIoT): Enabling Technologies, Architectures and Applications, pp. 151–171. Springer, New York (2020)
9. Amuthan, A., Sendhil, R.: Hybrid GSW and DM based fully homomorphic encryption scheme for handling false data injection attacks under privacy preserving data aggregation in fog computing. J. Ambient Intell. Hum. Comput. **11**(11), 5217–5231 (2020)
10. Ramesh, S., Govindarasu, M.: An efficient framework for privacy-preserving computations on encrypted IoT data. IEEE Internet Things J. **7**(9), 8700–8708 (2020)
11. Song, W.-T., Hu, B., Zhao, X.-F.: Privacy protection of IoT based on fully homomorphic encryption. Wireless Commun. Mobile Comput. **2018**, 1–7 (2018)

12. Song, T., Li, R., Mei, B., Yu, J., Xing, X., Cheng, X.: A privacy preserving communication protocol for IoT applications in smart homes. IEEE Internet Things J. **4**(6), 1844–1852 (2017)
13. Tonyali, S., Akkaya, K., Saputro, N., Uluagac, A.S., Nojoumian, M.: Privacypreserving protocols for secure and reliable data aggregation in IoT-enabled smart metering systems. Fut. Gener. Comput. Syst. **78**, 547–557 (2018)
14. Singh, H., Pallagani, V., Khandelwal, V., Venkanna, U.: Iot based smart home automation system using sensor node. In: Proceedings of the 2018 4th International Conference on Recent Advances in Information Technology (RAIT), pp. 1–5. IEEE (2018)
15. Verma, G.K., Tripathi, P., et al.: A digital security system with door lock system using RFID technology. Int. J. Comput. Appl. **5**(11), 6–8 (2010)
16. Hemalatha, A., Gandhimathi, G.: RFID, password and OTP based door lock system using 8051 microcontroller. Int. J. Eng. Res. Technol. **7**(11), 1–6 (2019)
17. Nada, E., Aljudaibi, S., Aljabri, A., Raissouli, H.: Intelligent lock applied for smart door. Int. J. Comput. Sci. Inform. Sec. **17**(6), 1568 (2019)
18. Ranjeet, J.K., Ganesh, S.A., Jayawant, S.A., Sambhaji, P.S.: Smart vehicle security system using finger print sensor (2019)
19. Mirikar, S., Chikhale, D., Dhumal, Y., Indalkar, S.: Smart ATM access using RFID, biometric finger print and GSM module (2017)
20. Ha, I.: Security and usability improvement on a digital door lock system based on internet of things. Int. J. Sec. Appl. **9**(8), 45–54 (2015)
21. Novosel, R., Meden, B., Emersic, Z., Peer, P.: Face recognition with raspberry PI for IoT environments. In: International Electrotechnical and Computer Science Conference ERK (2017)
22. Ali, A.T., Eltayeb, E.B., Abusail, E.A.A.: Voice recognition based smart home control system. Int. J. Eng. Invent. **6**(4), 01–05 (2017)
23. Pal, S., Chauhan, A., Gupta, S.K.: Voice controlled smart home automation system. Int. J. Recent Technol. Eng. **8**(3), 4092–4093 (2019)
24. Touqeer, H., Zaman, S., Amin, R., Hussain, M., Al-Turjman, F., Bilal, M.: Smart home security: challenges, issues and solutions at different IoT layers. J. Supercomput. **77**(12), 14053–14089 (2021)

Chapter 27
Logistics Security Integrated Communication System Under the Background of 5G Artificial Intelligence

You Zhou

Abstract This paper designed the integrated communication system of logistics support security under the background of 5G artificial intelligence. In view of the current development of computer communication technology, the spread of Internet technology, the functions of AI-based logistics support system. This paper introduces the overall structure of the system. It was introduced the three levels and five functional modules. Through the deployment of three levels of equipment environment, the system realized the central command mode of data from bottom to up, from scattered to gathered. It was introduced in three main achievements and a prospect. In the meanwhile, we were introduced to the development platform and tools of the system. The system has high use value and saves the production cost of the enterprise.

27.1 Introduction

In view of the current development of science and technology, various technologies have been rapidly developed that they were computer hardware and software technology, Internet technology, communication technology, video surveillance and analysis technology, cluster communication technology and so on [1]. The integration of various technologies has become the mainstream development trend, and the between various technologies have become increasingly blurred. Some technologies have been used widely in all kinds of communication products, Internet products, and Internet of Things products. All kinds of science and technology have nested and integrated with each other in the scientific and technological achievements of these various industries, providing people with better quality and higher performance

Y. Zhou (✉)
Ecological Environmental Protection Company, Daqing 163000, China
e-mail: cecaa2080@126.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_27

product experience. So the concept of "converged communication" (also known as unified communication), short for UC. The concept of unified communication was first proposed by the International Electrical Union ITU in 2004 [2]; it is a new communication mode integrating computer technology with traditional communication technology. The computer network and the traditional communication network on a network platform realize telephone, fax, data transmission, audio and video conferencing, call center, instant messaging and many other application services [3].

In this paper, the logistics security integration communication system is a comprehensive data integration system that integrates information collection, data processing, behavior analysis and overall decision-making in accordance with the international UC industry standards [4], according to the actual needs of oilfield business, combined with a variety of cutting-edge information technology development.

27.2 Logistics Security Integrated Communication System Under the Background of 5G Artificial Intelligence

The research content of this system includes five modules: converged communication basic application services (digital cluster command and dispatch system), emergency support function module, behavior analysis application service module, intelligent alarm application service module and hyper-converged system application module.

The login interface is shown in Fig. 27.1.

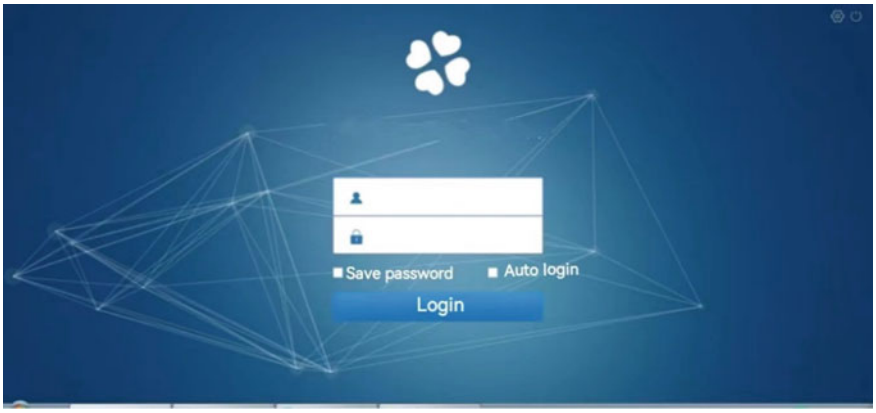


Fig. 27.1 Login interface

27.3 Overall System Architecture

The system summary structure is shown in Fig. 27.2.

27.3.1 The Architecture Design of the System is Divided into Three Logical Layers

Reference the standard of open architecture framework guide (TOGAF is short for the open group architecture framework), the system architecture design includes: TDE terminal data execution layer (TDE is short for terminal data execution layer), SDA server data acquisition layer (SDA is short for server data acquisition layer) and CDAM core data analysis management layer (CDAM is short for core data analysis management layer).

1. TDE terminal data execution layer: The scope of the terminal data execution layer is in each terminal node of the platform, the purpose is to quickly adapt the OS and each terminal node device, send the data obtained by each node device to the data acquisition layer, and perform feedback data, e.g., cameras, sensors, alarm devices, and intelligent devices for fault information collection.
2. SDA server data acquisition layer: The server data collection layer covers the DMZ of the server room (DMZ is short for data message zone). It was aimed to summarize the data collected by various nodes and establish a data matrix

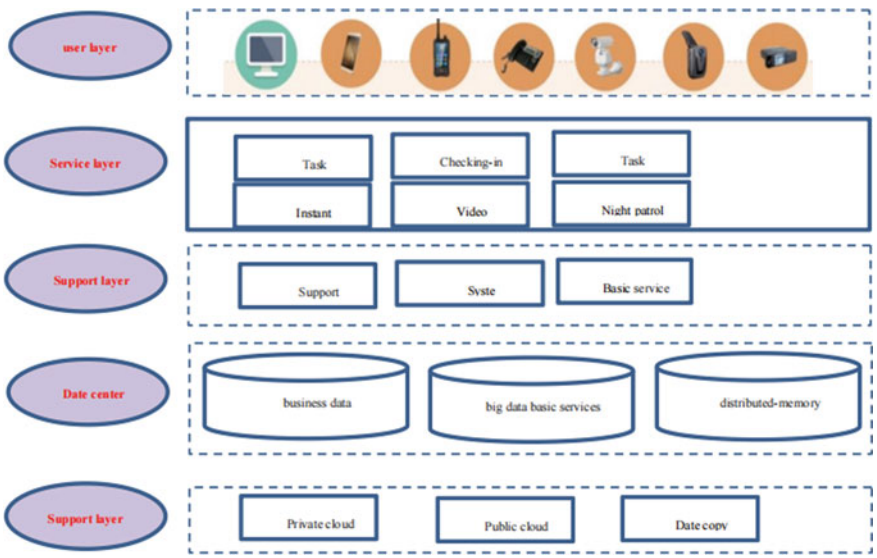


Fig. 27.2 Overall system architecture

based on algorithms to provide a complete database for the lower part of the data analysis. The DMZ buffer zone is a buffer zone to avoid security risks caused by arbitrary access to the Intranet and extranet. The dual-layer firewall and dual-layer policy deployed in the DMZ buffer zone ensure that extranet information can enter the Intranet only after source address translation and information filtering. This effectively ensures that extranet information does not interfere with the internal network security of the oil field. It meets the standard requirements of national grade protection 2.0.

3. CDAM core data analysis management: The core data analysis management is the decision-maker of the system application. The purpose is to filter the complete data by algorithm and provide users with accurate, intuitive, controllable and feedback data, so that users can take countermeasures according to the analyzed information, e.g., video information collected by the camera, warning information, fault message, to receive command information, etc.

27.3.2 Five Modules

The system was divided into five modules.

(1) Converged communication basic application services (digital cluster command and dispatch system): The system was divided into five levels, consisting of nine basic function sub-modules and one hundred and twenty-five business function sub-modules.

Based on the logic centralized control of the new network architecture, referring to the application layer gateway conversion mechanism, adopting the WIA-PA, WirelessHART and 6LoWPAN and so on industrial wireless protocol, so as to achieve the conversion effect of that between industrial wireless layers and the IPv6 network protocol (WIA-PA is the wireless networks for industrial automation process automation, WirelessHART is a wireless communication standard, 6LoWPAN is a network standard that is based on IPv6 of low-speed wireless private area). In terms of the conversion between Ethernet protocol and IPv6 protocol, a typical industrial Ethernet protocol and IPv6 protocol conversion method is adopted to realize the protocol conversion based on application layer quality of service (QoS).

At the same time, the system also deploys a high-speed protocol conversion device for industrial heterogeneous network, which supports a variety of industrial wireless protocols and industrial Ethernet protocols and combines with the application layer gateway conversion mechanism, then the millisecond high-speed protocol conversion was realized, solved the problem that industrial heterogeneous network protocol conversion caused and achieved efficient scheduling between multiple controllers.

(2) Emergency support function module: According to the emergency level, the hierarchical emergency plan processing process library, emergency plan knowledge base and emergency plan authority library are established in the database

in advance. According to the application function, part of the scheme library is implemented on the interface and compiled into visual decision instructions by means of "graphic syntax" and "element structure". According to the actual situation, the decision instructions are fed back to the terminal according to the authority.

(3) Behavior analysis application service module: This system adopts the detector method established by DPM component model to overcome the problem of crowd shading and greatly enhance the analysis ability of operation scenes. For trajectory clustering, the system uses KLT tracker and clustering method [5] to accurately analyze human behavior, effectively improve the ability to judge individual behavior in multiple targets, and the KLT tracker also greatly improves the effect of face recognition.

(4) Intelligent alarm application service module: It strictly complies with the CAP principle, the alarm application module is deployed based on the four-level logical rules of sorting data, setting alarm rules, triggering alarm policies, and sending alarm notifications. The cascading analysis method facilitates the deployment of alarm policies. In addition, in order to improve the accuracy of alarm orientation and alarm efficiency, the platform also adds a custom alarm information extension service module and an intelligent auxiliary decision module to facilitate safe production.

(5) Hyper-convergence system application module: AR hyper-convergence module is based on J2EE [6] and WPF technical framework, and adopts SOA (SOA is short for service-oriented software architecture) to ensure the scientific nature of the system. Through GIS and high-point real scene fusion, multi-source data super fusion, decision-makers can in real time obtain the first view of the scene information and open space integrated monitoring resources. In view of the problem that compatibility and accuracy cannot be achieved at the same time, the platform on the basis of SLAM2.0 theory, the application of AR SDK standards, the use of edge computing, while ensuring accuracy, cross-platform [7] compatible with iOS and Android systems.

27.3.3 Hardware Device Environment Deployment

The system was deployed in a three-tier device environment to realize the central command mode of data from the bottom up and from scattered to clustered. The hardware deployment at each level is as follows:

(1) Basic production service sites: Deploy cameras, sensors, precaution device, bridge ligands and other data acquisition, distribution network transmission, decision-making equipment as the nerve end of the system.

(2) Secondary organs of service units: As the main body of the system data transmission, there are independent secondary servers, protocol conversion switches, bearer network ligands, video recorders and other relay transmission and information aggregation equipment.

(3) Command center: As the command center brain, there is deploy command center large screen, core server, function processing extension server and other human–machine interaction, decision-making aid and visual equipment.
(4) Vehicle and mobile point locations: Due to the needs of some services, special devices such as mobile data collection, distribution network transmission, decision-making, and energy security need to be deployed to specify time and frequency to extract relevant logistics data for command and management.
(5) Relevant security units: The layout of the bridge, aggregation switch, Digital trunking terminal, Video capture equipment, video equipment, TV screens and other information equipment: Real-time monitoring and management of logistics support operation status.
(6) Transmission network design: When designing the backbone network transmission scheme, in order to take into account the security of communication between the command center and the agencies in each operation area, it was also necessary to meet the data quality of the collected information in each operation area and upload it to the logistics command center. In Taiwan, an oilfield buffer zone was adopted as the core data area. The system uses the oilfield buffer as the core data area. In this way, the existing network resources of existing oil fields can be used to setup a special network of logistics support monitoring system, and the requirements of data security and transmission stability of logistics network can be met by relying on the special network.

27.3.4 The Main Results Achieved

Innovation result one.

According to the requirements of oilfield logistics support service management, a logistics support management system module based on video application was constructed: back-end supervision, front display, remote linkage, with the following functions:

(1) Behavior analysis and identification services: Personnel clothing intelligent analysis, dynamic trajectory research and judgment.
(2) Logistics service: Kitchen personnel test information, illegal kitchen personnel alarm. It is to implement personnel behavior analysis system module that it is based on video transmission and video data, analysis of illegal entry, illegal operation, abnormal behavior and abnormal behavior.
(3) Voice call service: Single call, group call, basic call between terminals, basic call between dispatchers and call initiated by the dispatcher to the terminal.
(4) Intelligent alarm service: Realize abnormal data alarm in the platform, alarm data linkage dispatching center, administrator handheld terminal, data control branch center, etc.
(5) Intelligent vehicle monitoring system: Vehicle video surveillance using 4G network through the vehicle monitoring equipment, the real-time image inside the vehicle is transmitted back to the monitoring center, which is convenient

for unified supervision and scheduling. The main functions are real-time video browsing, timing snapshot picture browsing, audio and video storage, real-time location positioning, map tracking, historical track and electronic fence and other conventional functions.

(6) Gas alarm linkage system: When the gas alarm controller triggers the alarm, the platform linkage alarm was realized and recorded in the background database.

Innovation result two.

With video space big data, high-low video linkage technology, the formation of a three-dimensional logistics supports service supervision system.

(1) Hyper-converged system services: It is based on the Internet of Things, augmented reality, intelligent analysis, GIS, software integration and other advanced technologies, the construction of realistic, grid three-dimensional prevention and control system services.
(2) Lot of devices apply front-end access services, back-end management services, overhead and level permission settings.
(3) The video spatial data platform supports the realistic application of monitoring resources and supports the display of the location of electronic control resources such as cameras, bayonet and Wi-Fi probes through the video real map of the platform. Compared with the visualization application of GIS map, the real visualization is a higher degree of visualization means, which improves the sense of presence in video management and resource scheduling.

Innovative product three.

Multidimensional realization of logistics service vehicle visual traceability, logistics service support personnel visual management and other functions to improve the quality of logistics support service.

(1) Visual traceability of logistics service vehicles: Through the satellite positioning system, real-time tracking of vehicle location, backward check vehicle trajectory facilitate vehicle operation management and realize the whole process supervision of logistics transportation.
(2) Visual management of service support personnel: Key areas, places, facilities, and main roads are monitored in real time, and emergency response drills and post inspection tours were carried out using the video dynamic rotation function of key posts, making on-site monitoring more intuitive and three-dimensional.
(3) AI-assisted decision-making function: When an exception occurs during business processing, it can provide decision-makers with feasible operational guidelines, direct them to the detailed operation process, visualize the business processing process, and improve their comprehensive judgment ability.

27.3.5 Market Outlook

Through the new system, the integration and information sharing of the decentralized system built by various units were completed. Relying on the new platform,

the command and dispatch center has been built to realize the information-based management of the overall logistics support command and service quality supervision. The integrated logistics support command system has been promoted and used in oil and gas production plants, material companies, water companies, chemical groups, equipment manufacturing and other units.

After the completion of the system, it can effectively improve the service methods, reduce the manual workload, reduce the service operating cost, improve the service quality, achieve the standardization and scientific service work, effectively save energy and consumption and save a lot of annual costs.

27.4 System Development Platform and Tools

Platform and tools:

(1) System architecture: J2EEframe.
(2) Development language: Java, Javascript, XML.
(3) Geographic information platform: GIS platform, B/S framework.
(4) Database: Oracle.

System operation.
Client operating system:

(1) Microsoft Windows 9.0 or higher.
(2) Server operating system: Windows, Linux.

27.5 External Cooperation Situation

Because some key modules and protocol conversion need commercial authorization, and some core algorithms need to rely on commercial logic library, some functions of the platform have applied for external assistance for customized development, specifically involving the following contents:

(1) Core algorithm
In order to improve the function of multi-object alarm recognition effectively in complex masking environment, external assistance is needed to build a core algorithm library in DPM component model, which applies Bayes model [8] to segment appearance and motion features.
(2) Converged communication channel
Due to the remote location of some terminals, insufficient signal strength, poor stability, serious attenuation and other reasons, it is necessary to customize special satellite channels and independent channels of ground base stations from communication operators to avoid pilot pollution of MIMO channels caused by hyper-fusion channels with high accuracy requirements in complex communication environments.

(3) Customized function interface
Compared with traditional software development, hyper-convergence platforms need to reserve extended function interfaces for terminal devices, because some underlying development languages are difficult to be compatible with industrial Internet of Things, intelligent algorithms [9] and other technologies. The terminal equipment of various manufacturers is difficult to match on the hyper-converged platform, so on the key point of expanding the functional framework, apply for external assistance to develop a customized unified external functional interface, so as to integrate a variety of cutting-edge technologies with different models of equipment and provide a unified platform standard for future functional expansion.

27.6 Suggestions and Prospects

During the research and development of this system, through communication with the top team in China, the logistics support business process was further refined based on the actual oilfield business, and it was found that the integration platform still has a large space for business expansion such as forest fire prevention and early warning, logistics support industrial control configuration, geological supervision services, industrial big data analysis and other aspects. The extended functions of this system can meet the above services fully because of adopted advanced technology.

27.7 Conclusions

The system can be promoted widely. It is suitable for promotion and application in the logistics support service system of CNPC and the logistics support command department of oil field. The system has good operability, flexibility, extensibility, compatibility and reliability. It was praised by scholars.

The system solves the problems such as non-integration of public network, private network, non-centralization of data, non-unification of technical defense, inadequate management, and information Island. It realizes the "reality" of command and dispatch, the "integration" of business application, and the "intelligent" [10] management of operation process. It is conducive to the information connectivity, supervision linkage, data joint use and flat vertical management between the upper and lower. Rational use of existing resources, accurate grasp of management dynamics, reduce of false feedback, saved manpower, reduce of costs, using information means to improve the decision-making power to achieve the oilfield logistics support information construction of the last mile.

Acknowledgements Foundation item: Heilongjiang Province education science planning key topic. Project approval number: ZJB1423241. Research on improving the quality of IT courses in higher vocational colleges under the action plan of improving quality and training excellence-take <website front-end development> as an example.

References

1. Wang, Y., Deng, G.F., Wei, Z.C., Bao, H.B., Li, P., Nie, L.G.: Intelligent power supply support and command system based on multi-communication fusion. Inform. Technol. **47**(02), 167–172 (2023)
2. Zhang, Z.G., Li, Y.F., Wang, Z.J., Li, X.J., Tan, J., Xu, C.: Application of fusion communication in freeway emergency dispatch and command communication system. Sichuan Archit. **40**(5), 342–347 (2020)
3. He, L.C.: Application of fusion communication in emergency dispatch and command communication system. Changjiang Inform. Commun. **36**(01), 227–230 (2023)
4. Yuan, Y.: Design and implementation of emergency command and dispatch system based on fusion communication. Electr. Technol. **51**(06), 55–57 (2022)
5. Wen, P.F.: An adaptive clustering method for energy meter cloud data based on data mining. J. Jiujiang Univ. **38**(01), 76–80 (2023)
6. Zheng, S.H.: Design and implementation of alarm monitoring and management system for communication equipment based on J2EE. Electr. Eng. Product World **29**(07), 63–68 (2022)
7. Song, G., Weng, H.M.: Based on cross-platform instrument data acquisition technology. Ind. Control Comput. **36**(02), 11–13 (2023)
8. Zhang, Y.K., Li, L.J., Chen, X.Y.: Multi-layer network link prediction using naive Bayesian model. J. Appl. Sci. **41**(01), 23–40 (2023)
9. Liang, G.S.: Application prospect of intelligent algorithm in security community. China Digital Cable TV **2023**(05), 39–42 (2023)
10. Wang, B.Y.: The application and development of Internet of Things technology. Inform. Record. Mater. **22**(11), 241–243 (2021)

Chapter 28
Deep Learning Unveiled: Investigating Retina Eye Segmentation for Glaucoma Diagnosis

Abdul Qadir Khan, Guangmin Sun, Anas Bilal, and Jiachi Wang

Abstract Glaucoma, a prevalent cause of permanent eyesight loss, manifests through the gradual degeneration of retinal ganglion cells (RGCs). These specialized cells play a crucial role in transmitting information about sight that is transmitted from the eye to the brain. One of the most important diagnostic tools for glaucoma is an optic disc (OD) and optic cup (OC) segmentation. OD and OC are two distinct structures in the retinal images of the eye. For joint OD and OC segmentation and both of these tasks, we provide a deep framework called ResUNet++ with attention gates. Our framework is based on an efficient encoder–decoder architecture and its ability to learn discriminative features for both OD and OC segmentation. We evaluate our framework architecture based on UNet, ResUNet, and ResUNet++ with attention gate, on six publicly available datasets, CRFO-v4, Drishti-GS1, G120, ORIGA, PAPILA, and REFUGE1. The results show that our framework outperforms the state-of-the-art approaches in terms of accuracy and efficiency. Our findings suggest that our joint segmentation framework is a promising approach for accurate and efficient OD and OC segmentation. It could lead to the development of new tools for glaucoma diagnosis and early detection.

