Ruidan Su
Han Liu   *Editors*

Medical Imaging and Computer-Aided Diagnosis

Proceeding of 2020 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2020)

MOREMEDIA ▶

≋ Springer

Lecture Notes in Electrical Engineering

Volume 633

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Associate Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Executive Editor (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Web of Science and Springerlink **

More information about this series at http://www.springer.com/series/7818

Ruidan Su · Han Liu
Editors

Medical Imaging and Computer-Aided Diagnosis

Proceeding of 2020 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2020)

Springer

Editors
Ruidan Su
Shanghai Advanced Research Institute
Chinese Academy of Sciences
Shanghai, China

Han Liu
Cardiff University
Cardiff, UK

Preface

Welcome to the proceedings of the International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2020) which was held in Oxford, UK, from January 20–21, 2020. MICAD is an annual conference which aims to provide a communication platform for top scholars, engineers, scientists, as well as graduate students to share ideas and discuss the latest technology in medical imaging and computer-aided diagnosis or related fields such as artificial intelligence and machine learning, to encourage growth, raising the profile of this multidisciplinary field with an ever increasing real-world applicability.

The diverse range of topics reflects the growth in the development and application of medical imaging and computer-aided diagnosis. The main topics covered in the proceedings are (i) optical imaging, (ii) image segmentation, (iii) biostatistics, (iv) applications of machine learning in imaging, (v) enhancement and reconstruction, (vi) diagnosis, classification, and treatment, and (vii) ophthalmology.

MICAD 2020 received submissions from 12 countries, in total, 56 full papers and 11 abstracts (the latter not considered for inclusion in this volume), and each paper was reviewed by at least three reviewers in a standard peer-review process. Based on the recommendation by three independent referees, finally 24 papers were accepted for MICAD 2020 (acceptance rate of 45%).

Many people have collaborated and worked hard to produce a successful MICAD 2020 conference. First, we would like to thank all the authors for submitting their papers to the conference, for their presentations and discussions during the conference. Our thanks go to the program committee members and reviewers, who carried out the most difficult work by carefully evaluating the submitted papers. Our special thanks to Yu-Dong Zhang, University of Leicester, UK, Manuchehr Soleimani, University of Bath, UK, Richard Jiang, Lancaster University, UK, and Edwin Abdurakman, City, University of London, UK, for the exciting keynote talks. We express our sincere thanks to the organizing committee chairs for helping us to formulate a rich technical program. We hope you enjoy the proceedings of MICAD 2020.

Ruidan Su

Organization

Honorable Chair

Marek Michalewicz   Deputy Director of the Interdisciplinary Center
                                for Mathematical and Computational
                                Modelling (ICM), University of Warsaw,
                                Poland

General Chair

Ruidan Su   Shanghai Advanced Research Institute, Chinese
                                Academy of Sciences, Shanghai, China

Program Chairs

Han Liu   Cardiff University, UK
Feng Dong   University of Strathclyde, UK
Zakaria Belhachmi   University of Haute-Alsace, France

Publication Chair

Ruidan Su   Shanghai Advanced Research Institute, Chinese
                                Academy of Sciences, Shanghai, China

Keynote Speakers

Yu-Dong Zhang   University of Leicester, UK
Manuchehr Soleimani   University of Bath, UK
Richard Jiang   Lancaster University, UK
Edwin Abdurakman   City, University of London, UK

Technical Program Committee

Smain Femmam (IEEE Senior Member)	University of Haute-Alsace France, France
Xin-She Yang	Middlesex University, UK
Mbaitiga Zacharie	Okinawa National College of Technology, Japan
Muthukumar Subramanian	National Institute of Technology, Puducherry, India
Jagath C. Rajapakse (IEEE Fellow)	Nanyang Technological University, Singapore
Sourav Dhar	Sikkim Manipal University, India
Feng Dong	University of Strathclyde, UK
Zakaria Belhachmi	Université Haute-Alsace, France
Sivarama Krishnan Rajaraman	Lister Hill National Center for Biomedical Communications (LHNCBC), National Library of Medicine (NLM), National Institutes of Health (NIH), India
Mahsa Mohaghegh	Auckland University of Technology, New Zealand
Joanna Collingwood	University of Warwick, UK
Gunasekar Thangarasu	Linton University College, Malaysia
Shou-Hsiung Cheng	Chienkuo Technology University, Taiwan
Dawar Khan	Interactive Media Design Laboratory, Nara Institute of Science and Technology, Japan
Vinesh Sukumar	University of Idaho, USA
Ehsan Kamrani	Bio-Optics Laboratory, Harvard Medical School, USA
Na Ma	Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China

Contents

Computer Modeling and Laser Stereolithography in Cranio-Orbital Reconstructive Surgery

Sergey A. Eolchiyan[1], Mikhail M. Novikov[2(✉)], and Svetlana A. Cherebylo[2]

[1] Neurotrauma Department, N. N. Burdenko National Medical Research Center of Neurosurgery, 4-th Tverskaya-Yamskaya Street, 16, 125 047 Moscow, Russia
Seolch@nsi.ru

[2] Institute on Laser and Information Technologies of RAS – Branch of Federal Scientific Research Centre "Crystallography and Photonics" of RAS, Shatura, Moscow Region, Russia
{novikov,Svetlana.cherebylo}@rambler.ru

Abstract. Digital technologies, namely computer modeling and laser stereolithography, were used to facilitate reconstructive surgery by means of precise individual 3D digital models, surgical templates and implant molds in 191 patients with complex cranio-orbital defects and deformities. Of these, 149 underwent reconstruction of thin-walled structures - orbital floor and medial wall. The patients aged between 5 and 51, were operated on in the period of 2002–2018, their management was unified. Special technique ("virtual pencil"), permitting 2D images of intact and damaged orbital thin walls editing, was developed. It enabled to avoid further data processing errors, while converting DICOM files to 3D model, and obtain real defects dimensions and geometry of the damaged area. Thus, pre-op planning, particular surgical procedure stages simulating, precise patient-specific implants fabrication for cranio-orbital and face skeleton reconstruction became possible. Our experience proves that computer-aided design systems and additive technologies enable optimal surgical treatment of improved efficacy.