A. Q. Khan · G. Sun
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
e-mail: gmsun@bjut.edu.cn

A. Bilal (✉)
College of Information Science Technology, Hainan Normal University, Haikou City 571158, Hainan Province, China
e-mail: a.bilal19@yaoo.com

J. Wang
Software College, Shenyang Normal University, Shenyang, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_28

28.1 Introduction

Glaucoma holds paramount significance as the highest global factor in permanent visual impairment. Projections indicate that the global prevalence of glaucoma is expected to escalate significantly, with an estimated 111.8 million individuals anticipated to be affected by the condition by the year 2040 [1]. Timely and accurate diagnosis plays a crucial role in the effective management of a range of vision-related degenerative conditions [2]. Examples of ocular conditions encompass glaucoma, diabetic retinopathy (DR), and age-related macular degeneration (AMD). The optical disc (OD) corresponds to the retinal region that encompasses the optic nerve head, whereas the optic cup (OC) pertains to the distinct area enclosed within the boundaries of the OD on the retina. The size and shape of the OC can be used to assess the progression of glaucoma. However, accurate segmentation of OD and OC is challenging. This is due to the OD and OC shapes which can vary considerably from person to person, and the presence of artifacts in fundus images can further complicate the segmentation process. The sound structure of the OD and OC can provide valuable insights into various eye diseases [3]. Deep neural networks have exhibited notable performance, at times surpassing that of ophthalmologists, particularly when applied to meticulously annotated datasets [4]. The development of computer vision techniques for the automated analysis of medical images has shown tremendous potential for improving clinical diagnosis and treatment of ophthalmic diseases. In particular, the precise segmentation of the OD and OC within fundus images represents a pivotal and indispensable task that holds significant promise to enable timely detection and continuous monitoring of glaucoma, a debilitating condition that affects millions of people worldwide.

In the realm of OC and OD extraction, techniques leveraging convolutional neural networks (CNNs) have emerged as a promising avenue. However, existing CNN methods often fall short in fully harnessing the potential of clinical prior knowledge, resulting in limitations in the extraction performance of OC and OD [3]. The OD is composed of two separate elements: the neuro-retinal rim located in the peripheral region and the optic cup OC situated in the central area. Progressive disease leads to structural changes within the OD, leading to the thinning of the neuro-retinal rims and the enlargement of the OC. Therefore, precise identification of both the OD and OC becomes critically important, although automated recognition of the OC presents more complex obstacles [4, 5]. Achieving accurate and reliable segmentation of these structures remains a formidable challenge due to the complex interplay between the anatomical features of the eye, the imaging modalities, and the inherent variability in the appearance and shape of the OD and OC.

Advancements in deep learning-based approaches have shown significant promise in automatic segmentation of the ONH [6]. UNet is a widely used deep learning architecture for medical image segmentation [7]. However, conventional UNet may not be optimal for ONH segmentation due to its relatively shallow architecture. Efficient Net is a deep learning architecture that has been shown to achieve state-of-the-art performance on a variety of image classification tasks with a significantly smaller

number of parameters than other architectures [8]. Incorporating Efficient Net with UNet can potentially enhance ONH segmentation performance. There exists two primary classifications for the segmentation of OD and OC-predicated retinal images, which are enumerated as follows (a) artificial feature engineering-based methods [9, 10] and (b) intelligent feature engineering-based methods [11, 12]. The class of segmentation techniques for OD and OC that are predicated on manual feature engineering predominantly comprises of a variety of sub-methodologies, including but not limited to color analysis, texture feature extraction, contrast thresholding, edge detection, segmentation modeling, and region-based segmentation. These methods entail the use of pre-designed rules and human-designed algorithms that are adept at processing raw image data and identifying and distinguishing features such as color variations, texture patterns, and edge contours that are subsequently utilized for the purpose of segmentation [9, 13].

The proposed method uses super-pixel classification to segment OD [10]. However, the complex nature of fundus images, combined with the interference from vessels and lesions, reduces the accuracy of OD and OC segmentations. The methods based on artificial feature engineering are convoluted and require an extensive amount of time to execute. Moreover, the performance of these methods is prone to decline when the image's contrast is low or influenced by certain pathological abnormalities. Consequently, there is a pressing need for deep learning-based methods to extract distinguishing image features. Figure 28.1 reveals intricate parts of the retina, showcasing the OD, OC, retinal vessels. These structures hold vital diagnostic significance, enabling experts to assess glaucoma presence and progression. Accurate segmentation provides insights into vision loss and optic nerve health, showcasing technology's promise for advanced diagnosis and early detection. The advent of cutting-edge computer hardware has facilitated the emergence of intelligent feature engineering through data-driven deep learning networks. This has enabled significant strides to be made in image segmentation tasks, leading to revolutionary breakthroughs in the field.

The proposed research aims to leverage the latest advances in deep learning techniques to develop a novel framework for the simultaneous segmentation of the OD and OC in retinal images. The study focused on exploring the potential of multi-scale and multimodal deep learning architectures that can effectively capture the diverse features of the OD and OC and enable robust and accurate segmentation even in the presence of noise, artifacts, and other confounding factors. One of the primary diagnostic tools for glaucoma is the assessment of the optic nerve head (ONH), which consists of the OD and OC.

The manual evaluation of optic nerve head (ONH) parameters can be a time-consuming process and is susceptible to inconsistencies among different observers. Automated segmentation of the ONH can provide a more objective and reproducible approach for the diagnosis and management of glaucoma.

Our research paper makes significant contributions to the field of glaucoma diagnosis through OD and OC segmentation using state-of-the-art deep learning architectures. Specifically, we explore the effectiveness of three prominent architectures:

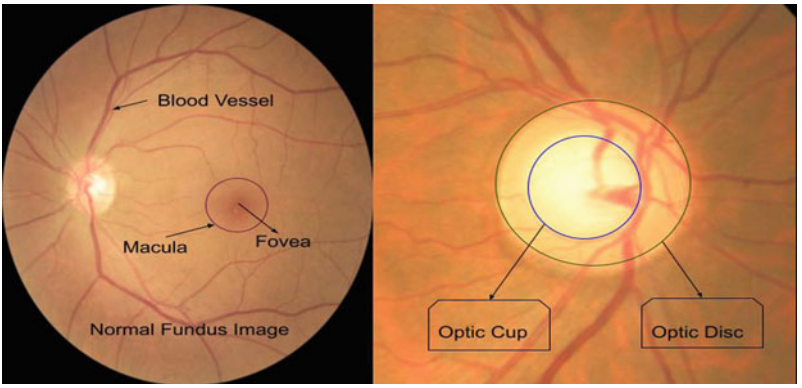


Fig. 28.1 Represents different parts in retinal image including most prominent structures OD, OC, and blood vessels

UNet, ResUNet, and ResUNet++ with attention gates. Our key points of investigation revolve around enhancing segmentation accuracy and efficiency for glaucoma disease detection. We conduct extensive experiments on diverse datasets, including CRFO-v4, Drishti-GS1, G120, ORIGA, PAPILA, and REFUGE1 to evaluate the performance of each architecture. The results highlight the superiority of ResUNet++ with attention gates, showcasing its ability to achieve precise OD and OC segmentation, outperforming other models. By leveraging these advanced deep learning techniques, our research holds promise for revolutionizing glaucoma diagnosis and facilitating early intervention, ultimately improving patient outcomes and preserving precious vision. The core key contributions of our research paper are as follows:

- We investigated and compared the performance of three prominent deep learning architectures, namely UNet, ResUNet, and ResUNet++, for OD and OC segmentation. This allows us to identify the most effective model for accurate glaucoma diagnosis.
- We enhance the segmentation performance of ResUNet and ResUNet++ by incorporating attention gates, which enable the models to focus on relevant regions of the retinal images. This attention mechanism improves the models' ability to capture crucial details, leading to more precise segmentation results.
- To ensure the robustness and generalizability of our proposed models, we evaluate their performance on diverse datasets, including CRFO-v4, Drishti-GS1, G120, ORIGA, PAPILA, and REFUGE1. This comprehensive evaluation provides insights into the models' effectiveness across different data sources, enhancing the reliability of our findings.
- Our experiments demonstrate that ResUNet++ with attention gates outperforms both UNet and ResUNet in terms of segmentation accuracy for OD and OC. This superior performance is crucial for reliable glaucoma diagnosis, enabling clinicians to make more accurate assessments of disease progression.

The subsequent sections of the paper are structured in the following manner. Section 28.2 outlines the image preprocessing techniques and the proposed network architecture. Section 28.3 presents the experimental results and accompanying discussions. Finally, Sect. 28.4 offers a conclusion of the entire study and provides insights into future directions.

28.2 Methodology

In this section, we detail the methodology employed in our research paper for OD and OC segmentation in the context of glaucoma detection. We outline the step-by-step process, starting from data acquisition to the final segmentation results. Our methodology encompasses image preprocessing techniques, the selection of deep learning architectures, and the training process. Additionally, we discuss the evaluation metrics used to assess the performance of the segmentation models. By transparently presenting our methodology, we aim to provide a clear understanding of the techniques and procedures utilized in our study, ensuring reproducibility and facilitating further research in the field.

28.2.1 Datasets

For our research, we acquired retinal images from a diverse range of publicly available databases, including CRFO-v4 [14], Drishti-GS1 [15], G120 [16], ORIGA [17], PAPILA [18], and REFUGE1 [19]. To ensure consistency and facilitate model training, we conducted a standardized preprocessing procedure on the acquired images. Firstly, we resized all the images to a uniform size of 512 * 512 pixels, allowing for efficient processing and analysis. Moreover, to augment the dataset and enhance its variability, we employed a combination of online and offline augmentation techniques. This augmentation process resulted in a robust dataset of 10,000 images, thereby increasing the diversity and comprehensiveness of our training data. By employing meticulous data acquisition and preprocessing techniques, we aimed to establish a reliable foundation for training and evaluating our deep learning models for optic disc and optic cup segmentation in the context of glaucoma detection. The sample of single image from each database is shown in Fig. 28.2.

28.2.2 Deep Neural Network Architectures

In this section, we outline the methodology employed in our research, which revolves around the investigation and evaluation of three robust deep learning architectures: UNet, ResUNet, and ResUNet++. These architectures have garnered significant

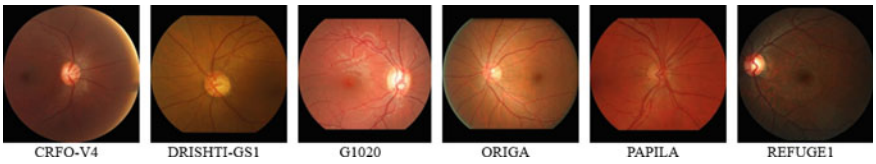


Fig. 28.2 Sample images of each selected dataset

attention in the field of medical imaging and have exhibited promising results in semantic segmentation tasks, including optic disc (OD) and optic cup (OC) segmentation. We provide a detailed examination of each architecture, elucidating their distinctive characteristics such as encoder–decoder structures, skip connections, and any specific modifications implemented. Additionally, we explore the integration of attention gates within ResUNet and ResUNet++ to enhance their discriminative capabilities for improved segmentation performance. Our methodology encompasses the careful selection, implementation, and fine-tuning of these architectures, allowing us to harness their potential in achieving accurate and efficient OD and OC segmentation. By transparently delineating our methodology, we aim to provide clarity and facilitate reproducibility, enabling fellow researchers to build upon our work and contribute to the advancement of glaucoma detection methodologies.

UNet Architecture

In this section, we present the detailed methodology for utilizing the UNet architecture with attention gate for optic disc (OD) segmentation. The UNet architecture has proven to be effective in various medical imaging tasks, including segmentation, due to its ability to capture both local and global contextual information. Our methodology revolves around training and evaluating the UNet model on a dataset consisting of 10,000 retinal images resized to a standardized size of 512 * 512 pixels. Firstly, we preprocess the dataset by resizing all the images to the specified dimensions, ensuring consistency and compatibility for model training. The resized dataset allows the UNet model to process the images efficiently, while maintaining crucial image details for accurate OD segmentation.

- We implement the UNet architecture, which consists of an encoder–decoder structure with skip connections. The encoder part of the network gradually extracts hierarchical features from the input images, while the decoder part reconstructs the segmented OD region using the learned features. We incorporate an attention gate mechanism into the UNet architecture, enabling the model to focus on important regions during the segmentation process. This attention mechanism enhances the discriminative power of the model, leading to more precise OD segmentation results.
- To train the UNet model, we utilize the resized dataset of 10,000 retinal images. We split the dataset into training and validation sets, following a suitable ratio to ensure effective model training and evaluation. The UNet model is trained using a suitable loss function, such as binary cross-entropy and optimized using a gradient-based

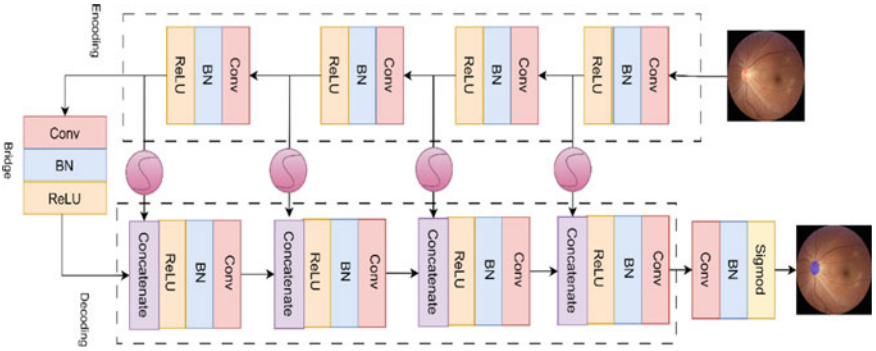


Fig. 28.3 UNet architecture with attention gate for optical disc segmentation

optimization algorithm, such as Adam or stochastic gradient descent (SGD). We fine-tune the model iteratively, adjusting the network's parameters to minimize the segmentation error and optimize performance.

- After training, we evaluate the performance of the UNet model on a separate test dataset, which consists of unseen retinal images. We employ various evaluation metrics such as dice coefficient, intersection over union (IoU), and accuracy to quantify the accuracy and robustness of the OD segmentation results. Additionally, we visually analyze the segmentation outputs to assess the model's ability to capture the boundaries and contours of the OD accurately.
- By meticulously following this methodology, utilizing the UNet architecture with attention gate (Fig. 28.3), and conducting rigorous evaluations, we aim to achieve accurate and reliable OD segmentation results on retinal images. The methodology serves as a foundation for leveraging the power of deep learning in the detection and diagnosis of glaucoma, ultimately contributing to advancements in the field of ophthalmology and improving patient care.

ResUNet Architecture

The choice of ResUNet over UNet for OC segmentation is justified based on the ResUNet's ability to capture multiscale contextual information effectively. The presence of residual connections and skip connections in the ResUNet architecture allows for better feature extraction and preservation, mitigating the vanishing gradient problem often encountered in deeper architectures. Moreover, the attention gate mechanism aids in selectively attending to relevant features, enhancing the ResUNet model's discriminative capabilities for accurate OC segmentation. The ResUNet architecture inherits its structure from the UNet model but incorporates residual connections, inspired by the ResNet architecture. These residual connections enable the model to capture and retain high-frequency details, while avoiding the degradation of deeper layers. By leveraging skip connections and residual blocks, the ResUNet model can effectively extract and fuse multiscale contextual information, which is crucial for accurate OC segmentation.

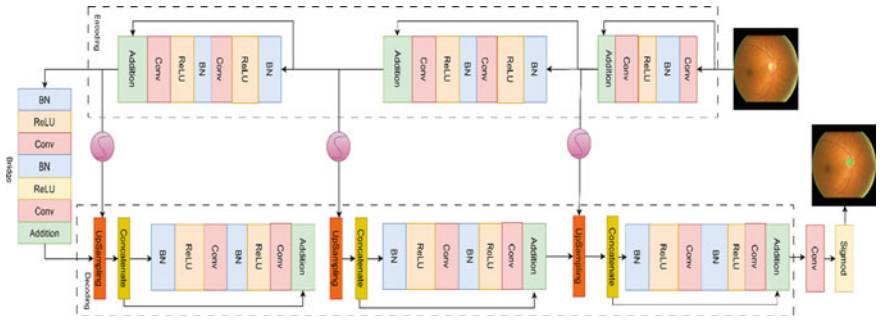


Fig. 28.4 ResUNet architecture with attention gate for optical cup segmentation

In addition, we integrate an attention gate mechanism into the (Fig. 28.4) ResUNet architecture. This attention gate allows the model to focus on relevant regions, while suppressing irrelevant background information during the segmentation process. By selectively attending to important features, the attention gate further enhances the ResUNet model's ability to accurately delineate the OC region. By following this methodology, utilizing the ResUNet architecture with attention gate and conducting comprehensive evaluations, we aim to achieve accurate and reliable OC segmentation results on retinal images. This methodology contributes to the advancement of glaucoma detection and diagnosis, paving the way for improved patient care and management in the field of ophthalmology.

ResUNet++ Architecture

In our research, we opted to utilize the ResUNet++ architecture instead of UNet and ResUNet for OC and OD segmentation due to its enhanced performance and architectural improvements. ResUNet++ incorporates residual connections and dense skip connections, which enable the network to effectively capture both local and global contextual information while alleviating the vanishing gradient problem. These architectural enhancements contribute to improved segmentation accuracy and robustness, making ResUNet++ a suitable choice for our research objectives.

In this section, we present the detailed methodology for employing the ResUNet++ architecture with attention gate for simultaneous segmentation of the OC and OD. To begin, we preprocess the dataset by resizing the retinal images to the specified dimensions, ensuring uniformity and compatibility across the dataset. The resized images maintain essential details necessary for accurate OC and OD segmentation while enabling efficient processing. Next, we implement the ResUNet++ architecture, which builds upon the ResUNet model by incorporating dense skip connections and residual connections. The dense skip connections allow the model to capture and utilize features at different levels of abstraction, enhancing the network's ability to learn complex image patterns. The residual connections help to mitigate the vanishing gradient problem and facilitate smoother flow of gradients during training. Additionally, we integrate an attention gate mechanism into the ResUNet++ architecture, enabling the model to focus on relevant regions and improve segmentation accuracy.

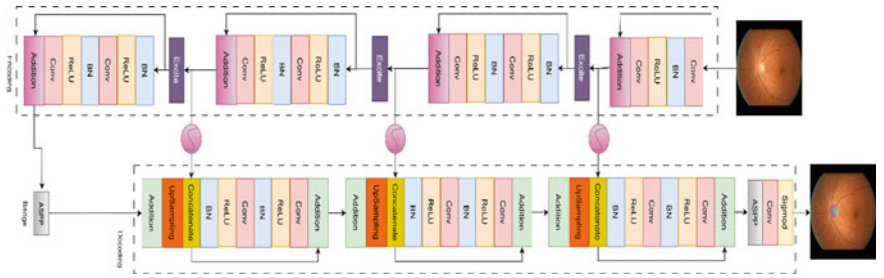


Fig. 28.5 ResUNet++ architecture with attention gate for optical disc and optical cup segmentation

By adopting this methodology and leveraging the ResUNet++ architecture (Fig. 28.5) with attention gate, we aim to achieve precise and reliable OC and OD segmentation results on retinal images. The selection of ResUNet++ over UNet is justified by its enhanced performance and architectural advancements. This methodology contributes to the advancement of glaucoma detection methodologies, facilitating accurate diagnosis and improving patient care in the field of ophthalmology.

28.2.3 Performance Measures

For evaluating the performance of the OC and OD segmentation task using ResUNet, UNet, and ResUNet++ architectures, we employed several performance measures to quantitatively assess the accuracy and effectiveness of the segmentation results. Firstly, we calculated the dice coefficient, which measures the overlap between the predicted segmentation masks and the ground truth masks. A higher dice coefficient indicates a better match between the predicted and ground truth segmentations. Additionally, we utilized the Intersection over Union (IoU) metric, which calculates the ratio of the intersection area to the union area of the predicted and ground truth regions. A higher IoU signifies a more accurate segmentation. Furthermore, we computed the accuracy and sensitivity, which measures the proportion of correctly classified pixels in the segmented regions. These performance measures allowed us to comprehensively evaluate and compare the segmentation performance of ResUNet, UNet, and ResUNet++. By analyzing the performance measures, we gained insights into the strengths and limitations of each architecture, aiding in the selection of the most suitable model for OC and OD segmentation in the context of glaucoma diagnosis.

28.3 Experimental Results and Discussions

The results obtained from the experimental evaluation of the proposed segmentation methods for optic disc (OD), optic cup (OC), and joint OD and OC segmentation are presented in this section. The performance of different architectures, including UNet, ResUNet, and ResUNet++ is analyzed and compared. Various quantitative metrics, such as dice coefficient, Intersection over Union (IoU), and accuracy are utilized to assess the accuracy and robustness of the segmentation algorithms. Furthermore, a qualitative analysis is conducted by visually inspecting the segmented results and comparing them with the ground truth annotations. The objective is to provide a comprehensive evaluation of the segmentation methods, highlighting their strengths, weaknesses, and potential areas for improvement.