Keywords: Laser stereolithography · Computer tomography · Cranio-orbital defects · Reconstructive surgery · Computer-aided design (CAD) · Computer-aided surgery (CAS)

1 Introduction

Orbital walls' combined or isolated fractures are not rear in cranio-facial trauma patients, and successful cranio-orbital reconstruction is still a challenge even for good trained and experienced surgeons. Restoration of injured structures' anatomy and functions demands precise tree-dimensional model of damage area, essential for biocompatible implants design, production and modeling [1]. Increasing quality of life requirements encourage surgeons to search for new opportunities to improve the accuracy of pre-op planning and surgery [2]. Computed tomography is diagnostic method of choice for cranio-orbital

injury patients' evaluation and surgery planning, as it provides adequate information concerning surgical anatomy in a multiplanar, high-precise three-dimensional reconstruction (individual digital model) provides spatial visualization of the damaged area and allows stereolithographic physical model production as well as patient-specific implant design and 3D mold model creation for its fabrication using biocompatible materials [3]. Modern high-resolution computed tomography provides axial X-rays series (computer slices) of specified step (0.2–1 mm). Each computer slice has limited spatial resolution equal to one pixel (up to 0.2 mm). Thus, orbital walls less than 0.1–0.2 mm thick are not automatically recognized as bone structures by image processing programs and corresponding data are not converted into 3D digital model [4, 5]. Discontinuous CT image of orbital floor, roof (partially) and medial wall can lead to diagnostic mistakes [6]. Thus, 3D digital model construction creates false holes with diameter of 10 mm and more. Diameter size depends on computer tomograph capabilities and settings and soft tissue cutoff threshold setting. One can estimate real boundaries of orbital wall fracture or defect only after these false gaps are removed.

This paper presents the technique developed for personal 3D digital model creation using edited 2D images of intact and damaged orbital thin walls. It enabled production of stereolithographic model with correct damaged area display, pre-op surgical planning and manufacturing of precise implants to restore original anatomy of cranial vault and adjacent orbital and facial skeleton structures.

2   Materials and Methods

Totally 191 patients with post-traumatic cranio-orbital defects and deformities aged between 5 and 51, underwent reconstructive surgery utilizing computer modeling and laser stereolithography in Burdenko National Medical Research Center of Neurosurgery from 2002 to 2018. All of them were operated on by the first author.

The decision on the reconstructive operation was made by the surgeon after evaluation of patient objective and CT data. CT data in DICOM format were sent to ILIT RAS to technical specialists (second and third authors) via internet. To create the most informative digital model first of all, each CT slice with orbital thin-walled structures image was sequentially edited using the so-called virtual pencil. The boundaries of medial, lower and upper orbital walls bone structures were drawn to increase their thickness in different projections (axial, sagittal and coronary) on both sides (intact and damaged). In true defect area there are no boundaries of dark and light gray shades. Then CT data were processed using Mimics software (Materialise, Belgium) to convert to a 3D model in the STL format used in laser stereolithography facilities.

Patient-specific implant templates were created based on the mirror mapping of the intact side to the damaged side and model boolean operation. The next step – precise patient's skull plastic copy reproducing defects and deformations and high-precision implants' molds laser stereolithography manufacturing. Plastic models and implant's molds were sterilized and used during the operation.

3   Results

DICOM files editing allowed false gaps elimination and direct visualization of orbital walls defects true boundaries on a personal digital model (Fig. 1).
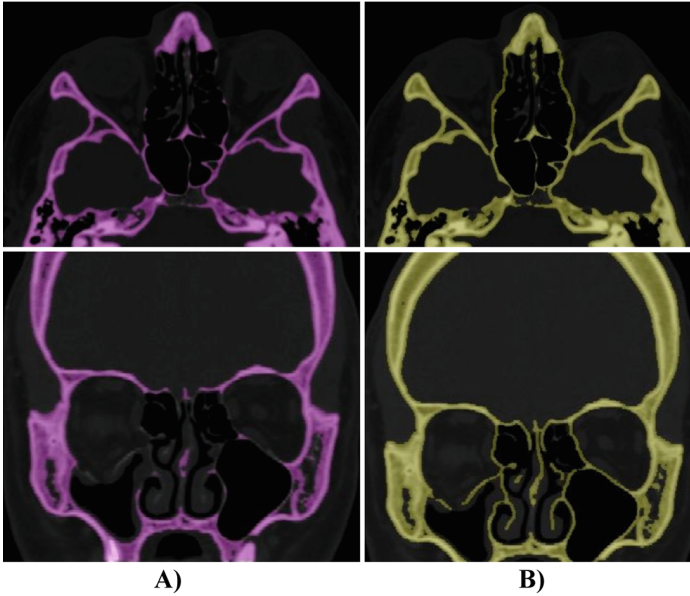
Fig. 1. Axial and sagittal CT images, A) before orbital walls editing, B) after orbital walls editing.
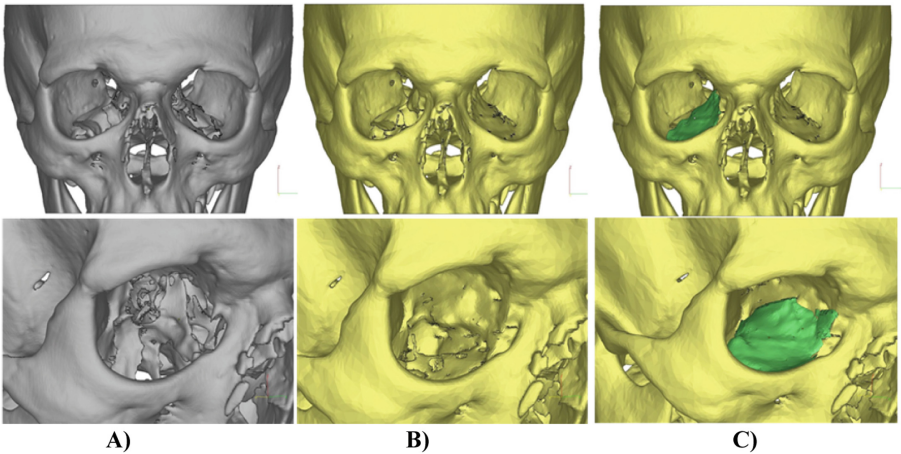
Fig. 2. Three-dimensional digital models created using CT data: A) before orbital walls editing, B) after orbital walls editing with increased thickness, C) mirror mapping of the intact orbital floor to the damaged side.
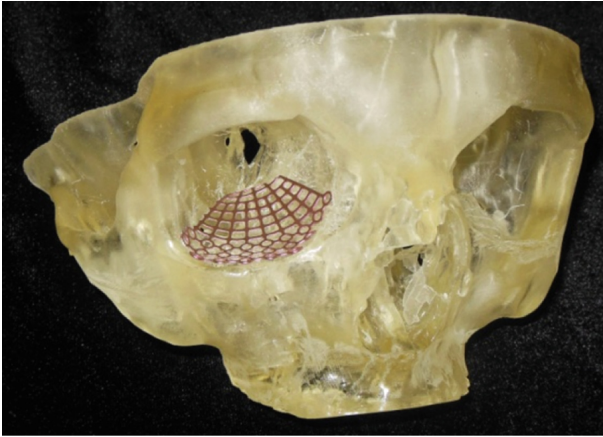
Fig. 3. Fitting a simulated titanium implant on a stereolithographic model.

A)   **B)**   **C)**
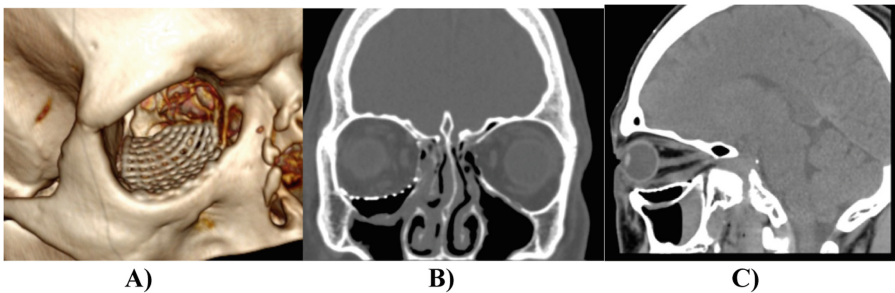
Fig. 4. Postoperative CT examination: A) 3D reconstruction. B) Frontal sections. C) Sagittal section.

Figure 2 shows the difference between 3D digital models created using the standard methods before and after CT data editing. Orbital walls exact geometry restoration enables individual implant of correct shape design and fabrication well in advance before or during the operation, facilitating surgery, reducing its duration and improving surgical treatment quality.

Different surgical manipulations, such as osteotomy lines defining, bone fragments movement, reposition and fixation were imitated by the surgeon utilizing 3D digital model and its plastic copy in pre-op planning and preparation (Fig. 3). Titanium mesh implants and templates for cranial bone autografts of different shape and size, depending on defect geometry, harvesting were made also. Directly during intervention surgical templates and molds were used for polymethyl methacrylate or titanium mesh plate individual implants fabrication. High information content of full-size skull stereolithographic models, created using edited DICOM files and realistically displaying orbital thin-walled bone structures lesions and defects was confirmed during surgery in all cases (Fig. 4).

4   Discussion

In the mid-1990s, new diagnostic methods (computer and magnetic resonance imaging) as well as computer modeling and additive manufacturing systems development enabled new surgical treatment technologies emergence and introduction into clinical practice. Originally they were called Image Guided Surgery and afterwards Computer Aided or Assisted Surgery (CAS). CAS-technologies solve the problems of the surgeon's work automated support and increase its efficiency. One of the most promising CAS-technologies application fields lies in skull and facial skeleton reconstruction. The surgeons and technical specialists together prepare CASE-technologies-facilitated surgical intervention. During preoperative planning special high-resolution CT data processing programs are used. Different software applications for automated modeling and manufacturing, including additive technologies (CAD/CAM-AM) are used for damaged anatomical region physical models, surgical templates, individual implants and individual implants molds production [7, 8].

This approach to surgical treatment increase patient's safety level and reduce postoperative complications risks [9, 10]. In complex acquired cranio-orbital defects and deformities computer modeling and laser stereolithography have an advantage over CT-examination for reconstructive surgery pre-op planning and performing and enables surgical treatment quality improvement [7, 11]. Sufficient accuracy of personal digital 3D model and its plastic copy is essential. Currently, software systems allowing two-dimensional computer slices of Dicom format transition to 3D modeling are widely used. One of spread programs is Mimics (Materialise). It allows to download grayscale images of patient's CT tomograms and using the grayscale selection functions (Hounsfield scale), select fragments with values corresponding to the density of the human body bone tissue alone.

Unfortunately, the Mimics program and its analogues do not allow correct reproduction of the orbital medial, upper and lower walls entire surface in automatic mode, selecting the parameters of the gray value. Since these walls are quite thin, there are false gaps in their display. One can see defect true boundaries or fracture fragments displacement only by their removing. Sequential editing of orbital thin-walled structures with artificial thickening enables construction of highly informative three-dimensional computer model with clear contours, allowing getting its real copy without false holes preventing defect examination and implant creation.

5   Conclusion

Proposed technique of thin-walled orbital structures CT data editing avoided errors due to each slice limited spatial resolution, exceeding orbital floor, roof (partially) and medial wall thickness. It allows higher precision of orbital thin-walled structures' defects dimensions and geometry reproduction resulting in patient-specific implants higher accuracy enabling predictable good functional and esthetic outcome of surgical treatment.

Conflict of Interest.   The authors declare that they have no conflict of interest.

References

1. Stoor, P., et al.: Rapid prototyped patient specific implants for reconstruction of orbital wall defects. J. Craniomaxillofac. Surg. **42**, 1644–1649 (2014)
2. Sharaf, B., et al.: Importance of computer- aided design and manufacturing technology in the multidisciplinary approach to head and neck reconstruction. J. Craniofac. Sur. **21**(4), 1277–1280 (2010)
3. Eolchiyan, S.A.: Complex skull defects reconstruction with CAD/CAM titanium and polyetheretherketone (PEEK) implants. Zh Vopr Neirokhir Im N N Burdenko **78**(4), 3–12 (2014)
4. Jones, D., Evans, J.: Blow-out fractures of the orbit: an investigation into their anatomical basis. J. Laryngol. Otol. **81**(10), 1109–1120 (1967). https://doi.org/10.1017/S00222151000 68122
5. Straker, C., Hill, J.: Management of orbital blow-out fractures. S. Afr. Med. J. **76**, 535–537 (1989). http://archive.samj.org.za/1989%20VOL%20LXXVI%20Jul-Dec/Articles/11% 20November/2.5%20MANAGEMENT%20OF%20ORBITAL%20BLOW-OUT%20FRAC TURES.%20C.A.%20Straker%20and%20J.C.%20Hill.pdf
6. Radnionok, A., et al.: Algorithm of processing microspiral ct-scan results for constructing a three-dimensional model of orbit thin bones. J. Eng. Sci. **25**(4), 39–46 (2018)
7. Huotilainen, E., et al.: Imaging requirements for medical applications of additive manufacturing. Acta Radiol. **55**(1), 78–85 (2014)
8. Tack, P., et al.: 3D-printing techniques in a medical setting: a systematic literature review. BioMed. Eng. OnLine (2016). https://doi.org/10.1186/s12938-016-0236-4
9. Potapov, A.A., et al.: Modern technology in the surgical treatment of head injury sequelae. Ann. Russian Acad. Med. Sci. **67**(9), 31–38 (2012). (In Russia)
10. Eropkin, S.V., et al.: Prospects for reconstructive skull surgery using computed tomography and stereolithography Burdenko's. J. Neurosurg. **2**, 53–57 (2002)
11. Kravchuk, A.D., et al.: Additive technologies in neurosurgery Burdenko's. J. Neurosurg. **82**(6), 96–103 (2018)

Sparse Representation Label Fusion Method Combining Pixel Grayscale Weight for Brain MR Segmentation

Pengcheng Li and Monan Wang[(✉)]