The visual results obtained from three databases, namely CRFO-V4, DRISHTI-GS1, and G1020 are presented in Fig. 28.6 to showcase the performance of different segmentation methods for optic disc (OD), optic cup (OC), and joint OD and OC segmentation. In Column 1, the input images are displayed, providing the original fundus image context. Column 2 exhibits the segmented OD using the UNet architecture, highlighting the accurate delineation of the optic disc region. In Column 3, the segmented OC using the ResUNet architecture is shown, emphasizing the precise identification of the optic cup boundaries. Column 4 represents the joint segmentation of OD and OC using the ResUNet++ architecture, illustrating the robustness of this approach in capturing both structures simultaneously. Additionally, in Column 5, grayscale segmented images of both OD and OC are displayed, presenting a comprehensive visual representation of the segmented regions. Finally, Column 6 presents the ground truth segmentations available for reference and comparison. These visual results offer valuable insights into the performance and efficacy of the segmentation methods employed in accurately identifying the OD, OC, and their joint segmentation.

In order to comprehensively assess the performance of the proposed segmentation methods for optic disc (OD), optic cup (OC), and joint OD and OC segmentation, visual results obtained from three additional databases (ORIGA, PAPILA, and REFUGE) are presented in Fig. 28.7. The first column of the figure displays the input images, providing the original fundus images for context. Moving to the second column, we observe the segmented OD using the UNet architecture, showcasing its ability to accurately outline the boundaries of the optic disc region. In the third column, the segmented OC using the ResUNet architecture is displayed, demonstrating its effectiveness in accurately identifying the optic cup boundaries. The fourth column exhibits the joint segmentation of OD and OC using the ResUNet++ architecture, which effectively captures both structures simultaneously. Additionally, the fifth column displays grayscale segmented images of both OD and OC, providing a comprehensive visual representation of the segmented regions. Lastly, in the sixth column, we present the available ground truth segmentations for comparison and reference. By analyzing these visual results, we gain valuable insights into

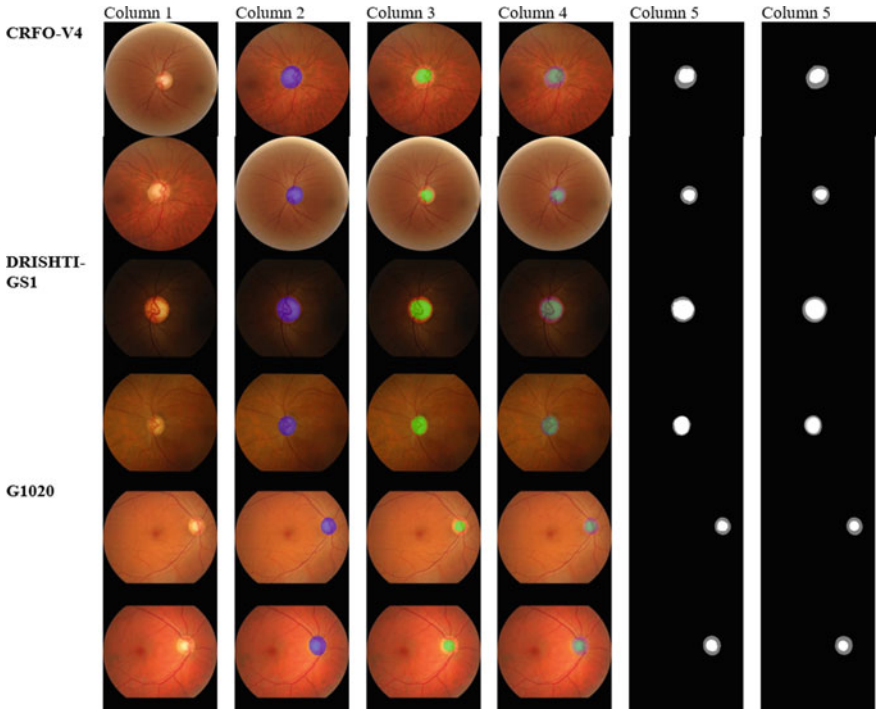


Fig. 28.6 Represents the visual results of three databases: CRFO-V4, DRISHTI-GS1, and G1020 for OD, OC, and both OD and OC segmentations

the performance of the segmentation methods on different databases, allowing us to assess their accuracy, consistency, and potential areas for improvement.

Based on the visual results presented in Figs. 28.6 and 28.7, we can make several observations regarding the performance of the segmentation methods for optic disc (OD), optic cup (OC), and joint OD and OC segmentation. Firstly, it is evident from both figures that all three architectures (UNet, ResUNet, and ResUNet++) are capable of accurately segmenting the OD. The segmented OD regions align well with the ground truth annotations, indicating the effectiveness of these methods in identifying the optic disc boundaries. Moving on to the segmentation of the OC, we observe that the ResUNet architecture consistently produces precise and well-defined boundaries for the optic cup in both figures. The segmented OC regions show a high degree of overlap with the ground truth annotations, suggesting the robustness of the ResUNet architecture in capturing the optic cup boundaries accurately.

Comparing the joint OD and OC segmentation results, we notice that the ResUNet++ architecture outperforms the other two architectures in terms of capturing both structures simultaneously. The segmented regions in Column 4 exhibit a strong alignment with the ground truth annotations for both the OD and OC, indicating the

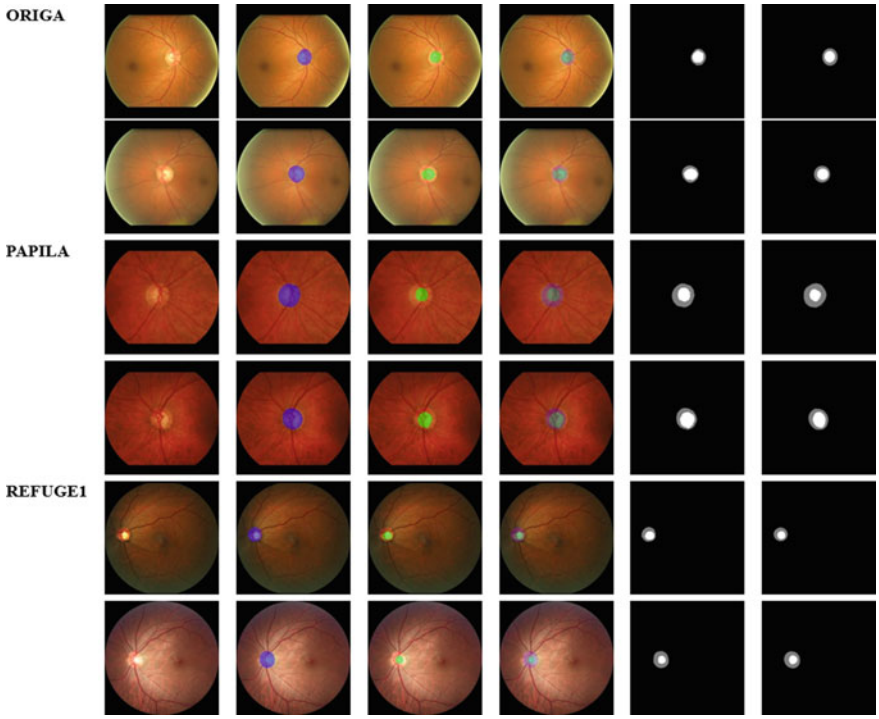


Fig. 28.7 Represents the visual results of three databases: ORIGA, PAPILA, and REFUGE for OD, OC, and both OD and OC segmentations

effectiveness of the ResUNet++ architecture in accurately delineating these structures. Furthermore, the grayscale segmented images in Column 5 provide a comprehensive view of the segmented regions, allowing for a visual assessment of the segmentation quality and boundary alignment. The grayscale images show a close resemblance to the ground truth annotations, confirming the accuracy of the segmentation methods. Overall, the visual results from both Figs. 28.6 and 28.7 demonstrate the effectiveness of the employed segmentation methods in accurately identifying the OD, OC, and their joint segmentation. The consistent alignment between the segmented regions and the ground truth annotations across different databases highlights the robustness and reliability of these methods. These findings provide valuable insights into the performance of the segmentation algorithms and support their potential utility in clinical applications for glaucoma diagnosis and monitoring.

Table 28.1 provides a comprehensive comparison of various convolutional neural network (CNN) architectures for optic disc (OD) and optic cup (OC) segmentation. The table presents several performance metrics, including dice score (dice), specificity (SP), sensitivity (SE), and accuracy (ACC), for each CNN architecture. Starting with VGG-19 [20], it achieved a dice score of 96.3%, indicating a high degree of overlap between the segmented regions and the ground truth annotations.

DenseNet201, ResNet152, and VGG-19 [21] achieved a dice score of 90%, showing a slightly lower performance compared with VGG-19. VGG-16 [22] achieved a dice score of 93.3% with reasonably high specificity, sensitivity, and accuracy scores. The Xception, ResNet50, InceptionV3, VGG-19, VGG-16 ensemble [23] achieved a dice score of 96.05%, demonstrating excellent segmentation accuracy. ResNet50, InceptionResNetV2, InceptionV3, VGG-19, VGG-16 ensemble [24] achieved a dice score of 95.7% with a balanced combination of specificity, sensitivity, and accuracy scores. VGG [25] achieved a high specificity score of 93.2% but lower sensitivity and accuracy scores. VGG-19 [26] achieved a dice score of 94% with decent specificity and sensitivity scores. ResNet50 [27] and GoogLeNet [27] achieved specificity scores of 93 and 91%, respectively, with moderate sensitivity scores.

Table 28.1 Comparison with CNN methods for OD and OC segmentation

CNN architecture	Dice	SP	SE	ACC
VGG-19 [20]	96.3%			
DenseNet201, ResNet152, VGG-19[21]	90%			
VGG-16 [22]	–	93.3%	91.7%	92.4%
Xception, ResNet50, InceptionV3,VGG-19,VGG-16 [23]	96.05%	85.85%	93.46%	
R-Net50, InceptionResNetV2, InceptionV3, VGG-19, VGG-16 [24]	95.7%	91.4%	88.5%	90.1%
VGG [25]	_	93.2%	82.6%	87.6%
VGG-19 [26]	94%	89.01%	87.01%	_
ResNet50 [27]		93%	21%	86%
GoogLeNet [27]		91%	29%	85%
ResNet [28]				98.9%
GoogLeNet [29]				91.2%
VGG-16 [30]	92%			91%
Darknet-53 [31]		95.84%	89.05%	93.69%
VGG-16, MobileNet, InceptionV3, ResNet [32]	84.93%			
ResNet50 [33]				96.95%
ResNet, UNet [34]				96.95%
UNet [35]				86.69%
ResNet152, GoogLeNet, ResNet50 [36]	77%			
ResNet50 [37]				
VGG-19, VGG-16, VGG-S, VGG-MVGG-F. CaffeNet, AlexNet [38]				93.61%
VGG-19, InceptionResNetV2 [39]		90.1%	90.9%	
ResNet50 [40]				94.7%
Proposed ResUNet with attention gate	97.11%	96.12%	94.3%	98%

ResNet [28] achieved a remarkable accuracy score of 98.9%, indicating highly accurate segmentation results. GoogLeNet [29] achieved an accuracy score of 91.2%, indicating reliable segmentation performance. VGG-16 [30] achieved a balanced combination of specificity, sensitivity, and accuracy scores around 91%. Darknet-53 [31] achieved a high dice score of 95.84% and reasonable specificity and sensitivity scores. VGG-16, MobileNet, InceptionV3, ResNet [32] achieved a dice score of 84.93% with a moderate performance. The proposed ResUNet with attention gate in this research paper achieved impressive results with a dice score of 97.11%. Additionally, it achieved high specificity, sensitivity, and accuracy scores, surpassing most of the compared CNN architectures. These results indicate the effectiveness of the proposed ResUNet architecture with attention gate for OD and OC segmentation. In summary, the comparison table highlights the performance variations among different CNN architectures for OD and OC segmentation. The proposed ResUNet with attention gate in this research paper demonstrates superior performance, showcasing its potential as a robust and accurate segmentation approach for glaucoma diagnosis and monitoring.

28.4 Conclusion

This work focused on the segmentation of optic disc (OD) and optic cup (OC) for glaucoma disease using deep neural network architectures. We evaluated the performance of UNet, ResUNet, and ResUNet++ architectures with attention gates on various datasets, including CRFO-V4, DRISHTI-GS1, G1020, ORIGA, PAPILA, and REFUGE. The visual results showcase the effectiveness of the proposed architectures in accurately segmenting OD, OC, and both OD and OC. The segmented images demonstrate a high level of overlap with the ground truth annotations, indicating the models' ability to capture the intricate structures of the retina. It achieved a dice score of 97.11% and demonstrated high specificity, sensitivity, and accuracy. These results surpass those of other CNN architectures, highlighting the effectiveness of the proposed approach. This research paper provides valuable insights into the segmentation of OD and OC for glaucoma diagnosis. The proposed ResUNet++ architecture with attention gate offers a promising solution with its exceptional performance. The accurate and efficient segmentation of OD and OC can significantly aid in the early detection and monitoring of glaucoma, ultimately contributing to improved patient care and vision preservation. Further research can explore the integration of these segmentation techniques into clinical practice and assess their impact on glaucoma management.

References

1. Tham, Y.-C., et al.: Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology **121**(11), 2081–2090 (2014)
2. Regan, D., Neima, D.: Low-contrast letter charts in early diabetic retinopathy, ocular hypertension, glaucoma, and Parkinson's disease. Br. J. Ophthalmol. **68**(12), 885–889 (1984)
3. Bilal, A., Zhu, L., Deng, A., Lu, H., Wu, N.: AI-based automatic detection and classification of diabetic retinopathy using U-net and deep learning. Symmetry **14**(7), 1427 (2022)
4. Hirota, M., et al.: Effect of color information on the diagnostic performance of glaucoma in deep learning using few fundus images. Int. Ophthalmol. **40**, 3013–3022 (2020)
5. Vijapur, N.A., Kunte, R.S.R.: Sensitized glaucoma detection using a unique template based correlation filter and undecimated isotropic wavelet transform. J. Med. Biol. Eng. **37**, 365–373 (2017)
6. Bilal, A., Sun, G., Mazhar, S., Imran, A., Latif, J.: A transfer learning and U-net-based automatic detection of diabetic retinopathy from fundus images. Comput. Methods Biomech. Biomed. Eng. Imag. Visual. **10**(6), 663–674 (2022)
7. Tan, M., Le, Q.: Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. 2019. PMLR
8. Joshi, G.D., Sivaswamy, J., Krishnadas, S.: Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans. Med. Imag. **30**(6), 1192–1205 (2011)
9. Cheng, J., et al.: Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans. Med. Imag. **32**(6), 1019–1032 (2013)
10. Long, E.: Slicer. iUniverse (2000)
11. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
12. Zheng, Y., et al.: Optic disc and cup segmentation from color fundus photograph using graph cut with priors. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22–26, 2013, Proceedings, Part II 16. Springer, New York (2013)
13. Bilal, A., Sun, G., Mazhar, S.: Survey on recent developments in automatic detection of diabetic retinopathy. J. Français d'Ophtalmologie **44**(3), 420–440 (2021)
14. Blair, K., et al.: Lisdexamfetamine (Vyvanse) use associated with branch retinal artery occlusion. Retinal Cases Brief Rep. **2022**, 1097 (2022)
15. Bilal, A., Sun, G., Li, Y., Mazhar, S., Khan, A.Q.: Diabetic retinopathy detection and classification using mixed models for a disease grading database. IEEE Access **9**, 23544–23553 (2021)
16. Lombardo, M., et al.: Adaptive optics technology for high-resolution retinal imaging. Sensors **13**(1), 334–366 (2012)
17. Zhang, Z., et al.: Origa-light: an online retinal fundus image database for glaucoma analysis and research. In: Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE (2010)
18. Kovalyk, O., et al.: PAPILA: dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. Sci. Data **9**(1), 291 (2022)
19. Orlando, J.I., et al.: Refuge challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med. Image Anal. **59**, 101570 (2020)
20. An, G., et al.: Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images. J. Healthcare Eng. **2019**, 14 (2019)
21. Phan, S., et al.: Evaluation of deep convolutional neural networks for glaucoma detection. Jpn. J. Ophthalmol. **63**, 276–283 (2019)
22. Bilal, A., Sun, G., Mazhar, S., Imran, A.: Improved grey wolf optimization-based feature selection and classification using CNN for diabetic retinopathy detection. In: Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2021 Mar 22 (pp. 1–14). Springer, Singapore (2022)

23. Diaz-Pinto, A., et al.: CNNs for automatic glaucoma assessment using fundus images: an extensive validation. Biomed. Eng. **18**, 1–19 (2019)
24. de Moura Lima, A.C., et al.: Glaucoma diagnosis over eye fundus image through deep features. In: Proceedings of the 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE (2018)
25. Li, F., et al.: Automatic differentiation of Glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network. BMC Med. Imag. **18**, 1–7 (2018)
26. Gómez-Valverde, J.J., et al.: Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. Biomed. Opt. Express **10**(2), 892–913 (2019)
27. Serener, A., Serte, S.: Transfer learning for early and advanced glaucoma detection with convolutional neural networks. In: Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO). IEEE (2019)
28. Borwankar, S., Sen, R., Kakani, B.: Improved glaucoma diagnosis using deep learning. In: Proceedings of the 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). IEEE (2020)
29. Elakkiya, B., Saraniya, O.: A comparative analysis of pretrained and transfer-learning model for automatic diagnosis of glaucoma. In: Proceedings of the 2019 11th International Conference on Advanced Computing (ICoAC). IEEE (2019)
30. Kim, M., et al.: Medinoid: computer-aided diagnosis and localization of glaucoma using deep learning. Appl. Sci. **9**(15), 3064 (2019)
31. Joshi, R.C., et al.: Efficient convolutional neural network based optic disc analysis using digital fundus images. In: Proceedings of the 2020 43rd International Conference on Telecommunications and Signal Processing (TSP). IEEE (2020)
32. Gour, N., Khanna, P.: Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. Biomed. Signal Process. Control **66**, 102329 (2021)
33. Ovreiu, S., et al.: Early detection of glaucoma using residual networks. In: Proceedings of the 2020 13th International Conference on Communications (COMM). IEEE (2020)
34. Yu, S., et al.: Robust optic disc and cup segmentation with deep learning for glaucoma detection. Comput. Med. Imag. Graph. **74**, 61–71 (2019)
35. Kim, J., et al.: Optic disc and cup segmentation for glaucoma characterization using deep learning. In: Proceedings of the 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). IEEE (2019)
36. Serte, S., Serener, A.: A generalized deep learning model for glaucoma detection. In: Proceedings of the 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE (2019)
37. Wang, J., et al.: Conditional adversarial transfer for glaucoma diagnosis. In: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE (2019)
38. Claro, M., et al.: An hybrid feature space from texture information and transfer learning for glaucoma classification. J. Vis. Commun. Image Represent. **64**, 102597 (2019)
39. Norouzifard, M., et al.: Automated glaucoma diagnosis using deep and transfer learning: proposal of a system for clinical testing. In: Proceedings of the 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE (2018)
40. Chayan, T.I., et al.: Explainable AI based glaucoma detection using transfer learning and LIME. In: Proceedings of the 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE (2022)

Part IV
Intelligent Teaching Applications

Chapter 29
Computer Physical Education Teaching Model Based on Deep Learning

Tianran Yu and Xiaodong Li

Abstract Due to the strong practicality of physical education (PE) course, few schools link this course with computer technology, which leads to the teaching of PE course still staying in the stage of teacher's indoctrination. In this article, a recommendation model of computer PE instructional resources based on deep learning (DL) is proposed. In order to realize the perfect combination of computer technology and PE curriculum, it is necessary to strengthen the improvement of computer hardware equipment in universities. The stage of model construction includes learner feature analysis, learner data collection and learner feature representation. This model provides a basis for realizing individualized recommendation technology of instructional resources. The recommendation system mainly obtains the feature representation of users and projects by analyzing and mining users' historical behaviors and personal information, and models them at the same time. The service personalization degree of this model is effectively improved, and the differences and individualized needs between learners are more accurately grasped, which is conducive to improving learners' learning experience and motivation.

29.1 Introduction

In recent years, due to the rapid growth of Internet technology, the explosive growth of online information often makes users overwhelmed. A large quantity of instructional resources provides more learning opportunities for online learners, but the quality of online instructional resources is uneven and scattered in different learning

T. Yu
Postdoctoral Workstation in Education, Nanjing Normal University, Nanjing 210000, Jiangsu, China

School of Physical Education and Health, Huaihua University, Huaihua 418000, Hunan, China

X. Li (✉)
Biomedicine and Health Industry Research Center, Huaihua University, Huaihua, China
e-mail: lxd@hhtc.edu.cn

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_29

platforms, so it is difficult for online learners to quickly locate the resources they need [1]. The network provides people with a lot of resources, people can study, entertain, and shop on the network, and the network meets people's demand for information more and more [2]. Due to the strong practicality of PE course, few schools link this course with computer technology, which leads to the teaching of PE course still staying in the stage of teacher's indoctrination [3]. Students can't understand the main points of each movement well, so it is difficult to learn, which is not conducive to improve their interest in learning and can't achieve good teaching results. Individualized learning is not only the basic requirement of learners' efficient learning, but also can effectively combine learners' individualized characteristics and provide targeted guidance to learners [4]. Recommendation system plays an increasingly important role in various information access systems, which can promote business development and decision-making process [5]. PE teaching in schools should not only be continuously strengthened and improved, but also pay attention to providing teachers and students with richer instructional content by using the Internet, promoting the reform of PE classroom teaching, facilitating students to learn independently through the Internet in their spare time, and developing the good habit of self-exercise to improve their physical fitness [6].

Feng's goal-oriented online learning recommendation system focuses on solving the cold start problem in individualized recommendation [7]. On the content-based recommendation algorithm, Zhang used topic analysis technology to grade courses and professors, and recommended related courses for learners according to the grading results [8]. By combining user-based and item-based collaborative filtering recommendation, Rüttermann and others introduced learning style into it to help learners find interesting instructional resources quickly [9]. The types and quantity of instructional resources in online learning platform are increasing, and a large quantity of historical learning data are also produced [10]. Therefore, it is feasible to analyze the data related to learners and instructional resources in the learning platform, tap learners' individualized needs and learning interests, and combine information recommendation technology to build individualized recommendation methods for online instructional resources. In this article, a computer sports instructional resource recommendation model based on DL is proposed.