School of Mechanical and Power Engineering, Harbin University of Science and Technology,
Harbin 150080, Heilongjiang, People's Republic of China
mnwang@hrbust.edu.cn

Abstract. Multi-atlas based segmentation (MAS) methods have demonstrated superior performance in the field of automatic image segmentation, and label fusion is an important part of MAS methods. In this paper, we propose a sparse representation label fusion (SRLF) method combining pixel grayscale weight. We adopt a strategy for solving sparse coefficients multiple times and introduce pixel grayscale weight information in the label fusion process. In order to verify the segmentation performance, we apply the proposed method to segment subcutaneous tissues in 3D brain MR images of the challenging publicly available IBSR datasets. The results show that our method effectively improves the defects of SRLF method and achieves higher segmentation accuracy. We also compared our methods with commonly used automatic segmentation tools and state-of-the-art methods, and the average Dice similarity coefficient (Dsc) of the subcutaneous tissues obtained by our method was significantly higher than that of the automatic segmentation tools and state-of-the-art methods.

Keywords: Image segmentation · Multi-atlas segmentation · Label fusion

1 Introduction

Brain MR image segmentation is widely used in clinical practice, and the segmentation accuracy directly affects the disease diagnosis results. Manual segmentation is a commonly used method, but it is time-consuming and subjective. In addition, a large number of brain images are generated every day in the hospital, and it is a big burden to doctors if they are processed by manual segmentation. Therefore, an automated image segmentation method is needed to achieve routine analysis of brain MR images in the clinic. FreeSurfer [1] and FIRST [2] are widely used automatic segmentation tools and can alleviate the burden of manual segmentation, but their accuracy still needs to be improved.

Some automated image segmentation methods [3] have been proposed in the past decade. And multi-atlas-based segmentation (MAS) methods showed better segmentation performance compared to other methods, which can realize the automatic segmentation of images by using the prior information of the atlas. The MAS methods

mainly consist of two steps. Firstly, each atlas image is registered with the target image to map the atlas information to the target image space. Secondly, a label fusion method is used to fuse the atlas label information, Subsequently, the target organization is classified according to the fusion result. There are many image registration methods with better registration results [4], and image registration methods proposed recently have not been significantly improved. Therefore, most MAS methods use existing nonlinear registration.

Label fusion is a key step in MAS methods, which can reduce registration errors and improves segmentation accuracy. Majority voting (MV) [5] and Weighted voting (WV) are widely used label fusion methods. The MV methods assign the most frequent label to the target pixel. The WV methods assign different weight values to different labels and can obtain better fusion results than the MV methods. The intensity similarity between atlas images and the target image is mainly used as the weight value, such as the absolute value of the intensity difference [6], the Gaussian function of the intensity difference [7], the local mutual information [8]. Lin et al. [9] proposed a label fusion method that combining registration error and intensity similarity.

In recent years, a multi-atlas patch-based label fusion method has been proposed [10]. The target voxel label is determined by fusing the corresponding voxel patches from each of the atlas label images. In this method, each voxel is represented by a feature vector which is constructed from a patch centered at this voxel. A common method is to construct feature vectors using patch pixel strength [11]. Bai et al. [12] extracted more information from the patch to build the feature vector, including intensity, gradient, and contextual information. Roy et al. [13] proposed a sparse representation label fusion (SRLF) method for image tissue classification. This method uses the sparse dictionary learning and the whole segmentation process does not require deformation registration of the atlas images. Tong et al. [14] combined a discriminative dictionary learning and sparse coding technique to design a fixed dictionary for hippocampus segmentation. Lee et al. [15] proposed a brain tumor image segmentation method using kernel dictionary learning. This method implicitly maps intensity features to high-dimensional features to make classification more accurate. However, this method is time-consuming and cannot be widely applied. In order to deal with this problem, a linearized kernel sparse representative classifier is proposed [16]. Another commonly used algorithm in the field of automatic image segmentation is Learning-based segmentation algorithm, such as random forest (RF) [17] and convolutional neural network (CNN) [18] methods. CNN method has excellent feature extraction capabilities. Kaisar et al. [19] used a CNN method to achieve segmentation of the subcutaneous tissue.

One limitation of the label fusion method based on sparse representation is that most patches are discarded during the fusion process. This is mainly because the number of non-zero coefficients in the sparse solution is much smaller than the number of patches. This will result in some pixels in the target organization not being accurately identified.

In this paper, in order to overcome the shortcomings of the SRLF method mentioned above, we adopt a strategy for solving sparse coefficients multiple times and introduce pixel grayscale weight information in the label fusion process. In order to verify the segmentation performance of our method. We tested the proposed method to segment subcutaneous tissues of the challenging publicly available IBSR dataset [20].

2   Methods

As shown in Fig. 1, the proposed method mainly includes four steps: atlas registration, sparse representation method, pixel grayscale weight setting, and label fusion.
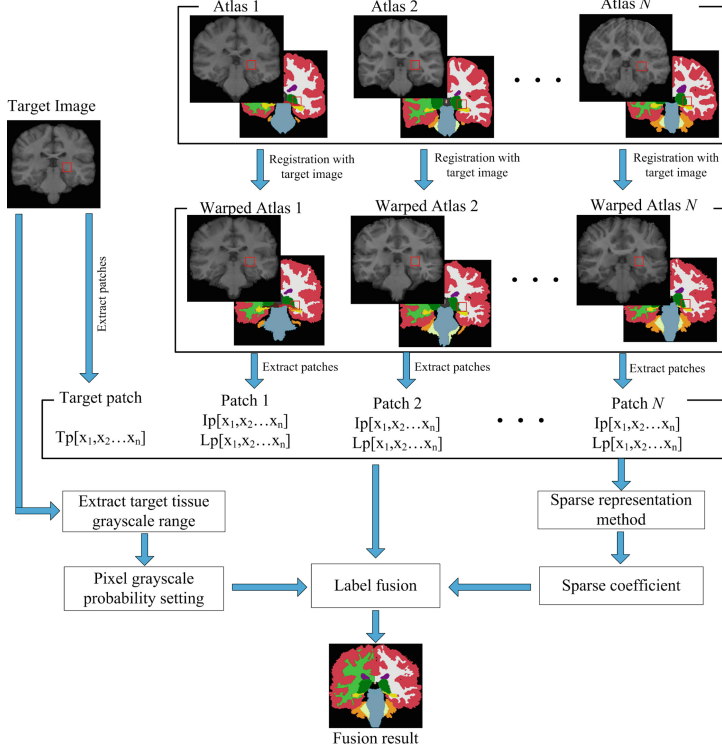
Fig. 1. Schematic diagram of the proposed method.