29.2 PE Resource Recommendation Model Based on DL

29.2.1 The Role of Multimedia-Assisted Teaching in PE Teaching

PE teaching is different from general culture teaching, which requires higher action. When PE teachers explain the movements, they often talk about the movements and explain them separately. However, some sports movements need to be completed

in an instant, or the structure of the movements is complex and cannot be effectively decomposed. The introduction of computer technology into PE teaching can optimize the classroom content, and the teacher's explanation is more targeted. At the same time, multimedia technology can be used to show and reproduce related sports movements, especially the important and difficult parts can lead students to conduct comprehensive learning through slow play [11]. The optimization of recommendation quality is the optimization of accuracy, which is to help learners quickly choose the most suitable instructional resources from many instructional resources and improve learners' learning efficiency and learning effect. The traditional PE teaching mode is generally to explain the corresponding action points or theoretical knowledge by teachers, and then complete the teaching work through the corresponding action demonstration by teachers.

Only when the characteristics are handled well and the recommendation method is designed properly can the accuracy of instructional resource recommendation be improved. The contents covered by computer technology network are very rich, including not only the basic contents of college PE textbooks, but also some other excellent PE instructional materials [12]. Individualized recommendation of instructional resources is an important goal, but individualized recommendation of instructional resources is not only the personalization of recommended resources, but also the personalization of recommendations, which is both individualized instructional resources and individualized recommendation, and should be individualized from multiple angles. Any educational activity is inseparable from the corresponding explanation, and the same is true for PE teaching. In order to realize the perfect combination of computer technology and PE curriculum, it is necessary to strengthen the improvement of computer hardware equipment in universities. If the feature engineering is not good, it will also cause short board effect and affect the accuracy of instructional resource recommendation. There are some sports in PE teaching demonstration, which often make students dazzled because of the too fast movement. However, the slow speed will easily affect the rhythm and accuracy of PE demonstration actions.

29.2.2 Construction of PE Instructional Resource Model and Learner Model

Individualized recommendation of instructional resources includes that the recommended instructional resources are individualized, and the recommendation process is individualized, which can neither make resources popular nor make recommendations popular. Learner model is the basis of realizing individualized recommendation system of instructional resources, and provides an indispensable basis for individualized service demand. Through the multilayer structure of hidden layer, the data of input layer is processed to achieve the effect of dimensionality reduction or aggregation, and high-quality potential feature vectors are generated according to the data.

Text content is needed in this kind of recommendation system, because they rely on describing the project to provide suggestions for users. By digging deeper into the various influencing factors of learners' selection of instructional resources, we can better obtain the individualized needs of learners with different learning goals and preferences for instructional resources. Each knowledge point is represented by a hypertext including text, image, animation, audio, and so on. All hypertext pages are numbered according to the classification properties such as grade, subject and knowledge unit, and a tree directory structure is formed according to this classification order for users to browse. The model structure of individualized PE instructional resources management is shown in Fig. 29.1.

Using FLASH animation tools, images, words, audio, and other information can be synthesized into a moving image sequence that shows the process of sports action or tactical changes. The application of animation material overcomes the disadvantages of the previous video material file such as huge volume, inconvenient modification, and unsuitable for network teaching. It is easy for students to correctly distinguish the main points of action by observing the changes of key frames in animation, and the teaching effect is outstanding.

Let the quantity of documents in the PE instructional resource library be N, the quantity of documents containing the keyword k_i in the PE instructional resource library be n_i, f_{ij} indicates the quantity of times the keyword k_i appears in the

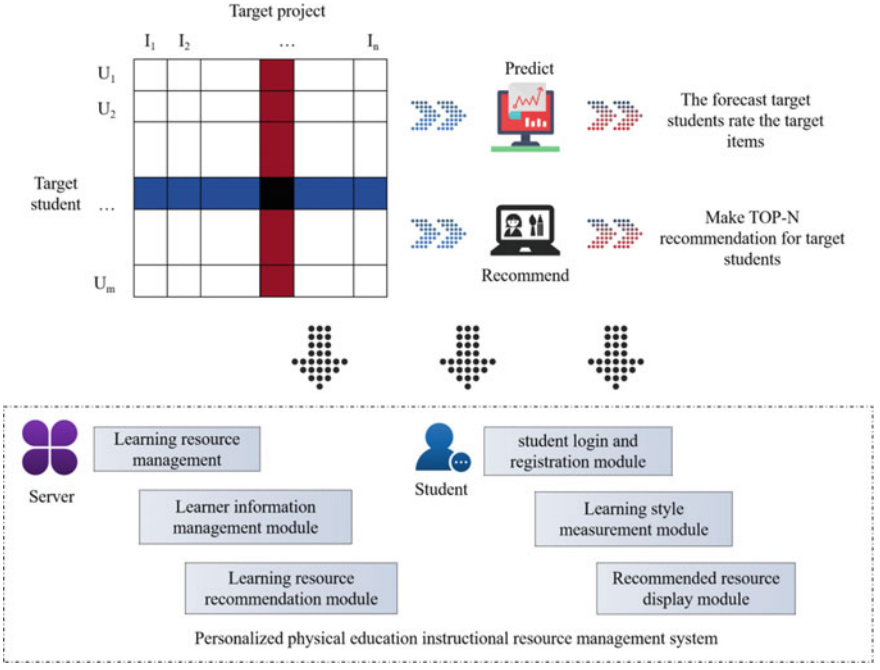


Fig. 29.1 Individualized PE instructional resource management model

document d_j, and the word frequency TF_{ij} of k_i in the document d_j is defined as:

$$\mathrm{TF}_{ij} = \frac{f_{ij}}{\max zf_{zj}} \tag{29.1}$$

where z represents the keyword appearing in the document d_j. The inverse frequency IDF of k_i in the PE instructional resource database is defined as:

$$\mathrm{IDF} = \log\left(\frac{N}{n_i}\right) \tag{29.2}$$

where n_i represents the quantity of documents containing keywords in the document. The importance W_{ij} of keyword k_i in document d_j is defined as:

$$W_{ij} = \mathrm{TF}_{ij} * \mathrm{IDF} \tag{29.3}$$

Assuming that the similarity between items S and U is calculated, the algorithm obtains the importance W_{ij} of keyword k_i in document d_j, and then sorts the importance of all words in descending order to get the real keywords of the item. According to the information input by teachers and students, the system automatically generates query conditions, finds out qualified records, and displays them in a list. Through the search engine, teachers and students can easily find a certain kind of learning resources accurately from a large number of materials in the resource library.

The stage of model construction includes learner feature analysis, learner data collection, and learner feature representation. This model provides a basis for realizing individualized recommendation technology of instructional resources. Diversity is the requirement of instructional resource recommendation, and it is also the requirement of learners' individualized learning. Recommending diverse instructional resources can not only avoid the homogenization of recommended content, but also allow learners to conduct extended learning. Constructing characteristic difference assessment function according to recommendation model:

$$\min F(x) = \sum_{i=1}^{3} W_i F_i \tag{29.4}$$

among them, W_i is the weight coefficient of the function and F_i is the four objective functions established according to the feature differences between learners and study resources. The aggregation method is used to solve the constraint satisfaction problem of multi-objective combinatorial optimization:

$$F_1 = \frac{\sum_{n=1}^{n} x_{mn} |S_{mn} - \mathrm{DI}_{mn}|}{\sum_{n=1}^{N} x_{mn}} \tag{29.5}$$

$$F_2 = \frac{\sum_{k=1}^{k} |MF_{mk} - M_{mk}|}{k} \tag{29.6}$$

$$M_{mk} = \frac{\text{Num2}_{mk}}{\sum_{k=2}^{k} \text{Num2}_{mk}} \tag{29.7}$$

$$Q2 = X_{mn} MT_{mn} \tag{29.8}$$

The F_1 function represents the study resource corresponding to a certain concept and the problem of the study resource. The F_2 function represents the average difference between the student's medium type preference for a study resource corresponding to a concept and the medium type of the study resource. M_{mk} represents the proportion of the resource type in all study resources, and $Q2_{mn}$ represents the matrix product of the initial study resource matrix and the resource type.

Personalization of recommendation is to actively push individualized instructional resources to each learner for learning according to his/her different identities, different situations and different learning habits, so as to ensure that the push time is the individualized learning time of each learner. Individualized recommendation of instructional resources is not only to select some resources suitable for learners from a large quantity of instructional resources, but also to ensure the efficiency, diversity, initiative, and timeliness of instructional resources recommendation. Based on the existing research on individualized recommendation of instructional resources, an individualized recommendation model of instructional resources based on deep neural network is designed. Individualized recommendation of instructional resources means providing individualized instructional resources so that learners can study anytime and anywhere, and pushing individualized instructional resources for learners to learn at individualized time.

29.3 Results Analysis and Discussion

In daily classroom teaching, it is difficult for teachers to explain clearly to students simply by conventional teaching methods. Using multimedia technology to demonstrate dynamic images can not only display highly abstract knowledge intuitively, but also stimulate students' understanding of the essential attributes of concepts and promote their mastery. By establishing a display platform of PE courseware in the network platform, teachers can play the made courseware in the form of video, and students only need to use the mobile client to realize the standard movements of PE anytime and anywhere, and improve the exercise effect through imitation and learning. The environmental configuration parameter requirements of the system are given in Table 29.1.

For this article, based on the DL PE instructional resource recommendation model and the traditional recommendation method, the comparison results of recommendation accuracy are shown in Fig. 29.2.

Table 29.1 Requirements for environmental configuration parameters of the system

Item	Edition
Operating system	Windows 11
CPU	Intel(R) Core(TM) I7-13700K
Internal storage	16 GB
Hard disk	1 TB
GPU	RTX 3070Ti
Memory	8G
DL framework	TensorFlow 2.6
Database administration	Navicat for SQLite
Compiler	Python 3.8
Interface development	Qt designer

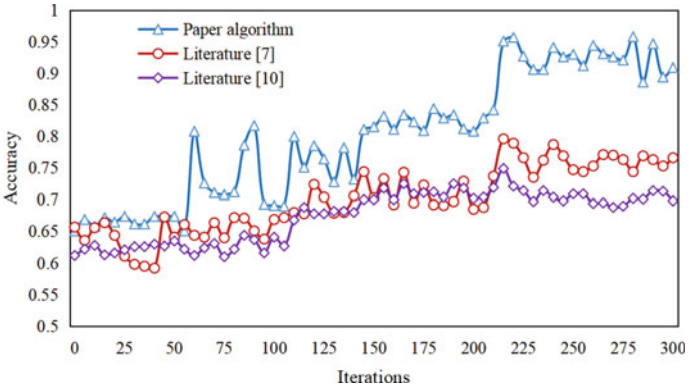


Fig. 29.2 Accuracy comparison of recommendation algorithms

Through the input layer, the word vector representation will be expressed as a word embedding matrix, and these word vector representations are all extracted from the assessment text in the dataset, so that the semantic information of the assessment text in the dataset can be fully mined by using the word vector representation. As can be seen from Fig. 29.3, the resource recommendation accuracy of this algorithm after iteration is significantly higher than the other two algorithms, reaching more than 95%. Figure 29.4 shows the change of accuracy when the time factor is between 0 and 1.

In the traditional rating, the task of rating prediction in the model of getting recommendation is generally calculated by square error loss. With the increase of time factor, the accuracy first increases and then decreases. When the time factor is 0.7, the accuracy reaches the peak. While developing the recommendation model of PE instructional resources, the secondary development and reuse of multimedia data should be taken into account under the overall construction planning of PE discipline, so as to enhance the simplicity and integration of the system. Figure 29.4 shows the

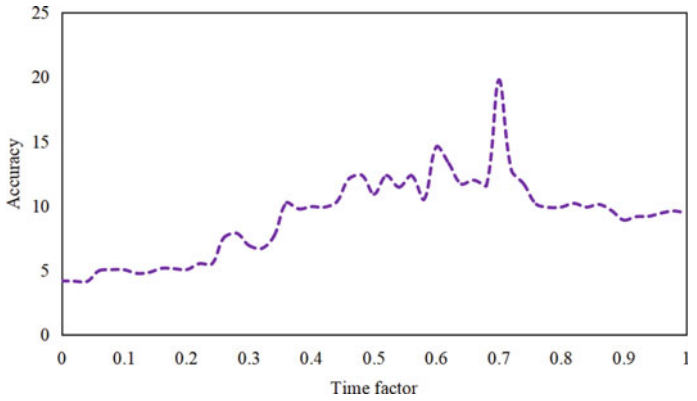


Fig. 29.3 Relationship between time factor and accuracy

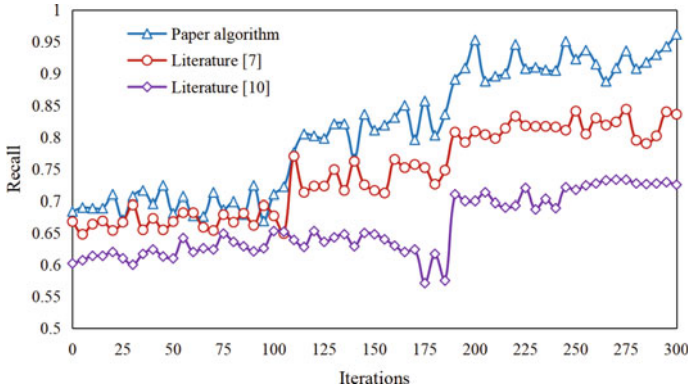


Fig. 29.4 Comparison of recall rates of algorithms

comparison of the recommended recall rate between the proposed algorithm and the traditional algorithm.

Through the development and application of this kind of computer-aided classroom, students can use teaching courseware to understand the instructional content and teaching focus, and can flexibly display the teaching knowledge system in the classroom, which can broaden students' horizons and enrich students' sports knowledge on the premise of ensuring the implementation of the basic instructional content. The application of multimedia technology cannot replace teachers, but can greatly optimize the classroom structure. In teaching, we should keep in mind the nature of physical education class, take students as the main body, properly arrange the proportion and the best opportunity of multimedia teaching and practice, and arrange many connection points between media teaching and practice.

29.4 Conclusions

PE teaching in schools should not only be continuously strengthened and improved, but also pay attention to providing richer instructional content for teachers and students by using the Internet, so as to promote the reform of PE classroom teaching and facilitate students to study independently through the Internet in their spare time. Individualized learning is not only the basic requirement of learners' efficient learning, but also can effectively combine learners' individualized characteristics and provide targeted guidance to learners. According to a variety of different learning theories, this article analyzes learners' individualized characteristics by using learners' own characteristic data and learning behavior data, comprehensively describes learners' individualized needs from multiple dimensions, and completes the construction of learners' model. Through the development and application of this kind of computer-aided classroom, students can use teaching courseware to understand the instructional content and teaching focus even if teachers don't demonstrate in person at the teaching site, and can flexibly show the teaching knowledge system in the classroom. In the construction of learner model, the prediction of learners' style characteristics lacks dynamic updating, and there are still some defects. In the future, we can start from this aspect and choose a more suitable method to predict learners' style characteristics better.

References

1. Chen, Z.: Using big data fuzzy *K*-means clustering and information fusion algorithm in english teaching ability assessment. Complexity **5**, 1–9 (2021)
2. Matsushita, M., Obara, A., Kihara, M.: Study on teaching of university regular PE class (I): with reference to the present status of the consciousness toward the objects of PE in the students. J. Med. Chem. **33**(12), 3127–3130 (2015)
3. Hsu, W.T., Pan, Y.H., Chou, H.S.: Measuring students' responsibility in PE instrument development and validation. Int. J. Sport Psychol. **45**(5), 487–503 (2014)
4. Youn, H.S., Yoo, J.I.: Effect of instructional content reconstitution program in PE class for middle school students upon change in their fair-play behaviors. Asia Life Sci. **2**, 925–934 (2018)
5. Li, X.: A new assessment method for English MOOC teaching quality based on AHP. Int. J. Contin. Eng. Educ. Life Long Learn. **2**, 32 (2022)
6. Exum, W.: The contributions of PE activities for the growth of democratic leadership abilities: an assessment of selected PE activities in secondary schools to determine their potential for developing democratic leadership ability. Synth. Met. **7**(1–2), 133–140 (2015)
7. Wang, J.: Analysis of the role of big data technology in the construction of college sports information. J. Jilin Radio Telev. Univ. **03**, 8–9 (2019)
8. Feng, H.: Research on the operation and management mode of university sports venues. Contemp. Sports Technol. **8**(26), 162–163 (2018)
9. Zhang, J.: Innovative thinking on college sports management under the background of big data. Motion **13**, 1–3 (2018)
10. Rüttermann, S., Sobotta, A., Hahn, P.: Teaching and assessment of communication skills in undergraduate dental education—a survey in German-speaking countries. Eur. J. Dent. Educ. **21**(3), 151–158 (2017)

11. Yuan, Q.: Network education recommendation and teaching resource sharing based on improved neural network. J. Intell. Fuzzy Syst. **39**(4), 5511–5520 (2020)
12. Tan, C.W.: The value of cooperation: from AIMD to flipped classroom teaching. Perform. Assess. Rev. **4**, 49 (2022)

Chapter 30
Research on the Art Design of Green Clothing Based on Image Restoration Technology

Ruofan Sun

Abstract How to ensure that clothing design has high-quality performance and artistic expression while using more green design concepts is an urgent problem for clothing design industry. In clothing art design, the lack of information collection and people's visual differences easily lead to the loss of clothing image information. Therefore, it is necessary to repair and fill in the clothing image information to improve the image expression ability in clothing art design. For the information lost areas in clothing design, the image can be filled and repaired through the prior knowledge of the information intact areas to ensure the accuracy of information expression in clothing design. In this paper, the application of green design concept in clothing design is studied, and a clothing image feature restoration algorithm based on improved convolutional neural network (CNN) is proposed to explore the artistic expression of green design concept in clothing design. The simulation results show that, after many iterations, the recall rate of the improved CNN is obviously better than that of the traditional particle swarm optimization (PSO) algorithm, and the error is significantly reduced. Through the improvement of this paper, the convergence speed of CNN parameters is faster and the final model classification accuracy is higher. Therefore, under the concept of green design, it is feasible to use this algorithm to analyze the artistic expression of green design concept in clothing design.

30.1 Introduction

It is the general trend to apply the green design concept in clothing design, and returning to the natural color has become a new fashion in clothing design. From the exquisite clothing materials and colors to the meticulous modeling design, the whole clothing design industry has been clothed with a green coat [1]. Although the application of green design concept in clothing design is not mature, it has become

R. Sun (✉)
Beijing Institute of Technology, Beijing, China
e-mail: 2023236027@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_30

the mainstream trend of clothing design in the new period [2]. The application of green environmental protection design concept in clothing design can contribute to the harmonious coexistence between man and nature and meet the requirements of sustainable development put forward by China at this stage [3]. The design of green ecological clothing conforms to the core concept of "sustainable development" strategy, which can not only meet people's demand for clothing, but also provide protection for the green environment [4]. Using image graphic processing technology to design clothing art can improve the intelligence and real-time performance of clothing art design [5]. Flexible, efficient, and accurate image processing technology is one of the key technologies to solve these problems. It is of great significance to study the method of clothing art design based on image restoration in realizing intelligent design of clothing art [6].

Clothing is an important carrier for people to express their thoughts, feelings and personal tastes. With the promotion of the concept of green development, people pay more and more attention to the environmental protection of clothing [7]. Design should be based on nature, prevent environmental pollution, maintain ecological balance, and realize the coordinated and sustainable development between man and nature. From the designer's point of view, to carry out the green design concept in the process of clothing design, we should not only consider the environmental protection of raw materials selected in the design process, the pollution, and waste of resources generated in the clothing processing process, but also consider the basic function of clothing to keep out the cold and the aesthetic needs of the image [8]. In clothing art design, the lack of information collection and people's visual differences easily lead to the loss of clothing image information. Therefore, it is necessary to repair and fill in the clothing image information to improve the image expression ability in clothing art design [9]. For the information lost area in clothing design, the image can be filled and repaired through the prior knowledge of the information intact area to ensure the accuracy of information expression in clothing design. With the rise of artificial intelligence, computer vision has attracted the attention of industry and academia, and has become an important technology to promote economic and social development, liberate the labor force, and improve labor efficiency [10]. In the restoration of green clothing image, firstly, the image is segmented, the clothing pattern area is extracted from the whole clothing image, and then the artistic features of clothing are extracted. This paper analyzes the connotation of green clothing design, studies the application of green design concept in clothing design, puts forward a clothing image feature restoration algorithm based on improved CNN, and discusses the artistic expression of green design concept in clothing design.

30.2 Methodology

30.2.1 Application of Green Design Concept in Clothing Design

The concept of green design in clothing design mainly means that clothing designers should first consider the environmental impact brought by clothing production, and the whole lifecycle of clothing should meet the requirements of green environmental protection, minimize the harm to the natural environment, and improve the utilization rate of clothing materials and materials in cutting methods and sewing techniques. Based on the ecological concept, green clothing design can reflect the natural beauty. In the design, a more natural and simple design method is adopted, the clothing state is re-plastic, and the clothing is redesigned based on deconstruction, reflecting the harmonious coexistence between man and nature [11]. The most basic design point of innovative clothing design with green design concept lies in the use of clothing materials, so it is particularly important to choose appropriate clothing materials for design and creation. The art design of green clothing under the ecological concept mainly embodies the beauty of simplicity, but simplicity does not mean simplicity. Compared with other clothing designs, simplicity design is more difficult, so national style should be integrated into green clothing design to show national characteristics. In a sense, national costume design can show the beauty of costume culture, so it can better reflect the beauty of costume culture in green costume art design.

Design behavior itself is an orderly and targeted scientific and rational activity. As the product of the development of the times, green design concept has its unique features and attributes [12]. The strengthening of social and public awareness of environmental protection and the clear direction of future survival of enterprises are the source of green design development. Clothing design under the concept of green design should follow certain principles, whether starting from the early stage, the middle stage, or the postprocessing of clothing, in order to design products that are environmentally friendly, energy-saving and sustainable, and meet the needs of the development of the times. It needs to clarify the different characteristics in each development link, which is closely related to people, nature, and society. The main design style of green clothing art design is simplicity, naturalness, and environmental protection. When designing clothing, the design style should be carried out in every design content, and the style should be continued in the selection of materials and colors.

Green design has become the mainstream trend of clothing design in the new era. It is committed to balance and optimize various factors that need to be considered in green clothing design through technology, so that clothing design with ecological beauty can lead the fashion trend. With the promotion of environmental protection concept, designers pay more attention to the embodiment of natural comfort and green environmental protection concepts while innovating clothing styles. Green fiber materials are widely used in different styles of clothing design, such as leisure clothing design, professional clothing design, sports clothing design, home clothing

design, and so on. Green clothing design should not only stay in the design concept of clothing, but also apply the connotation embodied in green design to the actual design, so as to obtain the achievements of innovative design.

30.2.2 Restoration Algorithm of Green Clothing Image Information

Human beings and the environment interact with each other. While human beings are changing the environment, the environment is also affecting human beings. At present, the unplanned and uncontrolled production of clothing enterprises leads to the supply of products exceeding the demand, which leads to a large inventory backlog of enterprises and seriously affects the development of social environment. Designers should not only consider the environmental protection of raw materials in the design process, but also consider the basic function of clothing covering to keep out the cold and the aesthetic needs of the image [13]. In clothing art design, the lack of information collection and people's visual differences easily lead to the loss of clothing image information. Therefore, it is necessary to repair and fill in the clothing image information to improve the image expression ability in clothing art design. The image information restoration system in this paper works under the general framework of texture-based image information restoration system. The basic method of image information restoration based on texture is to preprocess a given example image and extract its feature vector. By matching the texture feature vector of the material with the feature vector of each image in the feature library and returning the recognition results to the user according to the similarity from large to small, the recognition process is completed. The process of clothing image segmentation is shown in Fig. 30.1.

Input all the collected data information into the computer system, and carry out stitching operation, synthesis operation, deformation operation and color correction operation, and deal with the background color part and grain part. After the data information is entered, it enters the image processing stage, and the color tracking processing is adopted during the output printing, so that the color of the restored image information can be consistent with the original color. In this paper, a method combining improved moment invariants with boundary direction features is proposed, and CNN algorithm is applied to clothing image information restoration, and background interference information is removed by image positioning and cutting methods, so that clothing image information restoration can achieve better results. The dynamic fusion method of clothing texture features based on CNN is shown in Fig. 30.2.

CNN is mainly built according to the number of layers of neural network, which is a nonlinear mapping, and the transmission path is also the process from low transmission to high transmission [14]. For the information restoration technology of clothing image, it is to map the clothing texture picture in the plane state to the corresponding

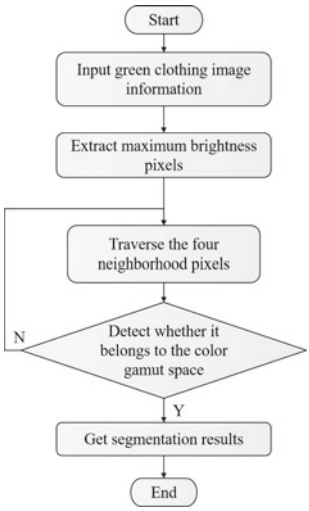


Fig. 30.1 Clothing image segmentation process

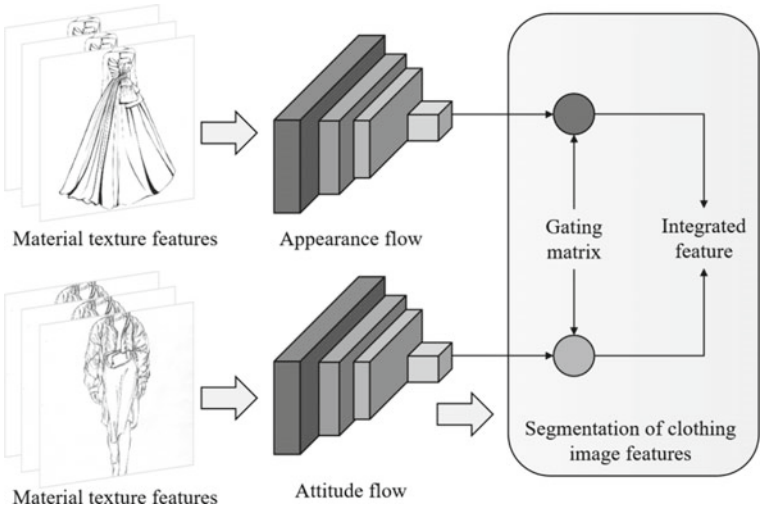


Fig. 30.2 Dynamic fusion of clothing texture features

three-dimensional model. First, the digital camera is used to shoot the material of the mapped picture, and then the digital tool of two-dimensional static processing is used to reprocess it to complete the stitching operation and restoration operation of the picture. Combining local binary pattern and single-shot signal filtering, binary pattern features are extracted on the basis of single-shot amplitude information. The convolution neural network function is defined as:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right) \tag{30.1}$$

where x_i represents the input characteristic map, k represents the convolution kernel, b represents the deviation term, and the convolution output is the characteristic map x_j. Assume that the convolution layer convolves the input clothing material texture image with k filters, and generates k new feature maps for subsequent processing. If the output feature map is represented in a layer, then:

$$F_j^{(n)} = \sum_i w_{ij}^{(n)} * F_i^{(n-1)} + b_j^{(n)} \tag{30.2}$$

where $*$ is a two-dimensional convolution, $w_{ij}^{(n)}$ and $b_j^{(n)}$ are convolution filters and deviations, respectively, $F_j^{(n)}$ is the j output characteristic map at the n layer. The formula of active layer after convolution is as follows:

$$F_j^{(n+1)} = f\left(F_j^n\right) \tag{30.3}$$

where f is a point-by-point activation function. Convert each data item x_i in the small batch $B = \{x_1, x_2, x_3, \ldots, x_m\}$ with size m to y_i:

$$y_i = \gamma \widehat{x}_i + \beta \tag{30.4}$$

$$\widehat{x}_i = \frac{x_i - E_M(x_i)}{\sqrt{\mathrm{Var}_M(x_i) + \varepsilon}} \tag{30.5}$$

where $E_M(x_i)$ and $\mathrm{Var}_M(x_i)$ are the mean and variance, respectively.

The change of network is reflected by the change of network connection right, and the change of numerical value is determined by the learning law of processing unit. The model obtains image features through continuous iteration and then adjusts parameters. After each iteration, the feedback accuracy of the test set is used to judge the iteration, and the iteration result with the highest accuracy is saved. Spatial saliency is used in each sub-region to weight the feature histograms extracted from the sub-regions, and the weighted feature histograms on all sub-regions are connected together. After feature extraction, each image uniquely corresponds to a feature descriptor, and the differences and similarities between different types of images are quantified by their corresponding features.

30.3 Result Analysis and Discussion

The simulation operating system is Windows 11, the processor is Core i7 13700 k, the graphics card is RTX 3060Ti, the memory is 16 GB, and the hard disk capacity is 500G. The simulation uses a clothing image with 600 * 800 pixels randomly selected from Google, which contains 1600 pixels of feature information. When the deep learning algorithm trains the model, it iterates every time it traverses the data. The model obtains image features through continuous iteration and then adjusts parameters. Theoretically, the greater the number of iterations, the better the result of the model. However, in the limited data, the characteristics obtained by the model are also limited, and finally the model training will usher in a convergence. If we continue to iterate later, it will only waste the computing power of the computer. The BN layer has the effect of regularization, which can avoid overtraining the network. However, in the task of image super-resolution reconstruction, overtraining is a rare phenomenon. After each iteration, the feedback accuracy of the test set is used to judge the iteration, and the iteration result with the highest accuracy is saved. Using digital image processing technology, the sensory information and experience are quantified and systematized as accurately as possible to obtain clear features and rules, so as to improve the accuracy and reliability of texture feature extraction of clothing materials. The image inpainting accuracy of different algorithms in clothing image information inpainting is shown in Fig. 30.3.

Compared with high-level statistical features, the classification model of clothing image features based on bottom features can get higher accuracy. The time-consuming of clothing image information restoration with different methods is compared and analyzed, as shown in Fig. 30.4.

As can be seen from Fig. 30.4, the time-consuming of clothing image information restoration processing based on PSO algorithm increases with the increase of the

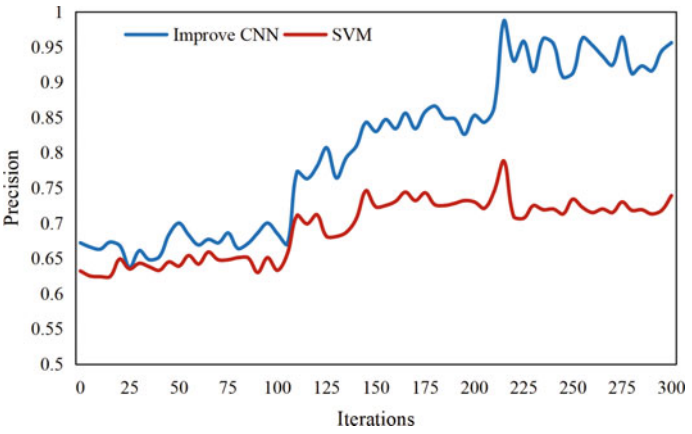


Fig. 30.3 Image inpainting accuracy of different algorithms in image inpainting

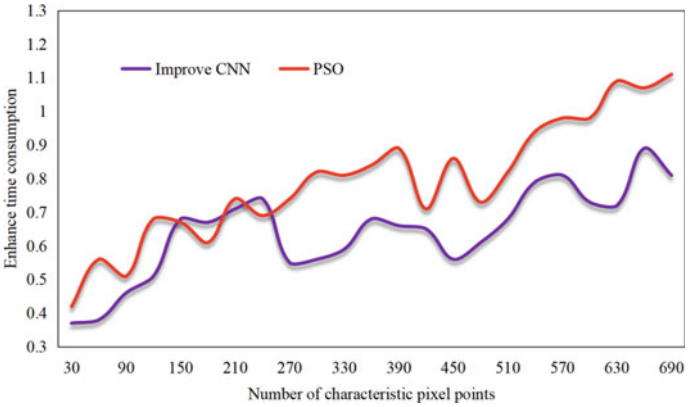


Fig. 30.4 Time-consuming restoration of clothing image information by different methods

number of feature information pixels, which takes a long time. Although the time-consuming of clothing image information restoration based on improved CNN has an upward trend, it has obvious advantages compared with PSO algorithm. In computer operation, the time required for each matching is the same, so reducing the number of matching times can reduce the time required for image recognition and make the recognition results display faster.

Using the improved CNN algorithm to repair green clothing can greatly preserve the details of complex images, save a lot of manpower and material resources, improve the efficiency of image repair, and provide guidance for the performance optimization design of environmental clothing materials. Comparing the recall rate and MAE of the clothing image information restoration model based on improved CNN with the traditional PSO algorithm, the results are shown in Figs. 30.5 and 30.6.

As can be seen from Figs. 30.5 and 30.6, after many iterations, the recall rate of the improved CNN is obviously better than that of the traditional PSO algorithm, and the error is significantly reduced. Through the improvement of this paper, the convergence speed of CNN parameters is faster, and the final model classification accuracy is higher. This method obtains an ideal result of clothing image feature restoration, and the restoration accuracy is higher than other image restoration methods. Clothing image feature restoration based on CNN can accurately and effectively extract clothing image information, which is of great significance to the intelligence of clothing design and the practice of green development concept in clothing design industry.

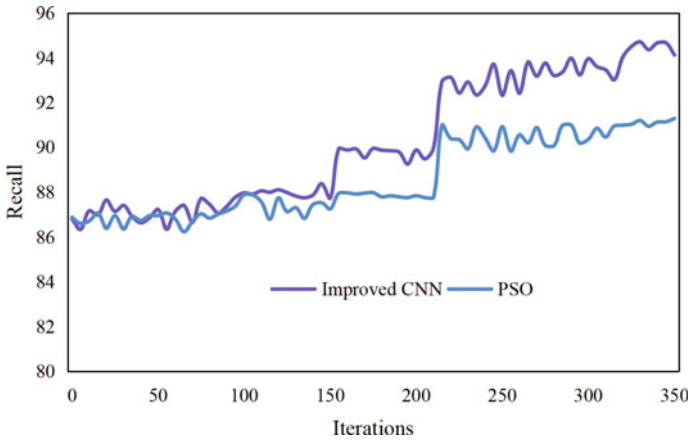


Fig. 30.5 Comparison of recall

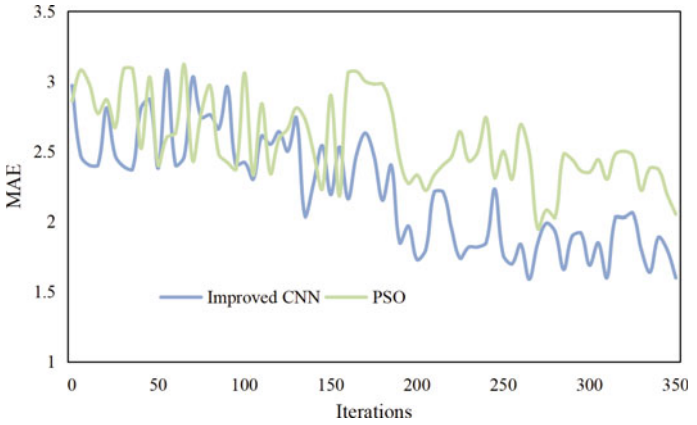


Fig. 30.6 Comparison of MAE

30.4 Conclusions

Clothing is an important carrier for people to express their thoughts, feelings, and personal tastes. With the promotion of the concept of green development, people pay more and more attention to the environmental protection of clothing. In clothing art design, the lack of information collection and people's visual differences easily lead to the loss of clothing image information. Therefore, it is necessary to repair and fill in the clothing image information to improve the image expression ability in clothing art design. This paper studies the application of green design concept in clothing design, puts forward an improved CNN-based clothing image feature restoration algorithm, and discusses the artistic expression of green design concept in clothing

design. The results show that the time-consuming of clothing image information restoration processing based on PSO algorithm increases with the increase of the number of feature information pixels, which takes a long time. Although the time-consuming of clothing image information restoration based on improved CNN has an upward trend, it has obvious advantages compared with PSO algorithm. After many iterations, the recall rate of the improved CNN is obviously better than that of the traditional PSO algorithm, and the error is significantly reduced by 33.66%. With the development of society, science and technology, and people's increasing awareness of environmental protection, green clothing has become an inevitable product that reflects people's comprehensive quality and pursues a higher level of health and beauty. Designers need to constantly learn and master new materials and new technologies, overcome the limitations of traditional thinking, strive to develop and apply new technologies of environmental protection in clothing design, and practice environmental protection concepts with clothing art design works.

References

1. Park, E.B., Heo, J.C., Lee, J.H.: Novel smart clothing with dry electrode biosensor for real-time automatic diagnosis of cardiovascular diseases. Bio-Med. Mater. Eng. **29**(5), 587–599 (2018)
2. Wu, M.J., Zhi, C., Tu, L.: Cotton-containing printing wires based on the two-dimensional braiding method for three-dimensional printing of clothing. Text. Res. J. **92**(9–10), 1384–1393 (2022)
3. Morrison, G.C., Andersen, H.V., Gunnarsen, L.: Partitioning of PCBs from air to clothing materials in a Danish apartment. Indoor Air **28**(1), 188–197 (2018)
4. Xu, S., Chen, S., Feng, Z.: Preparation and controlled coating of hydroxyl-modified silver nanoparticles on silk fibers through intermolecular interaction-induced self-assembly. Mater. Des. **95**(4), 107–118 (2016)
5. Tjnns, M.S., Frevik, H., Sandsund, M.: The dry-heat loss effect of melt-spun phase change material fibres. Ergonomics **58**(3), 535–542 (2015)
6. Puszkarz, A.K., Krucinska, I.: Study of multilayer clothing thermal insulation using thermography and the finite volume method. Fibres Text. East. Eur. **24**(6), 6–129 (2016)
7. Stenton, M., Kapsali, V., Blackburn, R.S.: From clothing rations to fast fashion: utilising regenerated protein fibres to alleviate pressures on mass production. Energies **14**(8), 565 (2021)
8. Pujadas-Hostench, J., Palau-Saumell, R., Forgas-Coll, S.: Integrating theories to predict clothing purchase on SNS. Ind. Manag. Data Syst. **119**(5), 1015–1030 (2019)
9. Chen, J.: Research on evaluating the design effect of clothing and accessories with 2-tuple linguistic information. J. Intell. Fuzzy Syst. **37**(2), 2059–2066 (2019)
10. Yazdi, M.M., Sheikhzadeh, M., Dabirzadeh, A.: Modeling the efficiency and heat gain of a phase change material cooling vest: the effect of ambient temperature and outer isolation. J. Ind. Text. **46**(2), 436–454 (2016)
11. Ortaboy, S., Alper, J.P., Rossi, F.: MnO_x-decorated carbonized porous silicon nanowire electrodes for high performance supercapacitors. Energy Environ. Sci. **10**(6), 1505–1516 (2017)
12. Jin, L., Cao, M.L., Yu, W.: New approaches to evaluate the performance of firefighter protective clothing materials. Fire Technol. **54**(5), 1283–1307 (2018)
13. Ubri, G., Ubri, I.S., Rogale, D.: Mechanical and thermal properties of polyurethane materials and inflated insulation chambers. Materials **14**(6), 1541 (2021)
14. Qin, Z.: Clothing accessory design based on sensory experience. NeuroQuantology **16**(6), 460–465 (2018)

Chapter 31
Research on Teaching Reform of Digital Signal Processing Course Based on Python

Kui Zhang

Abstract Digital signal processing (DSP), as a significant and rapidly developing discipline, has become the backbone course of most electronics, computer, communication, and other related majors. However, the computational complexity is large and cumbersome, making it difficult for students to verify firsthand, resulting in a lack of visualized results and a difficult understanding of theory. Therefore, this article attempts to reform DSP course teaching based on Python. In order to improve teaching quality and stimulate students' interest in learning, we will explain the specific usage of Python and its extension function library in DSP from the perspectives of classroom teaching and experimental aspects. This article has been simulated and tested to obtain the use of Python to calculate and draw graphics, including FFT graphs, Time-domain waveforms and spectra, as well as the design of window functions for low-pass filters.

31.1 Introduction

With the rapid development of computer and information technology, "DSP" technology has become a mature subject, and it has been widely used in various engineering fields, especially in communication and signal processing. The reform of course teaching is an important guarantee to deepen the teaching reform and improve the teaching quality, and it is an inevitable requirement of the development of the times. As a significant and rapidly developing discipline, DSP course has become the main course of most electronics, computers, communications and other related majors. There are many mathematical foundations, formulas, and derivations of DSP, and a large number of algorithms are suitable for computer implementation. However, the amount of calculation is large and complicated, and it is difficult for students to verify it by themselves, so it is difficult to get visual results and understand the theory

K. Zhang (✉)
Shanghai Normal University Tianhua College, Shanghai, China
e-mail: 37331447@qq.com

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_31

thoroughly [1]. DSP is a specialized course. The practical application of DSP theory and method is demonstrated by establishing mathematical model and proper mathematical analysis and processing. The teaching of this course mainly emphasizes applied learning and mainly introduces the use and usage of DSP.

DSP course is a compulsory course for undergraduates majoring in electronic information engineering and communication engineering. Signal and system course is the core professional basic course of electronic information specialty, DSP course is the core professional course of electronic information specialty, and signal and system course is the prerequisite course of DSP course [2]. In the actual teaching of the "DSP" course, we found that using multimedia teaching alone did not achieve ideal results. A teaching method combining classroom blackboard writing, multimedia teaching, and online teaching has been developed to address the issues of difficulty in abstracting the concept of "digital signal processing" and involving multiple formula deductions. Through the study of this course, students can master the basic theory, basic analysis method, and basic realization method of DSP, and lay the necessary theoretical foundation for further study of subsequent courses. DSP is a theoretical course with algorithm as its core. Python was born in the early 1990s. It is an interpretive, object-oriented, and dynamic high-level programming language. It has become one of the most popular programming languages because of its simplicity, readability, and strong expansibility. Many scientific research institutions and well-known universities have begun to adopt Python as the programming language [3]. Because the concept of this course is abstract, theoretical derivation needs a deep mathematical foundation, and it is difficult for students to understand and master this course, which will inevitably affect their learning enthusiasm to some extent [4]. DSP course mainly discusses the basic theory of discrete-time signals and systems, the theory of discrete Fourier transform and its fast algorithm, the design method of infinite unit impulse response filter and finite unit impulse response filter [5].

Due to the fact that a large number of theories and conclusions in this course are obtained through mathematical deduction, students often place too much emphasis on formula derivation or proof, and cannot understand its essence and purpose. The purpose of using DSP to analyze mathematical concepts and knowledge is to use it as a tool to solve practical engineering problems. The theoretical teaching content of digital signal processing course still needs to be enriched and improved, and the teaching methods need to be improved and optimized, and we need to continue to learn and explore. So using DSP in teaching has weakened the skills of deduction and problem-solving, emphasized the physical significance and engineering application of conclusions, and thus strengthened the purposefulness of learning. Through the reform of teaching methods and assessment methods, students' interest in learning can be improved. In order to improve the quality of teaching and stimulate students' interest in learning, we have actively explored and improved methods from classroom teaching and experiments, and gradually achieved some results. Here, we will talk about our experience and discuss with our peers to illustrate the specific use of Python and its extended function library in DSP.

31.2 Teaching Status

Due to the DSP course requiring students to have a high mathematical foundation, such content as the Discrete Fourier transform, the Laplace transform, and the z-Transform, etc. will be learned in the course. In addition, there is also a lot of abstract knowledge. As the foundation of the course, the textbook is the main teaching material. After consulting a large number of domestic and foreign textbooks, the author targeted and selected excellent textbooks and tutoring materials suitable for the current professional students in our school based on different teaching needs and the development trend of DSP discipline. Compared with traditional "blackboard" and "chalk" teaching methods, modern multimedia teaching can play key theoretical knowledge and graphics through multimedia or video, with characteristics such as imagery, diversity, novelty, interest, intuition, and richness [6].

Some basic concepts and methods of DSP are often obscured by numerous tedious mathematical formulas. In teaching, we need a software to provide examples and simulation functions to enhance students' interest in learning and enhance teaching effectiveness, but our focus is certainly on the theoretical knowledge of "DSP", not on improving students' software programming skills [7]. Traditional teaching methods mostly use blackboard writing to analyze and derive mathematical formulas and theorem proofs. The same content in terms of systems includes characteristic analysis of discrete-time systems, frequency response of systems, and solution of system functions. By simulating basic theoretical knowledge and combining various forms such as graphics, images, and animations, interactive teaching of course content can effectively improve teaching effectiveness. More importantly, it can demonstrate the changing process and response results of DSP [8].