2.1   Atlas Registration

The image to be segmented is taken as the target image T. We selected the SuperElastix registration tool [21] and set the affine and B-spline deformation method to register the target image and the atlas intensity images. Saving the warped atlas $A_1(I_1, L_1), A_2(I_2, L_2), \cdots A_n(I_n, L_n)$, I_i is the warped intensity images of atlas A_i, L_i is the warped label images of atlas A_i, and n is the number of atlas.

2.2   Parse Representation Method

Extract Patches: Extracting patch Tp_{xj} centering on the pixel x_j of the target image, and then extract patch Ip_i and Lp_i from the x_j position of the I_i and L_i.

Construction Vector: In this paper, we use $PT[pt_1, pt_2 \cdots pt_{len}]^T$ to represent Tp_{xj}. pt_i corresponds to each pixel in the patch Tp_{xj}, len is equal to the number of pixels in the Tp_{xj}. Ip_i is represented by a column vector $PI_i[pi_{1i}, pi_{2i} \cdots pi_{leni}]^T$, $i = 1, 2, \cdots m$, and the over-complete dictionary D_I is represented as $D_I[PI_1, PI_2 \cdots PI_m]$. Sparse coefficient $\alpha = [\alpha_1, \alpha_2 \cdots \alpha_m]^T \in Rn$. Each element of the PT can be represented by the following formula:

$$pt_j = pi_{j1}\alpha_1 + pi_{j2}\alpha_2 + \cdots pi_{jm}\alpha_m \tag{1}$$

$$PT = D_I\alpha \tag{2}$$

The idea of sparse representation is to represent signal PT by solving the most sparse solution α. The α can be obtained by solving the following equation:

$$\hat{\alpha} = \arg\min\|\alpha\|_0 \quad \text{subject to} \quad \|PT_i - D\alpha\|_2^2 \leq \varepsilon \tag{3}$$

The solution α is used as the weight w_i of WV, and then the label fusion is performed by using (1).

The number of non-zero parameters in the sparse solution obtained by (3) is less than or equal to *len*. Therefore, most patches are discarded during the fusion process, and this causes some pixels in the target organization that cannot be accurately identified. To address this problem, we remove the elements corresponding to the coefficients that are not equal to 0 from the over-complete dictionary after a sparse coefficient is solved. And then construct a new sparse representation equation.

The number of patches participating in the fusion will be expanded by solving sparse coefficients multiple times.

2.3 Pixel Grayscale Weight Setting

In order to make full use of the prior information of atlas, we introduce a grayscale weight in label fusion method.

Target Tissue Grayscale Range Setting. a. After the images are normalized, the pixel value ranges from 0 to 1. Divide the pixel value range 0–1 into N_I intervals and then extract the grayscale distribution map. Judge the target tissue distribution range based on the peak value of the grayscale map. If the number of peaks is 2 or 3, the upper boundary value of the target organization is set to $ISN_{max} + Ov$, where ISN_{max} is the maximum interval serial number corresponding to the peak and Ov is the offset value.

b. If the number of peaks is 1, then divide the pixel value range 0–1 into $2 \times N_I$ intervals and repeat step 1.

c. If the number of peaks is greater than 3, then divide the pixel value range 0–1 into $N_I/2$ intervals and repeat step 1.

Pixel Grayscale Weight Setting. As shown in Fig. 2, the pixel grey value span of subcutaneous tissue is between 0.3 and 0.6. In this paper, we set the pixel value span of thalamus to 0.6, and set the pixel value span of other subcutaneous tissues to 0.5 by analyzing the grayscale distribution of the target tissue. And the maximum pixel

values of thalamus, hippocampus, caudate, putamen, pallidum and amygdala are set to $ISN_{max} + 0.15$, ISN_{max}, ISN_{max}, $ISN_{max} + 0.1$, $ISN_{max} + 0.15$ and $ISN_{max} - 0.05$, respectively.

It can be seen from Fig. 2, the target tissue gray distribution curve approximates the positive distribution. In this paper, we divide the pixel value span of the thalamus into 12 intervals, and the pixel grayscale weight of each interval is set to [0.8, 0.8, 0.9, 0.9, 1, 1, 1, 1, 1, 1, 0.9, 0.9]. We divide the pixel value span of hippocampus, caudate, putamen, pallidum and amygdala into 10 intervals, and the pixel grayscale weight of each interval is set to [0.8, 0.9, 0.9, 1, 1, 1, 1, 1, 1, 0.9].
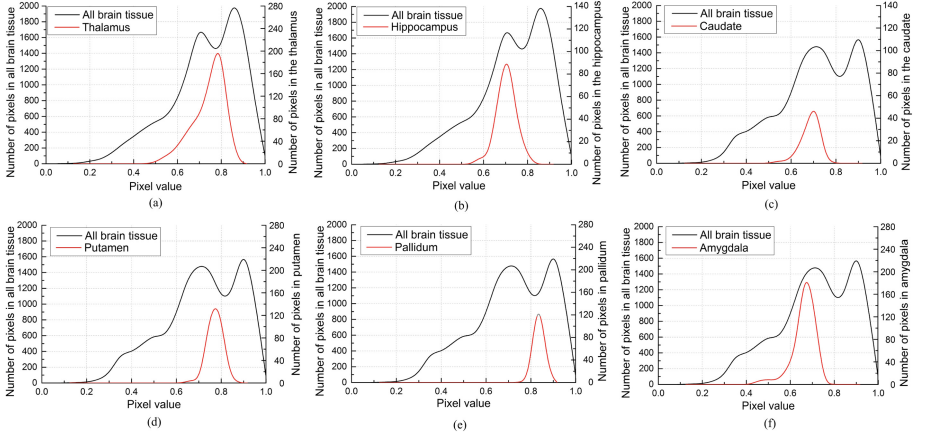
Fig. 2. Subcutaneous tissues grayscale distribution map, (a) Thalamus, (b) Hippocampus, (c) Caudate, (d) Putamen, (e) Pallidum, (f) Amygdala.

2.4 Label Fusion

A label fusion is performed by combining the sparse coefficient and the pixel grayscale weight obtained by the above steps. The fusion formula is as follows:

$$Fv(x) = \beta \times P(x) \times \frac{\sum_{i=0}^{n} w_i L p_i}{\sum_{i=0}^{n} w_i} \tag{4}$$

$$L_T(x_j) = \begin{cases} 1, & \text{if } Fv(x) \geq 0.5 \\ 0 & \text{if } Fv(x) < 0.5 \end{cases} \quad j = 1, 2 \cdots m \tag{5}$$

Where $Fv(x_j)$ is the fusion result, $P(x)$ is the pixel grayscale weight, w_i is the sparse coefficient. If $L_T(x_j) = 0$, the pixel x_j belongs to background, and if $L_T(x_j) = 1$, the pixel x_j belongs to target organization. m is the number of target image pixels.

3 Experiments and Results

In this section, we use the leave-one-out cross-validation method to verify the segmentation performance and robustness of the proposed method.

3.1   Segmentation Evaluation Index

We applied the Dice similarity coefficient (Dsc) and Hausdorff distance (HD) segmentation metrics to measure the segmentation result of the proposed method. The metrics are defined as follows:

$$Dsc(T, F) = \frac{2V(T \cap F)}{V(T) + V(F)} \tag{6}$$

$$HD(T, F) = \max(H_1(T, F), H_2(T, F)) \tag{7}$$

$$H_1(T, F) = \max_{p_t \in T} \left(\min_{p_f \in F} d\left(p_t, p_f\right) \right) \tag{8}$$

$$H_2(F, T) = \max_{p_f \in F} \left(\min_{p_t \in T} d\left(p_f, p_t\right) \right) \tag{9}$$

Where T is the target tissue pixel set in the target label image, F is the segmented target tissue pixel set and $d\left(p_t, p_f\right)$ is the distance between pixels p_t and p_f.

3.2   Influence of the Iterations

The choice of patches has a greater impact on the segmentation results. In this experiment, we select the top 80 patches with high similarity to build the over-complete dictionary. The subcutaneous tissues belong to the gray matter (GM) or white matter (WM). Therefore, so we use the GM and WM as the target tissues to test the influence of iterations on the segmentation results. As shown in Fig. 3, the highest WM and GM Dsc were obtained when iterations is 5. For each additional iterations, the label fusion time is only increased by a few seconds. So we set the iterations to 5.
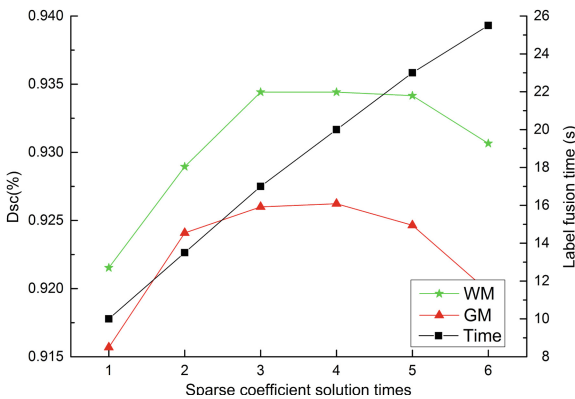
Fig. 3. Segmentation results of different iterations. The green line is WM segmentation result, the red line is GM segmentation result, and the black line is the registration time.

3.3   Detailed Segmentation Results

We used the proposed method to segment subcutaneous tissues including thalamus, hippocampus, caudate, putamen, pallidum and amygdala. According to the segmentation test, the fusion weight β of thalamus and pallidum is set to 1.25, the fusion weight β of hippocampus, caudate and putamen is set to 1.12, the fusion weight β of amygdala is set to 1.43.
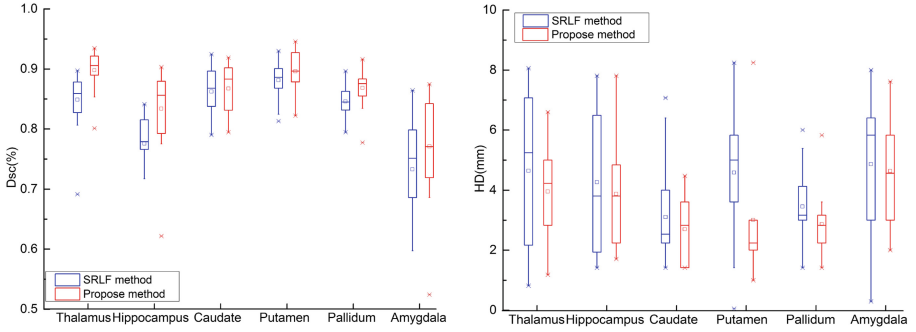
Fig. 4. Subcutaneous tissue segmentation results of the proposed method and SRLF method.
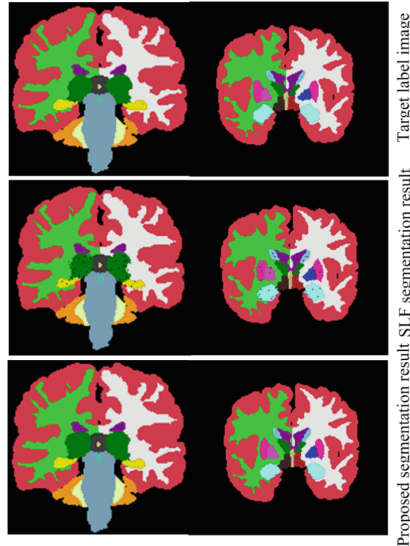
Fig. 5. Sample segmentation results of the proposed method and SRLF method.

Figure 4, shows the subcutaneous tissue segmentation results of the proposed method and SRLF method. It can be seen from the figure. Compared to the SRLF method, our method obtained higher Dsc and lower HD for the segmentation of all six subcutaneous

tissues. This shows that the proposed method can effectively improve segmentation accuracy. Figure 5, is the sample segmentation results of the proposed method and SRLF method. As shown in this figure, many pixels in the target organization which segmented by the SRLF method are not accurately identified. The proposed method effectively overcomes this shortcoming and obtains a segmentation result more similar to the target label image.

Table 1. Comparison of the proposed method with state-of-the-art methods on IBSR dataset in terms of Dsc.

Target tissues	FIRST [2]	FreeSurfer [1]	Lin's method [9]	Liu's method [16]	Kaisar's method [19]	Proposed method
Thalamus	0.889	0.840	0.870	0.904	**0.912**	0.898
Hippocampus	0.810	0.789	0.770	0.829	0.831	**0.834**
Caudate	0.817	0.805	0.800	0.862	0.866	**0.867**
Putamen	0.868	0.785	0.840	0.897	**0.902**	0.896
Pallidum	0.807	0.743	0.730	0.814	0.827	**0.869**
Amygdala	0.714	0.581	0.690	0.753	0.765	**0.771**
Average	0.818	0.757	0.783	0.843	0.851	**0.856**

We also compared the proposed method with commonly used automatic segmentation tools and state-of-the-art methods on the IBSR dataset. As shown in Table 1, the proposed method showed a better performance in comparison to both FIRST and FreeSurfer methods for the six subcutaneous tissues. The overall Dsc means of our method were significantly higher than both of the methods, with mean Dsc of 0.818, 0.757 and 0.856 for FIRST, FreeSurfer and the proposed method. In thalamic segmentation, Kaisar's method obtained the highest segmentation accuracy with Dsc of 0.912 and we obtained a Dsc of 0.898. In putamen segmentation, our method obtains a similar segmentation result with Kaisar's method. In hippocampus, pallidum and amygdala segmentation, the proposed method obtained the highest Dsc value. In summary, the proposed method achieved a better segmentation performance compared to other methods.