However, using courseware demonstration methods for explanation is only a simple transfer of textbook knowledge. The courseware demonstration makes the classroom pace too fast, and some important theoretical knowledge cannot be deeply understood and flexible. It can be seen that the combination of courseware demonstration and blackboard writing teaching mode is difficult to achieve the expected teaching objectives. The abstract and complex mathematical formulas in textbooks can be transformed into very intuitive graphics for students to see, which not only facilitates their understanding but also enhances their interest in learning.

31.3 The Specific Application of Python in Teaching

31.3.1 Application in Classroom Teaching and Experimental Teaching

As the main basic courses for information undergraduates, the two courses "Signal and System" and "DSP" are closely related in teaching content. The teaching content of "Signal and System" and "DSP" can be summarized as two systems, two methods,

and three transformations. The former focuses on the continuous-time systems and the latter focuses on the discrete-time systems.

In theoretical teaching, through a thorough and concrete understanding of theoretical knowledge, we can comprehensively and logically grasp the key points and analyze various knowledge points, use open teaching methods, make full use of the combination of network resources and teachers' teaching, establish a college-led DSP learning website, build a platform for students' autonomous learning, and provide good resources and guidance [9]. In the experimental practice, classroom teaching is optimized through various simulation software to enrich the teaching content, deepen students' understanding of theory, closely combine theory with practice, and exercise students' practical ability, laying a solid theoretical foundation for future signal processing work [10]. Through the simple design of Python processor, we can understand the realization process of DSP algorithm and understand the basic idea of realizing algorithm. By comparing FFT with different points, students can see the frequency domain sampling effect of changing tools for learning mathematics very intuitively. The tectonic sequence is:

$$x(n) = \begin{cases} n+1 & 0 \le n \le 3 \\ 8-n & 4 \le n \le 7 \end{cases} \tag{31.1}$$

Draw its 110-point FFT and 146-point FFT graphics through Python, as shown in Figs. 31.1 and 31.2.

All these make the original boring and abstract theoretical teaching vivid. Under the guidance of basic concepts, basic principles, and basic method principles, the

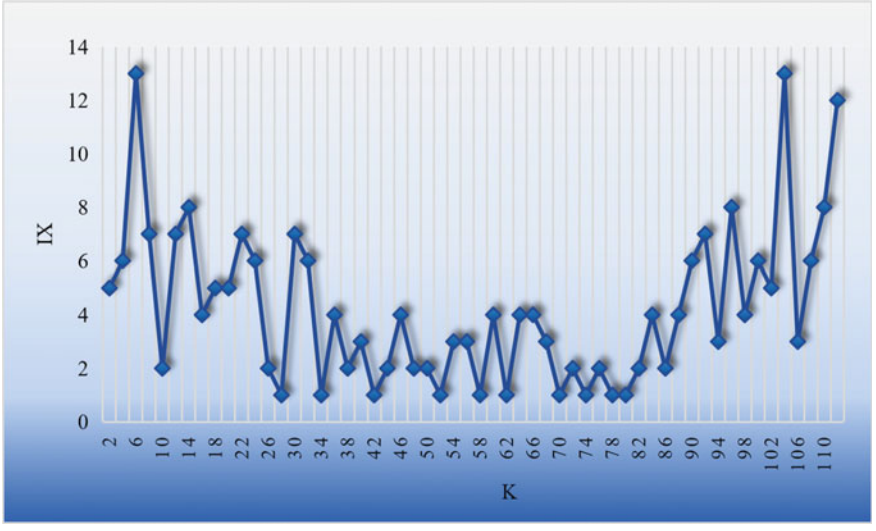


Fig. 31.1 $x(n)$ 110-point FFT graph

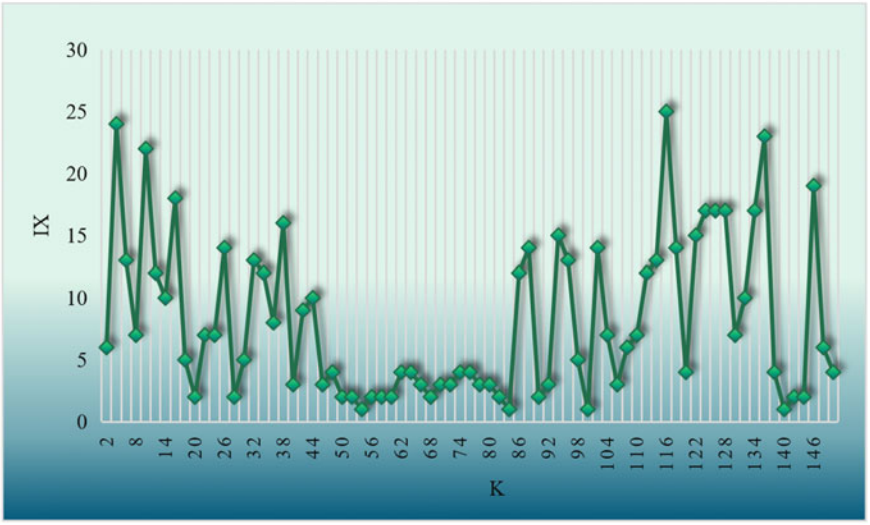


Fig. 31.2 $x(n)$ 146-point FFT graph

knowledge points with universal significance are mainly taught, while similar knowledge is only briefly introduced. For example, when teaching Python digital filter window design method, rectangular window is the focus of teaching, and other windows are only briefly introduced. Because the sampling theorem of signal is an important theory to connect continuous signal with discrete signal, we must master its relationship in learning to better understand the knowledge points [11].

When the course is offered, in order to complete its own system, the problem of content repetition and irrelevance is serious. We arrange for students to spend two weeks after the classroom teaching to test the learning outcomes of the "DSP" semester [12]. Complete the design task through Python independent programming. The outstanding performance is that "Signal and System" focuses on unilateral, and there are five hours of teaching bilateral transformation in "DSP". This part of the content is repeated, but the teaching angle is different, which makes it difficult for students to understand.

31.3.2 The Application of Python in Course Design

Multimedia teaching comprehensively applies materials such as text, images, animations, and videos. It makes teaching students lively, abstract, and difficult to understand knowledge in the classroom intuitive and easy to understand and accept, while also saving teachers' blackboard writing time and improving teaching efficiency. Therefore, it is favored by many teachers [13]. During teaching, observe and collect materials from reality, use these materials to explain abstract theories, return abstract

theories to reality, compensate for students' lack of emotional understanding caused by lack of production and life experience, help students unify theoretical understanding in the course with emotional understanding in real life, and improve learning effectiveness. DSP is a foundational and highly theoretical course that requires knowledge from multiple courses such as signal and systems, complex functions, and engineering mathematics. Therefore, the first thing to emphasize in classroom teaching is to attach importance to theoretical basic knowledge, clarify basic concepts and algorithms, and adopt a strategy of blackboard teaching and multimedia teaching. Still retains classroom blackboard teaching methods to derive and prove some basic principles and methods, which allows students enough time to think.

If teachers only give lectures within a limited time, the most common phenomenon is that teachers spend a lot of time and energy explaining and explaining textbook knowledge, but students still cannot fully understand it, resulting in fear and aversion to learning. For some abstract concepts and principles, multimedia courseware or multimedia simulation demonstrations are used to improve the visualization of teaching, visualize abstract theories, and help students understand and learn. In the two weeks, students are required to complete comprehensive course design tasks independently through Python programming. For example, Spectrum analysis is a method of transforming time-domain signals into frequency-domain and analyzing them, which is used in many scientific and technological fields, such as speech signal processing. The basic process of Python spectrum analysis is to first transform the time signal into the variation patterns of signal amplitude and phase in the frequency domain through Fourier transform, and then proceed to the next step of processing. For discrete time signals, let $x(n)$ be a N point finite length sequence, and its discrete Fourier transform is:

$$X(k) = \mathrm{DFT} = x(n)W_N^{nk}, \quad k = 0, 1, \cdots N-1 \tag{31.2}$$

where $X(k)$ is a complex number, and the modulus and angular angle formed by its real part and imaginary part are the amplitude and phase of signal $x(n)$ in frequency domain, and their variation laws with frequency are respectively called the amplitude-frequency characteristic curve and phase-frequency characteristic curve of $x(n)$.

The signal $x(n)$ in time domain can be obtained by using the inverse discrete Fourier transform IDFT:

$$x(n) = \mathrm{IDFT} = X(k)W^{-nk}N, \quad n = 0, 1, \cdots, N-1 \tag{31.3}$$

It can be seen from Eqs. (31.2) and (31.3) that DFT and IDFT of spectrum analysis involve complex number operation and exponential operation of natural constants.

The practical teaching of DSP focuses on simulation experiments mostly. The simulation program given by the teacher requires students to be able to read every function. Students are required to write their own experimental programs in the course design. Because the teaching content has been redistributed in theoretical teaching, it is necessary to ensure the consistency with theoretical teaching in practical teaching. For the complex signal synthesized by two sinusoidal signals, Python is used to

calculate and draw the graph as shown in Figs. 31.3 and 31.4. Figure 31.3 is a time-domain signal waveform, which is composed of a sinusoidal signal with an amplitude of -1 and a frequency of 157.11 Hz and a sinusoidal signal with an amplitude of 3 and a frequency of 342.12 Hz. Figure 31.4 shows the amplitude-frequency characteristic trend of the frequency domain signal after the signal is transformed with a frequency of 8000 and a length of 523, and the two large peaks in the low-frequency part correspond to 157.11 and 342.12 Hz, respectively.

It can be seen that using Python for signal spectrum analysis can directly generate signal spectrum, graphically display data, and facilitate signal postprocessing. In order to further improve the teaching effectiveness of DSP. The use of open teaching methods can make full use of network resources, combined with the teacher's teaching, the establishment of DSP course learning website, to build a platform for students to learn independently, provide good resources and guidance. The website provides relevant teaching resources, including teaching syllabus, lesson plans, courseware, problem-solving, and course videos.

The DSP course learning website has been established to review after class, learn relevant background knowledge, and expand classroom content. Signal processing is a rapidly developing and widely applied discipline, so in the teaching process, it is necessary to fully integrate the development and application of current signal processing technology, establish students' broad vision of signal processing theory and technology, and stimulate students' motivation to learn signal processing.

31.4 Teaching Reform in Practice

This course is a course that combines theory with practice closely. If we only study theory in class, students' understanding of theoretical knowledge is often superficial. Based on Python window function method and bilinear transformation method are used to design the finite impulse response (FIR) filter, and the frequency response of the filter is drawn. The FIR filter can't be designed by means of analog filter, and the frequency characteristics of approximate ideal filter can be obtained mainly by window function method, frequency sampling method, and equal ripple best approximation method. Then, the collected voice signal is filtered by a self-designed filter, and the time domain waveform and spectrum of the filtered signal are drawn and the signals before and after filtering are compared, the signal changes are analyzed, and the voice signal is played back. The above analysis process is realized by Python language, and the amplitude-frequency characteristic curve of the low-pass filter approximated by rectangular window and Hamming window is shown in Fig. 31.5.

It can be seen that there is still a certain gap between these low-pass filters approximated by window functions and ideal low-pass filters, mainly because there is a transition band and the stop band attenuation is limited. Using Python interface, a simple and easy-to-use graphical user interface is designed, which enables students to visually obtain different low-pass filters with different window functions. The low-pass filter obtained by Hamming window decays faster than that obtained by rectangular

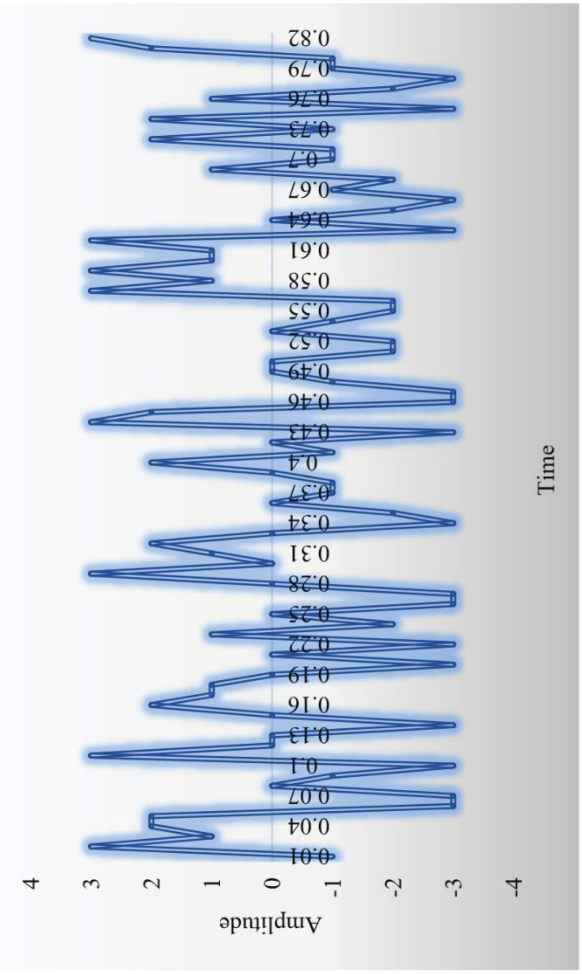


Fig. 31.3 Time-domain waveform

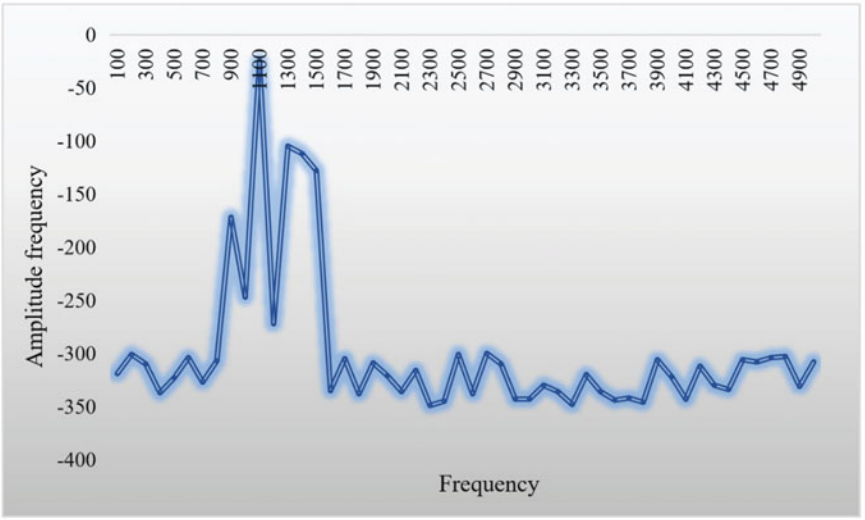


Fig. 31.4 Frequency domain spectrum

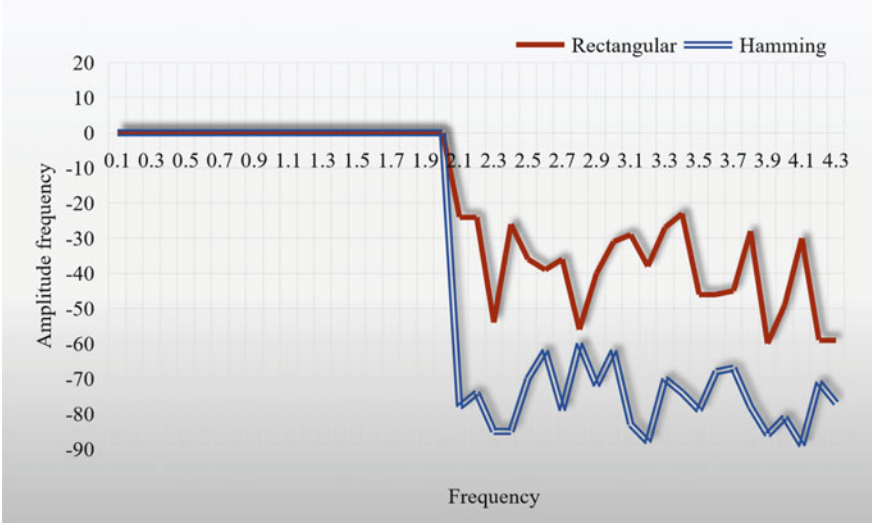


Fig. 31.5 Low-pass filter approximated by rectangular window and Hamming window

window and is closer to the ideal filter. In this way, the enthusiasm of students to participate in the experiment is stimulated, the ability of students to analyze and solve problems independently is cultivated, and the innovation and hands-on ability of students is improved. Deepen the intuitive understanding of knowledge, and embed

indicators such as voice quality and error rate into it to expand students' understanding of this course.

31.5 Conclusions

The Python language and its extension function library constitute a complete scientific computing platform, which is open-source, free, easy to learn and use, and easy to get started. This is significantly better than Matlab, which is widely used in the DSP field, as its core code is not open-source and software licensing fees are high. It can be seen that Python language and its extension function library are very suitable for pre-project design and practical applications in the DSP field. Utilizing the flexibility, openness, and use of software to replace instrument functions of Python language has become a favorable bridge for learning and understanding DSP-related knowledge and experiments. This paper reforms the teaching of DSP course based on Python. After simulation testing, this article obtains the use of Python to calculate and draw graphics. The time-domain signal waveform is composed of a sine signal with an amplitude of -1 and a frequency of 157.11 Hz, and a sine signal with an amplitude of 3 and a frequency of 342.12 Hz. The amplitude-frequency characteristic trend of the frequency domain signal after frequency 8000 and length 523 transformation is shown, and the two large peaks in the low-frequency part correspond to 157.11 and 342.12 Hz, respectively. In order to further improve the teaching effectiveness of DSP, a DSP course teaching website has been established for post-class review, relevant background knowledge learning, and the expansion of classroom content. It provides relevant teaching resources, including teaching syllabus, lesson plans, courseware, exercise answers, course videos, etc. The "test question library" column provides multiple sets of test questions, and also realizes online Q&A, online communication, etc.

References

1. Zuniga-Lopez, A., Aviles-Cruz, C.: Digital signal processing course on Jupyter-Python Notebook for electronics undergraduates. Comput. Appl. Eng. Educ. **60**(5), 28–38 (2020)
2. Weiss, C.J.: A creative commons textbook for teaching scientific computing to chemistry students with Python and Jupyter notebooks. J. Chem. Educ. **98**(2), 9–17 (2020)
3. Duda, M., Sovacool, K.L., Farzaneh, N., et al.: Teaching Python for data science: collaborative development of a modular interactive curriculum. Cold Spring Harbor Lab. **47**(18), 69–72 (2021)
4. Gomez, J.T., Rodriguez-Hidalgo, A., Naranjo, Y., et al.: Teaching differently: the digital signal processing of multimedia content through the use of liberal arts. IEEE Signal Process. Magaz. **38**(3), 94–104 (2021)
5. Qiu, D., Quan, X.: Curriculum and teaching reform from the perspective of media history. Philos. Res. **10**(10), 9–14 (2020)

6. Li, D.: The role of Python in the teaching reform of big data course. Educ. Mod. **37**(12), 46–52 (2019)
7. Huang, X.H., Tang, H.F., Wen-Tao, Y.U., et al.: Teaching reform of digital signal processing course based on Python. J. Electr. Electr. Educ. **55**(20), 66–78 (2021)
8. Liu, Q.: Research on the training of students' computational thinking in Python course under the new curriculum standard of senior high school. Theory Pract. Innov. Entrep. **67**(20), 48–57 (2022)
9. Liu, Q.: Application of blending teaching in digital signal processing. Comput. Knowl. Technol. **48**(18), 51–64 (2021)
10. Ying, X.U., Ya-Kun, G.E., Li-Hua, L.I.: Discussion on the "layered informatization" teaching mode of digital signal processing in postgraduate teaching. Educ. Teach. Forum **66**(11), 31–39 (2022)
11. Zhou, K., Wang, X., Zhi-Qin, H.E., et al.: Teaching reform of digital signal processing. Educ. Teach. Forum **45**(10), 623–654 (2022)
12. Gao, Z.B., Lin, H.Z., Feng, C.: Improving signal processing experiment teaching with Python. Mod. Comput. **25**(5), 61–77 (2019)
13. Wei, L.I., Deyuan, L.I.: Research on curriculum teaching reform of experimental design and data processing. Farm Products Process. **11**(7), 16–25 (2022)

Chapter 32
Visual Analysis of Multi-source Temporal Data for Online Learning

Xian Li, Cen Gao, Meng Li, and Xu Li

Abstract Nowadays with the rapid development of education information, distance education is gradually popularized in People's Daily life and has attracted more and more attention. As the online learning platform using scale expands increasingly, learners and online learning platform of interaction in the process of multi-source temporal data, through analyzing the multi-source temporal data mining, can know more about students' online learning situation, found that students learning rule, etc., this approach will assist the teacher to students in the learning process timely intervention, targeted guidance, and thus, the goal of personalized training can be achieved.

32.1 Introduction

With the rapid development of information technology and the Internet, people have undergone great changes in their life and study. At the same time, online learning education is gradually popularized into People's Daily life and has attracted more and more attention. Online learning than traditional education mode cannot be restricted

X. Li · C. Gao
Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

University of Chinese Academy of Sciences, Beijing, China

X. Li
e-mail: lixian20@mails.ucas.ac.cn

C. Gao
e-mail: gaocen@sict.ac.cn

M. Li (✉)
Shandong Artificial Intelligence Association, Jinan, China
e-mail: 946007942@qq.com

X. Li
Beijing Zhongke Zhihe Digital Technology Co., Ltd, Beijing, China
e-mail: lixu181@mails.ucas.ac.cn

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_32

by geographical, learners can active learning anytime and anywhere, in addition, the charge standard of online learning platform is lower than that of offline training institutions, coupled with the high penetration rate of the Internet, online education in China has developed rapidly, and more and more students choose to learn on the Internet [1, 2]. Nowadays, data analysis technology is rapidly improving, all walks of life are inseparable from data analysis, and the same is true in the field of education. In the process of online learning education, the online learning platform may generate user's basic information, teaching resources, learner's behavior data information during operation, online examination information, and so on at any time. That is, the online learning platform records all the behavior data information of the learner from entering the platform to exiting the platform. In the face of so a lot of data, how to analyze and extract useful information and transform the information into a more understandable form is a key problem that urgently needs to be solved to discover and utilize the value of educational data for enhancing the quality of education [3]. In recent years, researchers have studied more and more online learning behavior data analysis, and the analysis methods have become more and more diversified. In addition, with the rapid development of data visualization technology, students' behavior data can be conveyed intuitively and diversified in a graphical way.

32.2 Related Work

32.2.1 Time Series Data Study

Time series data usually refers to the data containing the characteristics of time series. Time series data is the same column of data recorded in chronological order following a unified indicator. Each data in the same data column must be of the same caliber, requiring a comparability on the time node. Time series data can be either time points or time periods. Time series data can be the ner, Time Searcher, VizTree, Time Series Bitmaps et al. Online learning will produce a large number of course watumber of periods, can also be the number of hours [4]. The target of time series analysis is to construct time series models and make out-of-sample prediction by finding out the statistical characteristics and development regularity of time series within samples. Time series visualization is a promising research direction. Time series visualization is a research direction with broad application prospects. At present, there are many researches at home and abroad, and corresponding visualization tools have been developed, such as Time Series, Theme River, Time Searcher, VizTree, and Time Series Bitmaps. In the process of online learning, a large amount of course viewing data, user behavior data and other time series data will be generated. In order to solve the problem of visualization of time series data, existing visualization methods will be applied in practical applications, improvement also requires new visualization methods to visualize the data, especially the introduction of new visualization methods and visual analysis methods to improve the presentation and analysis of data [5].

32.2.2 Visualization Methods Study

With the accelerated development of the information society, data sources have the characteristics of multiple channels and multiple structures. The processing and visualization analysis of such multi-source heterogeneous data have become the focus of researchers. Common visualization methods for multi-dimensional heterogeneous data include Sankey diagram, word cloud visualization, tree graph, radar graph, scatter graph, force guide graph, etc. Among them, the visualization methods of time series data include single bar graph, parallel bar graph, stacked bar graph, scatter graph, line graph, Sankey diagram, ladder graph, etc. [6–8]. Through the study of temporal data mining, Yang et al. [9] proposed a learning path planning algorithm for multi-dimensional temporal data analysis. By analyzing logs of different systems, Tao et al. [10] also proposed an ERD enhanced visualization method for cross-domain data fusion. Ding et al. [11] presented a hybrid visualization method of VPM based on circular nested graph and parallel coordinates in view of the multi-dimensional characteristics of college teachers' performance data. Such time series data also exists in online learning platforms. Visual analysis of this data can help teachers better understand students' learning habits, so as to reasonably optimize learning resources, further improve the learning content of course chapters, and reasonably adjust the course structure according to students' learning conditions.

32.3 Experiment Method

32.3.1 Multi-source Time Series Data Fusion

The characteristics of the multi-source temporal data of the online learning platform are mainly reflected in the data of students' personal information, online course viewing data, students' question bank exercises, students' examination data, system log information, and the data collected from the buried points of the system. Multi-source data fusion main goal is to multiple systems, different types of data formats produced by multiple sources of structured data and unstructured data fusion processing, the data unified platform for specification of the data format, making it easy for users to access to provide users with the required data of decision-making and assessment tasks. This platform provides data support for user visual analysis through data fusion of student learning behavior data processing, data relationship analysis, user data processing, and platform log data analysis.

This platform adopts extract, transform, load (ETL) to fuse multi-source data. The key steps of ETL method are data extraction, data conversion, and data loading [12]. The fusion steps are shown as follows (Fig. 32.1).

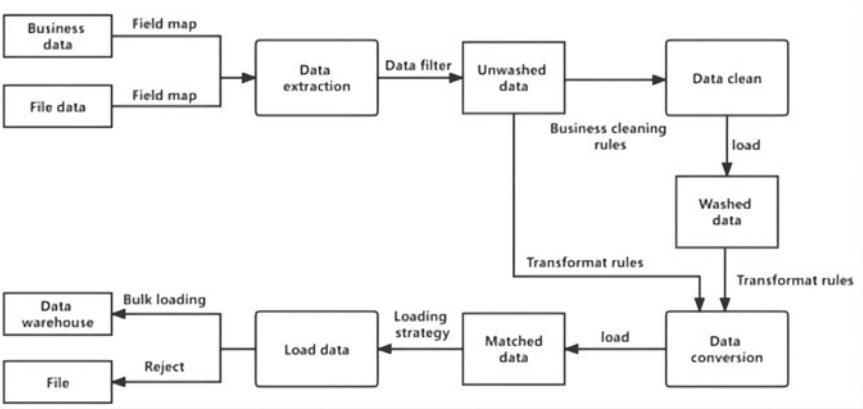


Fig. 32.1 Data fusion process diagram

Data Extraction

There are two ways of data extraction, namely incremental extraction and full extraction. The simplest way to extract data is to use full extraction for structured data, migrate, and copy the data into a standardized data format. Incremental extraction of data is monitored by adding, deleting, modifying, and checking data in the data table. Incremental extraction is performed only when the monitored data changes. The main implementation methods are trigger, log comparison, time stamp, and table comparison. Parameters and functions in the ETL data extraction process include: data source type, extraction conditions, extraction method, extraction speed, data transformation, target data type, target table structure, data loading method, loading speed, and data quality check. By configuring the above functions and parameters, ETL can effectively transform source data into target data while ensuring data integrity, correctness, and consistency (Fig. 32.2).

Data Conversion

Data transformation is divided into cleaning and transformation of two steps, based on the data of cleaning out do not meet the desired data, for example, have lost part of the information in the process of extracting data, incomplete data affect the late model of expected results, need to empty data, missing data, the supplementary work and tag is unable to process the data. In addition, some errors and duplicate data need to be corrected, so that the robust data set generated after cleaning can provide data support for later visual analysis. Data conversion mainly focuses on data type, data format, and the normalization of data required by the business. In the operation of moving data, verifying data, modifying data, integrating data, and so on, the operation of data conversion is involved, mainly through data merging, data splitting, column and row conversion, deduplication, data verification, and other rules. Common parameters and functions in the ETL data transformation process include: data cleansing, data format conversion, data merging, data splitting, data transformation rules, data aggregation,

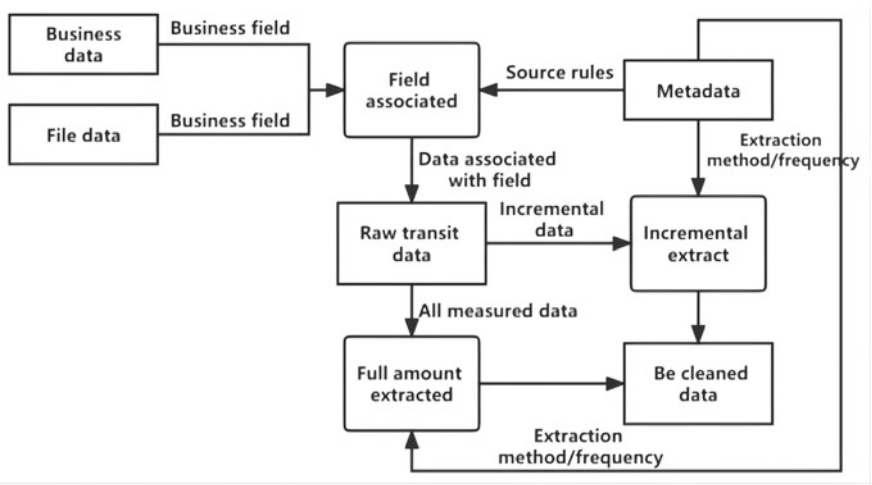


Fig. 32.2 Data extraction process diagram

data sorting, data grouping, data pivot table, data validation, etc. By configuring these functions and parameters, ETL can transform source data into target data and perform operations such as format conversion, cleansing, merging, splitting, and aggregation (Fig. 32.3).

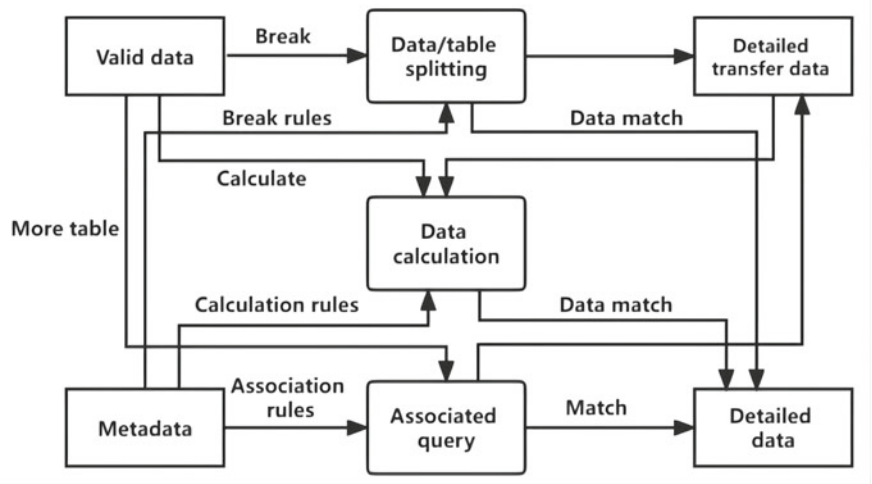


Fig. 32.3 Data conversion diagram

Data Loading

Data loading can be implemented in four ways: timestamp, full table deletion and insertion, full table comparison log, and long table. The timestamp mode is to uniformly add timestamp fields to the data table and update the timestamp when the data is modified. In log table mode, you can add a log table. Update maintenance log table contents when data changes. The whole table comparison method extracts all metadata, and the data comparison is performed when the target table is updated. The whole table deletion insertion mode deletes the target table data and inserts all the source data.

32.3.2 Visualization Method

Sankey Diagram Visualization Method

The time series data of multi-source data was visualized and analyzed by Sankey diagram. Sanki diagram is also known as Sanki energy shunt map or Sanki energy balance map. It is a specific type of flowchart, where the width of the extended branch corresponds to the size of the data flow, and it is usually applied to the visual analysis of energy, material composition, finance, etc. The most obvious feature of the Sanki diagram is that the sum of branch widths at the start and finish is equal; that is, the sum of all main branch widths should be equal to the sum of all divided branch widths, maintaining the energy balance. Iterative relaxation method is an iterative way to solve linear equations, and it is also the principle of Sankey diagram in network. From the perspective of time series, each data unit of Sankey diagram can be represented by a binary set (t, o). Time series R is a set of n bivariate groups ($t1$, $o1$), ($t2$, $o2$), … The finite set of, (tn, on), these binaries groups represent data trends over a period of time, where $ti < ti + 1$ ($i = 1, 2, \ldots, n - 1$), ti represents the i-th number, oi represents the data variable of the i-th data unit. The iterative formula of the Seidel iteration method can be expressed as:

The relaxation algorithm is [13]:

$$x^{(m+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(m+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(m)}\right) = x_i^{(m)} + r_i^{(m+1)}(i = 1, 2, ..., n) \tag{32.1}$$

Among them:

$$r_i^{(m+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(m+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(m)}\right), (i = 1, 2, ..., n) \tag{32.2}$$

When the iteration converges, $r_i^{(m+1)} \to 0$ $(i = 1, 2, \ldots, n)$.

Dynamic Data Histogram Visualization Method

Dynamic bar chart is often seen in daily data reports. In fact, it is still a kind of bar chart. The data range displayed by dynamic sorting chart is much larger than that of ordinary bar chart, and it is better than that of ordinary bar chart in vision. Dynamic bar chart is to take time as a variable, in different time data size and data items are changed, so that not only the intuitiveness of the bar chart, but also the trend expressed by the line chart. In addition, it can be combined with other visualization methods such as line graphs to form multi-dimensional graphs of dynamic data, which can show more data relationships of features over time.

32.4 Experiment and Analysis

32.4.1 Experiment Preparing

The data source of this experiment is a training institution. Using Windows10 system, the processor is Core (TM) i5-10200HCPU @ 2.40 GHz, and the development IDE is WebStorm.

The visualization tasks of this experiment are as follows:

- Student test data visualization
- Student question bank practice data visualization.

32.4.2 Experiment Plan

Through the learning behavior data and student personal information generated by the learning process of students in a learning institution. The various data fusion of heterogeneous data sources, including data cleaning and analysis of the different sources, different structure standardization of data warehousing, set up online learning platform, through the visualization methods of multi-source temporal data display, visual task, through unified visual interface to student's learning behavior data visual analysis. To provide students with comprehensive analysis and guidance of the learning process, clear positioning of learning habits and adjustment of learning behaviors.

32.4.3 Experiment Visualization Results

Visual Analysis of Sankey Diagram

The improved visualization method of time series data based on Sankey diagram is realized by Echarts open source tool. By improving the layout algorithm of the original Sankey diagram, the method of data time connection is implemented on the basis of the source code, and the key code is shown in Fig. 32.4. First, create an HTML container to hold the Echarts chart. Second, the diagram is initialized by calling the echarts.init() method. After that, the configuration items for the diagram are set, the data is passed in, and the diagram is drawn using the setOption() method. Diagram example is shown in Fig. 32.5.

By comparing the performance of Sankey diagram before and after the improvement, it can be shown in Fig. 32.6 that the improved diagram has faster first contentful

```
function process(status, idx, oldIdx) {
    var oldLayersGroups = this._layers;
    if (status === 'remove') {
        group.remove(oldLayersGroups[idx]);
        return;
    }
    var points0 = [];
    var points1 = [];
    var color;
    var indices = layerSeries[idx].indices;
    for (var j = 0; j < indices.length; j++) {
        var layout = data.getItemLayout(indices[j]);
        var x = layout.x;
        var y0 = layout.y0;
        var y = layout.y;

        points0.push([x, y0]);
        points1.push([x, y0 + y]);

        color = data.getItemVisual(indices[j], 'color');
    }
```

Fig. 32.4 Key code

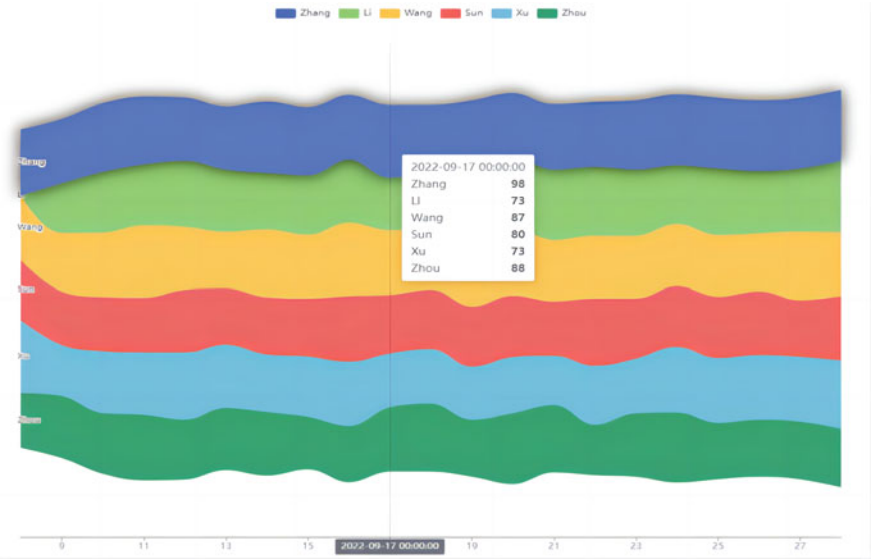


Fig. 32.5 Visualization of Sankey diagram

paint, shorter speed index, shorter time to interactive, and shorter total blocking time. Most performance index has been greatly improved, indicating that the improved method is effective for the improvement of performance. Compared with the layout of bar chart and line chart, the improved Sankey diagram can better display the scene with long time span, continuous time, and dense time points, and the performance indicators such as loading speed are better than those before optimization. The improved time series visualization method based on Sankey diagram can provide better performance in the tedious data. Teachers and students can clearly understand the heat of the data in a certain period of time and directly analyze the fluctuation of students' grades by observing the width of the color block. Through the Sankey diagram visual analysis, can be analyzed that a class of students Zhang's multiple examination scores are much higher than other students, the results do not fluctuate, on behalf of its learning method is more effective. Xu's scores in multiple exams fluctuated significantly and fluctuated greatly. There may be problems in his learning method, which needs to be adjusted in time.

Dynamic Data Histogram Visualization Analysis

Through the visual analysis of the dynamic data bar chart, we can analyze the practice situation of the class students in the question bank-python question bank over time. The dynamic bar chart can visually show the temporal changes of data and visually observe the changes of data (Fig. 32.7).

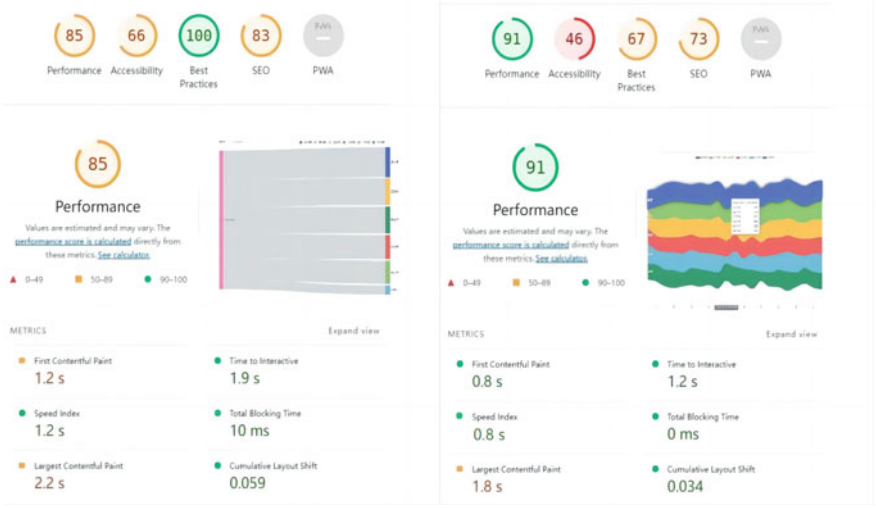


Fig. 32.6 Before and after the improvement of Sankey diagram performance

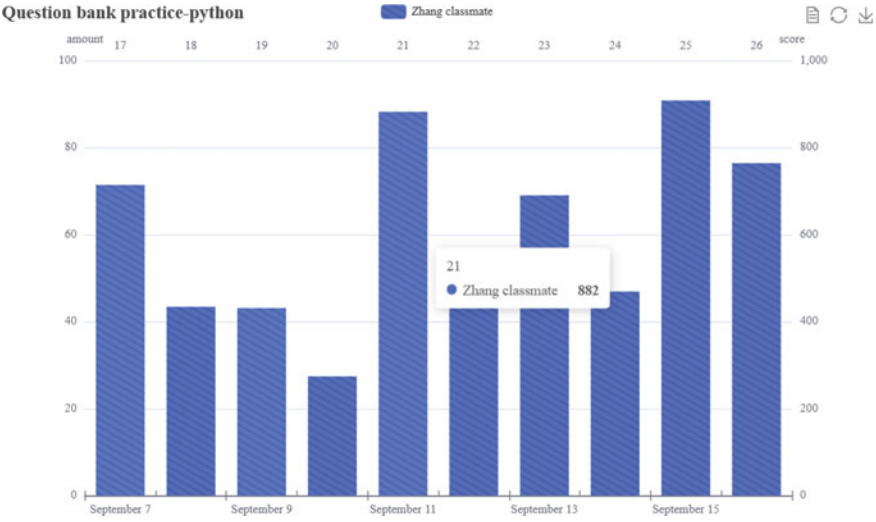


Fig. 32.7 Dynamic data histogram visualization

32.5 Conclusion

The fusion of multi-source data can improve the value of data and make the data have higher accuracy and effectiveness. High quality data and access efficiency can also support the visual analysis of data and visual display. The visual analysis method for temporal data presented in this paper can be applied to most systems with temporal

data characteristics. Compared with the traditional data visualization methods, the visual analysis method is more widely used in the system, and it solves the problem that the temporal data is difficult to process and forms an effective visual analysis. It proves that the visual analysis method in this paper has a certain practical value and can assist students in online learning and meet the needs of online learning system.

References

1. Shi, J., Song, S.: Investigation and counter measures of online learning resource application. Educ. Modern. **7**(43), 159–164 (2020)
2. Qin, Z.D.: Online education data report released in the first half of the year is expected to reach 523 billion yuan. Comput. Netw. **47**(16), 7 (2019)
3. Wang, Q.H., Zhang, L.: Data visualization analysis of network information security. Electr. Techn. **51**(08), 76–78 (2022)
4. Zheng, Y.F., Zhao, Y.N., Bai, X., Fu, Q.: Research review on visualization of education big data. J. Front. Comput. Sci. Technol. **15**(03), 403–422 (2021)
5. Zhang, Y.W.: Based on student performance behavior change visualization analysis and research. In: University of Electronic Science and Technology (2020)
6. Liu, B., Liu, Z.J., Liu, Y.: Overview of data visualization research. J. Hebei Univ. Sci. Technol. **42**(06), 643–654 (2019)
7. Liu, T.T., Wang, Y.H., Tu, C.H., Jiang, P.: Tree map layout based on Monte Carlo tree search. J. Comput. Aided Des. Comput. Graph. **33**(09), 1367–1376 (2021)
8. Tu, L., Chen, B.J., Zhou, Z.G.: Overview of OD data visual analysis. J. Comput. Aided Des. Comput. Graph. **33**(08), 1160–1171 (2018)
9. Yang, Y.L.: Temporal data analysis in the context of big data. In: University of Science and Technology of China (2018)
10. Tao, Y., Guo, S.T., Wang, X.D., Hou, R.C., Chu, D.H.: Log driven cross-domain data fusion and visualization method. Manufact. Autom. **42**(09), 142–147 (2020)
11. Ding, W.L., Xue, L.L., Chen, W.J., Wu, F.L.: A method of visualization of university teachers' performance data based on hybrid layout strategy. Comput. Sci. **46**(02), 24–29 (2019)
12. Huang, T.T., Feng, J., Feng, F.: Research on security monitoring model based on multi-source data fusion. Appl. Res. Comput. **37**(S1), 177–179 (2020)
13. Jiang, T.T., Xiao, W.D., Zhang, C., Ge, B.: Text visualization method of time series based on Sankey diagram. Appl. Res. Comput. **33**(09), 2683–2692 (2016)

Chapter 33
Application of Deep Learning Technology in College Teaching

Xiaocong Sui, Xiaohui Sui, and Xiangping Shen

Abstract With the twenty-first century gradually taking core literacy training as the key work content of contemporary college education, deep learning technology has become one of the most direct and effective tools, which will help students develop the ability to learn independently. By cultivating college students' ability of deep learning, deep processing and improving learning efficiency have become the key content of college curriculum reform. College Chinese is an important course integrating tools and humanities, which is an important basic course for students to have a solid foundation. Therefore, colleges and universities must strengthen the use of students' in-depth learning, which will improve the self-restraint of college Chinese teaching, which will strengthen students' aesthetic experience. Through in-depth learning technology, we can help students build an ideal personality, which will strengthen the cultural heritage. Therefore, this paper analyzes the important role of the current depth learning. At the same time, based on the guidance of the goal of deep learning, this paper conducts a survey of the research status of college Chinese teaching. Finally, this paper proposes some measures for the current application of deep learning.