4   Discussion and Conclusion

In this paper, we propose an automatic image segmentation method using SRLF method combining pixel grayscale weights. One limitation of the SRLF method is that some target tissue pixels cannot be accurately identified. The proposed method can solve this defect by solving the sparse solution multiple times. In order to improve the segmentation accuracy, a pixel grayscale weight is introduced into the label fusion method. Extracting grayscale distribution range of the target tissues according to its grayscale map, and then

setting pixel grayscale weight according to the pixel grayscale distribution information of the target tissues.

In the experiment, amygdala obtained the lowest segmentation accuracy compared to other subcutaneous tissues. This is mainly because the amygdala's volume is the smallest, and its registration result is not as good as other subcutaneous organizations. In fact, the registration results obtained using the same registration method for different tissues or images are different. Therefore, selecting the appropriate registration method for each organization can effectively improve the segmentation accuracy of MAS methods.

Due to the large difference between images, it is not appropriate to set the same pixel grayscale weight for different images. In the proposed method, we set a pixel grayscale weight by considering all the atlas images, but this weight is not optimal for each image. Therefore, it is a good way to improve segmentation accuracy by setting appropriate weights for different images.

References

1. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
2. Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M.: A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage **56**(3), 907–922 (2011)
3. Sandra, G., et al.: A review on brain structures segmentation in magnetic resonance imaging. Artif. Intell. Med. **73**, 45–69 (2016)
4. Wang, M., Li, P.: A review of deformation models in medical image registration. J. Med. Biol. Eng. **39**(1), 1–17 (2018)
5. Collins, D.L., Pruessner, J.C.: Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. NeuroImage **52**(4), 1355–1366 (2010)
6. Isgum, I., et al.: Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. IEEE Trans. Med. Imaging **28**(7), 1000–1010 (2009)
7. Sabuncu, M.R., et al.: A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imaging **29**(10), 1714–1729 (2010)
8. Nie, J., Shen, D.: Automated segmentation of mouse brain images using multi-atlas multi-ROI deformation and label fusion. Neuroinformatics **11**(1), 35–45 (2013)
9. Lin, X.B., Li, X.X., Guo, D.M.: Registration error and intensity similarity based label fusion for segmentation. IRBM **40**(2), 78–85 (2019)
10. Sanroma, G., et al.: A transversal approach for patch-based label fusion via matrix completion. Med. Image Anal. **24**(1), 135–148 (2015)
11. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. IEEE Trans. Med. Imaging **30**(10), 1852–1862 (2011)
12. Bai, W., et al.: Multi-atlas segmentation with augmented features for cardiac MR images. Med. Image Anal. **19**(1), 98–109 (2015)
13. Roy, S., et al.: Subject-specific sparse dictionary learning for atlas-based brain MRI segmentation. IEEE J. Biomed. Health Inform. **19**(5), 1598–1609 (2015)
14. Tong, Y., et al.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage **76**(1), 11–23 (2013)
15. Lee, J., Kim, S.J., Chen, R., Herskovits, E.H.: Brain tumor image segmentation using kernel dictionary learning. In: Proceedings of 37th Annual International Conference, pp. 658–661. IEEE EMBC, Milan (2015)

16. Liu, Y., Wei, Y., Wang, C.: Subcortical brain segmentation based on atlas registration and linearized kernel sparse representative classifier. IEEE Access **7**, 31547–31557 (2019)
17. Zikic, D., Glocker, B., Criminisi, A.: Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. Med. Image Anal. **18**(9), 1262–1273 (2014)
18. Moeskops, P., et al.: Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans. Med. Imaging **35**(5), 1252–1261 (2016)
19. Kaisar, K., et al.: Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. Med. Image Anal. **48**, 177–186 (2018)
20. IBSR Homepage. https://www.nitrc.org/projects/ibsr. Accessed 6 Nov 2019
21. SuperElastix Homepage. https://github.com/SuperElastix/. Accessed 6 Nov 2019

Deep Learning for Mental Illness Detection Using Brain SPECT Imaging

Felisa J. Vázquez-Abad[1,2], Silvano Bernabel[1,2], Daniel Dufresne[3(✉)],
Rishi Sood[4], Thomas Ward[4], and Daniel Amen[5]

[1] Department of Computer Science, Hunter College CUNY, New York City, USA
[2] School of Computing and Information Systems, University of Melbourne,
Melbourne, Australia
[3] Department of Mathematics and Statistics, Concordia University,
Montreal, Canada
`daniel@ozdaniel.com`
[4] Amen Clinics, New York, NY, USA
[5] Amen Clinics, Costa Mesa, CA, USA

Abstract. We apply deep learning to the detection of mental illness
with meaningful results, using SPECT (Single Photon Emission Computed Tomography) images of the brain. The data consists in scans from patients with attention deficit hyperactivity disorder (ADHD), major depressive disorder (MDD) and obsessive compulsive disorder (OCD), plus scans of healthy brains. We focus here on the application of a deep convolutional neural network (CNN). The main challenge in using CNN models for medical diagnosis is often the number of samples not being sufficiently large to ensure high accuracy. We propose a soft classifier for using the machine. Instead of a binary output "yes/no" for each condition, we add an intermediate outcome, which says that the machine yields a weak result. The "Red Zone" corresponds to a positive result (condition is present) and the "Green Zone" corresponds to a negative, each with a preassigned statistical confidence level; the "Amber Zone" is an ambiguous outcome, where the scan is assigned a likelihood of having the condition. This information is then passed to the doctors for further analysis of patients.

Keywords: Deep learning · SPECT scans · Artificial intelligence · Mental health

1 Introduction

One tool used in psychiatry today is SPECT (Single Photon Emission Computed Tomography) brain images. Our research consists in using SPECT images for visual recognition; we apply artificial intelligence to determine if there is indeed important information embedded in these images that is critical for achieving

Daniel Amen and Thomas Ward are co-last authors.

accurate diagnosis. Our algorithms use **only** the SPECT images to arrive at a patient's diagnosis, while conventionally psychiatrists gather a clinical history using other tests besides brain scans.

SPECT imaging has been used to evaluate a variety of conditions since the 1970s. In this paper, we find that machine learning analysis of SPECT images on its own yields accuracies between 78% and 99% when diagnosing ADHD.

We believe our algorithms would provide a first indication to primary care doctors that would enable them to refer patients to psychiatrists earlier in the development of illnesses such as ADHD. We aim at building algorithms that will be easy to compute and can be attached to the SPECT scanners, providing the results as well as the images immediately. With this in mind, it is our goal to create algorithms that will help early detection, yet accept input of small dimensions and compute relatively fast.