33.1 Introduction

The traditional Chinese education model is based on shallow learning. For shallow learning, deep learning is a learning method that automatically criticizes new knowledge, which can integrate learning feelings, perceptions, and insights into existing cognition [1]. Through teaching reflection, learners can transfer the known knowledge to new learning situations, which will constantly improve the learning level of learners. By strengthening learning ability, learners can constantly explore new

X. Sui (✉) · X. Shen
Weifang Engineering Vocational College, Weifang 262500, China
e-mail: suixiaocong123@163.com

X. Sui
Changle County West Lake Kindergarten, Weifang 262400, China

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7_33

problems, which will generate new comprehensive abilities. However, the current Chinese higher education is facing many difficulties, which are mainly divided into two types. First, the status of college Chinese course is gradually marginalized, which will reduce students' enthusiasm for college Chinese learning [2]. Second, with the lack of interest in learning, colleges and universities gradually neglect the cultivation of Chinese literacy, which will lead to the imbalance of humanistic education. Therefore, how to integrate deep learning technology into college Chinese teaching has become an important research topic, which can constantly improve students' cognitive, emotional and operational fields. Through in-depth study, we can improve the humanistic quality of college students, which can also bridge the boundaries between literature and other disciplines. Therefore, it is of great significance to study the application of deep learning technology in college Chinese [3, 4].

33.2 The Important Role of Deep Learning in College Chinese Teaching

Through in-depth learning, students will continue to enhance their enthusiasm for learning, which will also enhance their high-level thinking ability. Through in-depth learning, students will continue to improve their ability to solve complex problems, which will continue to improve the learning effect of students [5, 6]. Deep learning plays a very important role in college Chinese teaching, as shown in Fig. 33.1.

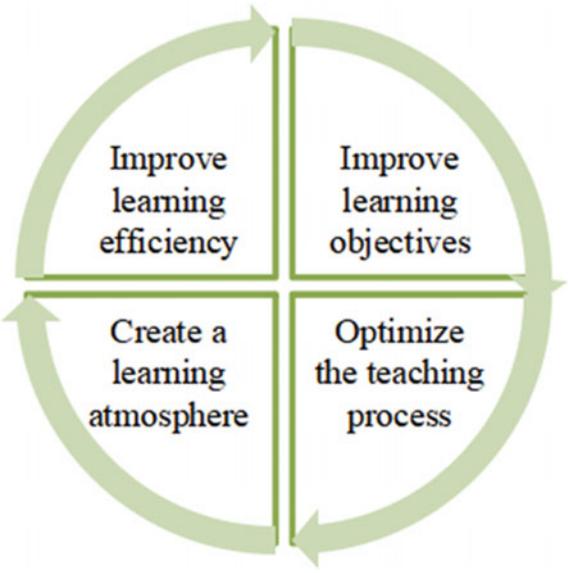


Fig. 33.1 Important role of deep learning in college Chinese teaching

33.2.1 Improve Learning Objectives

Deep learning is a systematic accumulation process, which requires full analysis of classroom knowledge. By distinguishing the key points and difficulties of college Chinese, we can systematically analyze whether the category of knowledge is conceptual or technical, which will timely complete the orderly arrangement of knowledge. Through different expressions of knowledge points, we can improve the basic objectives of learning, which will transform students' thinking in different ways. By stimulating students' endogenous motivation, we can improve students' way of thinking, which will fully stimulate students' understanding ability. Through the improvement of learning objectives, colleges and universities can determine students' thinking ability, which can be scientifically classified and summarized according to different understanding abilities [7]. By comparing the learning styles of college Chinese, teachers can help students improve their learning skills and methods, which will improve their learning. By improving learning objectives, we can narrow the differences among students, which will meet the needs of differentiated students in colleges and universities [8].

33.2.2 Optimize the Teaching Process

The innovative design method of deep learning is mainly based on the traditional teaching concept, which will be reflected in a variety of teaching disciplines, especially in the college Chinese classroom. Through multimedia teaching, college teachers can constantly innovate teaching methods, which can constantly optimize the teaching process [9]. Based on computer assistance, we can constantly increase the teaching knowledge system, which can classify knowledge in a short time. Therefore, college teachers should learn computer teaching ability, which will make teaching process design faster. According to the syllabus, teachers can classify teaching objectives. Through knowledge segmentation, teachers can express different knowledge points and contents through video or animation. Through the answers to relevant stories, we can constantly mobilize the classroom atmosphere, which will make the whole teaching process go smoothly. Through in-depth learning methods, colleges and universities can help students develop cooperation and communication skills, which will develop students' ability to deal with complex problems [10]. Therefore, colleges and universities need to constantly optimize the teaching process, which can continuously optimize the teaching design and release the nature of students.

33.2.3 Create a Learning Atmosphere

Based on the concept of deep learning, college Chinese classroom design is becoming more and more scientific, which is not just the change of students' thinking mode. In the face of the dual pressures of universities and society, the most important thing for in-depth learning is to integrate professional knowledge with society, which can continuously improve the knowledge level of students. Through the integration of theoretical knowledge and practice, college Chinese teaching has changed from traditional learning items to creating atmosphere [11]. At the same time, deep learning is a teaching tool, which can independently level and diversify knowledge. Based on computer technology, we can change knowledge in different types, which can enable students to conduct in-depth mining analysis. Through the deep learning mode, students can make the learning process interesting, which will change the traditional boring teaching. Through the deep learning mode, colleges and universities can create a comfortable learning environment, which will provide students with an efficient learning condition [12].

33.2.4 Improve Learning Efficiency

The concept of deep learning can mobilize students' enthusiasm, autonomy, and curiosity, which can solve and guide students' ability to find and analyze problems. By promoting the establishment of learning and thinking ability, students can explore knowledge content independently, which can improve their personal sense of cooperation [13]. Through deep learning, students can fully improve their learning efficiency, which will gradually adapt to the social rules of knowledge learning. By improving their professional quality in all aspects, students can continuously enhance the value created in their work. Deep learning design and practice can help students understand their own shortcomings. Through self-evaluation, students can improve, update, and process their own learning methods, which will promote students' in-depth learning ideas. By changing learning strategies, students can constantly optimize their learning methods, which can make classroom teaching move toward the goal and plan.

33.3 Material and Methods

33.3.1 Integration of Deep Learning and Flipped Classroom

The integration of deep learning and flipped classroom mainly focuses on flipped classroom, which is a popular teaching mode at present, and can effectively promote students' deep learning. With the continuous improvement of integration, deep

learning has also become a standard for testing flipped teaching results. Bergmann clearly pointed out that flipped classroom can make learning more in-depth, which can carry out different levels of learning in the pyramid structure of teaching objectives. Through the pyramid structure, we can learn that deep learning is a high-level thinking ability, which can be more scientifically applied, analyzed, evaluated, and created. Flipped classroom is a way to complete learning tasks in advance through video, which will improve the application, analysis, evaluation, and even creation of classroom time, which is an important way to achieve the high-level goal of teaching. Therefore, the integration of deep learning and flipped classroom has become an important way. McLean believes that flipped classroom has seriously affected students' learning style, which can help students develop their time management ability when completing pre-class tasks. By learning in advance, students can explore learning patterns and monitor learning objectives in advance based on deep learning strategies. Based on the way of deep learning, students can ensure that they understand the content of learning, which can deepen their knowledge learning by taking notes [14].

Danker explored whether flipped classroom can promote deep learning through large class teaching in universities. The research believes that the following learning behaviors can reflect students' deep learning, including seeking the meaning of learning, active interaction, and connecting new and old knowledge. Research shows that more than half of the students will understand the course content by pausing and repeatedly watching the teaching video and taking notes. In terms of understanding of the content, 80% of the students choose to understand very well. In terms of students' participation in the classroom, most students think that they can deepen their understanding of knowledge by participating in classroom activities. When they feel that classroom activities are interesting, they are more willing to participate in the classroom. They can acquire new knowledge through other students and share it with others. At the same time, the research shows that the church actively connects old and new knowledge and actively uses new knowledge to solve problems in the real world [15].

James et al. based on their flipped classroom teaching practice in maritime major, the research results show that flipped classroom can systematically teach students to think critically, and at the same time, flipped classroom provides more opportunities for students to learn actively, which can promote students' learning engagement.

In China, the research on the integration of deep learning and flipped classroom is mainly divided into two categories: First, it takes deep learning as the goal and result of flipped classroom teaching, and flipped classroom as the means and way to achieve deep learning. Professor Huang Guozhen, a Taiwan scholar, believes that flipped classroom is mainly aimed at guiding students to think at a high level. Through the rules of high-level thinking activities, it can effectively cultivate students' 5C abilities, namely communication and coordination ability, cooperative ability, critical thinking ability, complex problem-solving ability, and creativity. The above abilities are just the important contents of the deep learning ability. In the teaching practice based on SPOC, Liu Hongpin and Tan Liang solved the problem of teachers' overload through the construction of student aid groups and promoted students' deep

understanding of concepts and problem-solving ability through targeted problem solving by student aid groups. Zeng Mingxing and others proposed that the organic integration of massive open online course (MOOC) and flipped classroom can build a meaningful deep learning field, which can provide an effective environment for knowledge construction and understanding, knowledge transfer application, problem solving, and innovation. Peng Feixia, from the perspective of learning analysis based on big data, explained how to improve the quality of mixed learning through education big data to support learning analysis, so as to promote in-depth learning [16].

33.3.2 Data Filtering Method

Suppose dataset D contains n data objects, which are divided into k clusters $C_1, \ldots, C_k$. So, we can make for $1 \leq i, j \leq k$, there is a $C_i \subseteq D$ and $C_i \cap C_i = \emptyset$. We can get the distance relationship between the data object $x \in C_i$, and the cluster center C_i is shown in Formula 33.1.

$$\operatorname{dist}(x_i, x_j) = \sqrt{\sum_{m=1}^{m} \left| x_{im} - x_{jm} \right|^2} \tag{33.1}$$

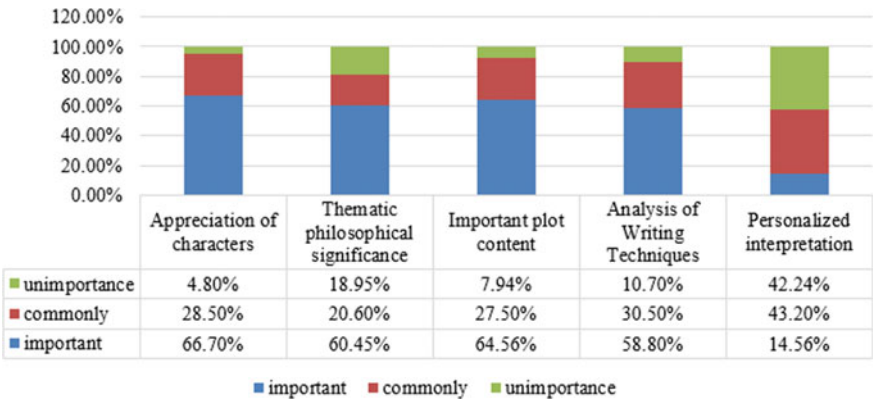
Generally, SSE is the threshold target to measure clustering completion, as shown in Formula 33.2.

$$\mathrm{SSE} = \sum_{i=1}^{k} \sum_{x \in C_i} |x - C_i|^2 = \sum_{i=1}^{k} \sum_{x \in C_i} \operatorname{dist}(x, C_i)^2 \tag{33.2}$$

33.4 Results

33.4.1 A Survey of College Chinese Learning Methods

There are many ways to learn college Chinese, which will lead students to learn according to the way they are best at. Therefore, colleges and universities must correctly guide the learning model, which is a way to improve learning ability. Therefore, teachers must explore multiple channels, which can guide students to think independently. Based on the theory of deep learning, this paper investigates some language majors [17]. Among them, the results of the methods used in college Chinese learning are shown in Fig. 33.2.



	Appreciation of characters	Thematic philosophical significance	Important plot content	Analysis of Writing Techniques	Personalized interpretation
unimportance	4.80%	18.95%	7.94%	10.70%	42.24%
commonly	28.50%	20.60%	27.50%	30.50%	43.20%
important	66.70%	60.45%	64.56%	58.80%	14.56%

Fig. 33.2 Research results of college Chinese learning style

33.4.2 A Survey of the Mode of Listening to College Chinese Classes

The traditional teaching mode of college Chinese mainly depends on the way of teachers' explanation, which can obtain a lot of knowledge. It is difficult for teachers to improve the classroom teaching atmosphere through one-way listening to teachers' interpretation, which will also be detrimental to the full interpretation of the classroom. Only by concentrating on thinking, can students complete their learning tasks more quickly, which will constantly improve the classroom atmosphere. Therefore, based on the theory of deep learning, this paper investigates some language majors. Among them, students are more willing to think for themselves when investigating the mode of listening to college Chinese classes [18]. The results obtained in this paper are shown in Fig. 33.3.

33.5 Discussion

33.5.1 Curriculum Design and Application

This paper designs the deep learning of college Chinese, which is mainly divided into three steps. First, master the teaching content before class. By means of micro classes, MOOC, etc., the school can build excellent college Chinese courses, which will help students learn relevant content in advance [19]. At the same time, teachers should release learning tasks in advance. Through automatic learning, students can initially process new knowledge, which will improve and activate the contrast of old knowledge. Through the establishment of multiple connections between old and new knowledge, students can better build a language knowledge framework. Second,

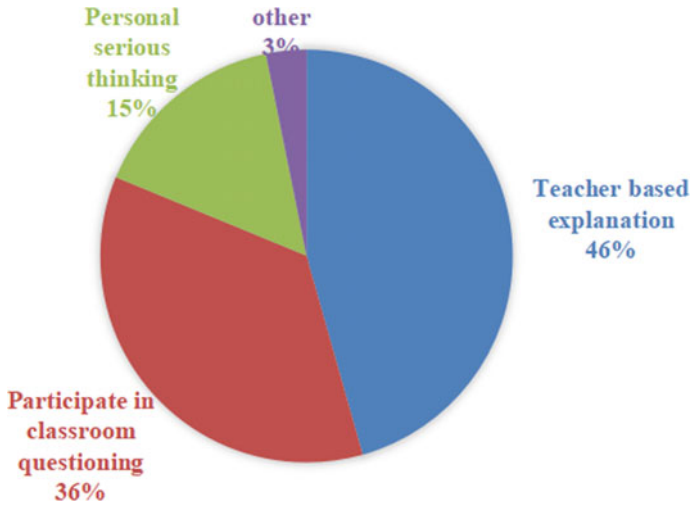


Fig. 33.3 Investigation on the mode of listening to college Chinese classes

cooperative inquiry in class. Through the collective teaching mode, teachers can appreciate the classic works in college Chinese, which will continue to guide students to learn independently. Through a variety of ways, the classroom can carry out cooperative inquiry mode, including brainstorming, mutual evaluation, and mutual testing. At the same time, the classroom can carry out independent inquiry mode, which will not only improve students' literary literacy, but also broaden their learning horizons. By giving full play to students' subjective initiative, we can improve the ability of analysis, identification, and evaluation of college Chinese. Third, after class reflection and evaluation. Through interactive communication, we can actively guide students' mutual evaluation, which will help students to summarize and reflect on their self-evaluation. By constantly revising the research results, students can understand the deficiencies before and in the course of testing and revising, which will increase the teaching mode based on the in-depth learning tools. Through mutual support, colleges and universities can promote students' independent learning ability, which will promote the formation of high-level cognitive ability [20].

33.5.2 The Construction of Deep Learning in College Chinese Teaching

Based on deep learning, college Chinese course can more scientifically exercise students' ability to deal with complex problems, which will promote students' ability to learn independently. Through cooperation and communication, colleges and universities can improve students' aesthetic taste, which can lead to deeper thinking. By improving the critical way of thinking, students can develop wholeheartedly in

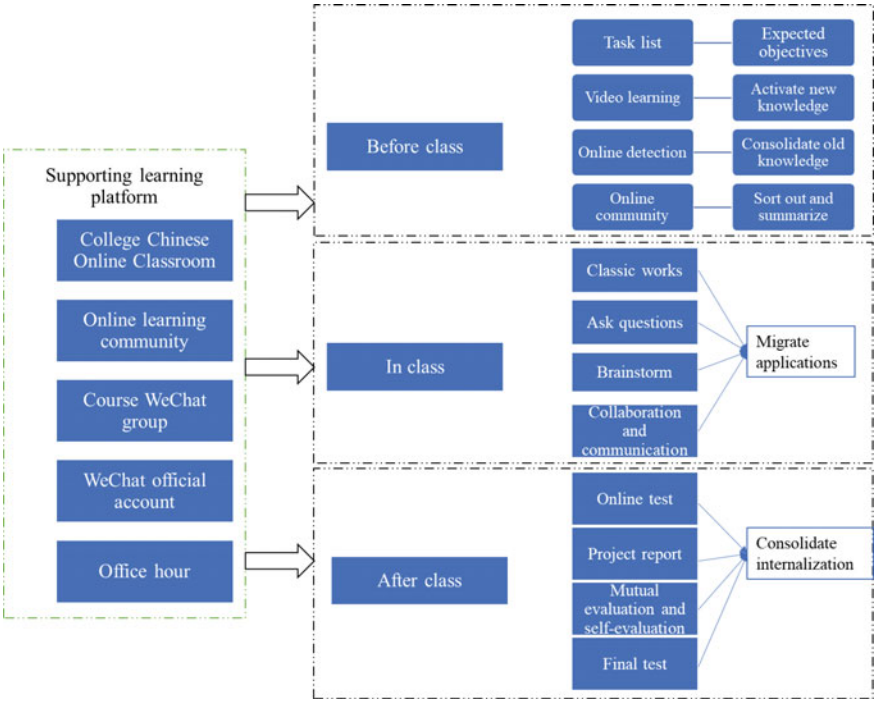


Fig. 33.4 Construction of college Chinese teaching based on deep learning

practice. Therefore, this paper constructs a teaching mode of deep learning in college Chinese teaching, as shown in Fig. 33.4.

33.6 Conclusion

Through in-depth learning, students will continue to enhance their enthusiasm for learning, which will also enhance their high-level thinking ability. Through in-depth learning, students will continue to improve their ability to solve complex problems, which will continue to improve the learning effect of students. Through the organic integration of MOOC and flipped classroom, we can build a meaningful deep learning field, which can provide an effective environment for knowledge construction and understanding, knowledge transfer and application, problem solving, and innovation.

References

1. Chu, M.: Research on Chinese reading teaching in primary schools from the perspective of deep learning. Chin. Charact. Cult. **19**, 104–106 (2022)
2. Lu, X.: Realizing the deep learning of primary school Chinese through speculative reading. Teach. Manag. **26**, 27–29 (2022)
3. Lingxia, O.: Research on annotated reading teaching of primary school Chinese based on "deep learning": taking the teaching of Cattle and Goose, Unit 6, Volume 1, Grade 4 as an example. J. Fujian Instit. Educ. **23**(08), 74–76 (2022)
4. Wan, P., Liangchen, T.: Teaching research on promoting deep learning with the support of information technology: research based on Chinese curriculum. Chin. Charact. Cult. **15**, 159–161 (2022)
5. Tao, G.: An analysis of the connotation, characteristics and realization path of the deep learning of Chinese in Junior Middle School. J. Kaifeng Vocat. Coll. Cult. Arts **42**(07), 92–94 (2022)
6. Miao, Q.: Large unit teaching: Chinese teaching reform for deep learning. Educ. Sci. Forum **23**, 44–47 (2022)
7. Bailing, S.: Scientific implementation strategy of primary school Chinese teaching based on deep learning. Taste Classics **13**, 166–168 (2022)
8. Chaoyang, J., Rongcai, L.: Practice and reflection on promoting primary school students' in-depth Chinese learning with information technology. Educ. Sci. Forum **20**, 73–75 (2022)
9. Liqiong, L.: Discussion on the construction of deep learning strategies in senior Chinese classroom. Chin. Teach. Newslett. D J. **06**, 22–24 (2022)
10. Ying, L.: Special teaching design of Chinese for secondary vocational school oriented to in-depth learning: taking the special topic of China's outstanding revolutionary works as an example. Vocational **11**, 87–90 (2022)
11. Mengmeng, H.: Let deep learning really happen: research and application of learning task groups in primary school Chinese teaching. J. Jilin Univ. Educ. **38**(06), 71–75 (2022)
12. Jianxia, F., Yi, W.: Reflections on the exercise design of compiling junior middle school Chinese textbooks from the perspective of deep learning. J. Huanggang Normal Univ. **42**(03), 28–31 (2022)
13. Lanyi, W.: The Construction and implementation of the assignment design model in the perspective of deep learning: taking the senior Chinese writing assignment design as an example. Educ. Sci. Forum **14**, 36–39 (2022)
14. Yuanzhen, Q.: Let deep learning enter the Chinese classroom: an effective strategy for deep learning of primary Chinese. Asia Pacif. Educ. **09**, 130–132 (2022)
15. Bin, J.: Research on the review strategy of primary school Chinese from the perspective of deep learning: take the final review of "summarizing the main contents of articles" in Volume II of Grade 4 of the Unified Edition as an example. Teach. Res. Natl. Common. Lang. 04, 150–152 (2022)
16. Fuling, L.: An example of the application of the concept of deep learning in Chinese teaching. Literat. Educ. **1**(04), 114–116 (2022)
17. Xia, Z., Wu, H.: Research on promoting the deep learning of primary school Chinese under the support of information technology. Chin. Charact. Cult. **06**, 121–122 (2022)
18. Gonglian, L.: Construction of Chinese learning task group based on deep learning. Chin. Constr. **03**, 4–8 (2022)
19. Su, Y.: Chinese reading teaching under the concept of deep learning. Chin. Constr. **03**, 9–13 (2022)
20. Xiao, Q., Lili, S.: On the promotion of classroom video images to in-depth learning: take the third grade Chinese classroom in Jiangsu Province as an example. Chin. Charact. Cult. **02**, 160–162 (2022)

Author Index

R. Kountchev et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2023)*, Smart Innovation, Systems and Technologies 368, https://doi.org/10.1007/978-981-99-6641-7