In deep learning, it is usually recommended to use at least 50,000 samples to achieve good accuracy. For example, in [2] brain scans are used to train a deep network that classifies tumor cells. Those authors use 2.5 and 3.4 million samples (each corresponds to a pixel in a brain image). In our project, each whole brain SPECT image is just one sample, because the goal is to detect mental illness of the whole brain, rather than classifying smaller regions as tumors as in [2]. The Amen Clinics has accumulated brain SPECT images over the past 28 years. The largest data set available corresponds to patients with ADHD, and it has images of 1,583 subjects. There are very few images of healthy brains (93). The first challenge is thus the small sample sizes available to train the machine. Furthermore, verifying accuracy with the test scans suffers from the imbalance between the number of healthy samples compared to the rest. For example, if the proportion of healthy samples in the test set is only 10%, then a learning algorithm that always predicts illness will achieve 90% accuracy (it only gets the 10% healthy diagnostics wrong).

The second challenge that we face is classification. Mathematically, a classification is well defined only when its classes are disjoint and cover the whole sample space (i.e. it forms a partition of the sample space). However, medical conditions are not mutually exclusive. That is, a patient with ADHD may also suffer from MDD, a suicidal attempter may also suffer from OCD, and so on. The average number of conditions per patient for the Amen Clinic is actually 4.2. The goal is to eventually include over 20 conditions. In theory, multiple conditions pose no problem, one simply considers all possible combinations of the medical conditions, which necessarily form a partition. However, the latter is not feasible in this study, because many classes would have extremely few samples.

Our contributions are:

– **Validation of the SPECT technology for diagnosis.** In this paper, we train a binary deep convolutional network for each of the three conditions ADHD, MDD and OCD. For deep learning we used compressed images from the scans exclusively. We did not use all of the 3D information from the SPECT scans, only the 2D surface images, to reduce the dimension of the input. Surprisingly, this seems to be enough for the deep CNN to distinguish

the condition being tested from healthy. The results show high accuracy in prediction, see next section.

– **Social and economic impacts.** Our results show that SPECT scans can help early detection of mental illness, which can then improve treatment. There are great social and economic benefits in having tools that allow earlier treatment of mental illness. In particular, our method will allow family doctors and school counselors to refer potential patients to appropriate treatment early on.

– **Amber zone.** In this paper, one binary classifier is trained to distinguish SPECT scans that show each particular condition from the healthy. We express the result of the algorithm as a raw score. Usually deep learning networks apply a final activation function to that raw score and then a threshold comparison; we do not perform this step. Instead, we keep the set of raw scores found when training the algorithm as a frequency distribution; when implementing the algorithm this frequency distribution is used to express the outcome as an estimate of the probability that the scan shows the condition. In our view, this is a more useful way to proceed than just providing a yes/no output. In practice, a patient's scan would be fed into several algorithms, each detecting a particular condition; the estimated probabilities obtained would give the doctor information about each condition and an approximate ranking of the conditions from which the patient most likely suffers. For each algorithm, if the probability is above a certain level (say, 80%) then the result is labeled a clear positive (the "Red" zone), if it is lower than another level (say 20%) it is a clear negative (the "Green" zone), and if the probability is anywhere in between then the result is ambiguous (the "Amber" Zone). This effectively builds "soft" classifiers to provide better information to the medical practitioner, rather than provide only a binary output. In this particular study this is a way to avoid using a multi-class algorithm, which as we have explained above is not possible here due to the small number of samples.

Related Work. An important difference between this and other research efforts for image processing is that we do not propose a "hard" classifier. The closest work that we have found to our CNN model is reported in [2]. Those authors use deep CNN models to classify each pixel of the brain scans in five classes: non-tumor, necrosis, edema, non-enhancing tumor or enhancing tumor (five classes). They propose an architecture with two scales (called the TwoPath model), and another one with a cascading model. Another one is Byra *et al.* [1], who studied automating and improving liver steatosis assessment. Their study compares their CNN model with two well-known methods of assessing liver steatosis in ultrasound images: the hepatorenal sonographic index (HI) and the Gray-level co-occurrence matrix (GLCM).

This paper extends a project that was submitted at the 2017 CUNY IBM-Watson[1] competition, and which was awarded Third Prize (the team leader was Thomas Ward, Felisa Vázquez-Abad was the academic mentor).

---

[1] Watson is a system developed by International Business Machines Corporation.

2   Main Results: CNN Models for Single Conditions

Each model's architecture presented below was chosen through cross-validation. Then the best performing design in terms of accuracy was selected. During the testing phase, since we had limited negative samples, we couldn't employ cross-validation, instead we used hold-out sample sets for the ADHD, MDD, and OCD models. The negative (healthy) samples are constant throughout all hold-out sets, and the positive samples correspond to the condition of the particular model. These hold-out sample sets were selected randomly once, and there was no human inspection of the test samples that would introduce bias into the models. Therefore the training phase never uses the test samples, and provides an accurate measure of their validity.

2.1   CNN Model

We built a standard CNN model for binary image classification, described in Fig. 1, using the Python package Keras, with Tensorflow as backend, and running on a GeForce GTX 1080 Ti GPU. The training set had 1,266 images of ADHD patients and 74 healthy controls, while the test set had 314 and 19, respectively. Training achieved 100% accuracy, with 21,759 parameters (execution time 8'44 min).
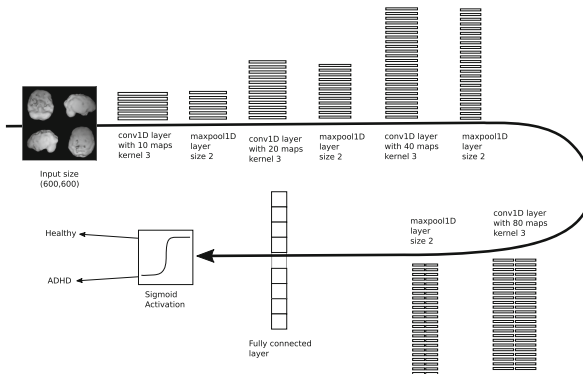
Fig. 1. Direct CNN model

The accuracy for the test sample is an important measure, as it provides information on how the algorithm performs when using it for classification: it provides the fraction of correct classifications. The available data is divided into a training set and a test set. The parameter values θ that are optimal for the training data are then fixed for the model (they are "the machine"). With these fixed parameters, each of the samples in the test set is used as input x to the algorithm, and a final raw score is obtained. This is a real number $v(x, \theta)$, positive or negative. The last step applies an activation function, in this case the sigmoid function:

$$\sigma(v(x,\theta)) \;=\; \frac{1}{1 + e^{-v(x,\theta)}}.$$

This is a number between 0 and 1, commonly called the "score". By default, a test image is classified as negative (resp. positive) if the score is less than (resp. larger than or equal to) 0.5.

To better understand the performance of the algorithm, given that we have significant imbalance between negatives (healthy) and positive (ill) images, we provide accuracies by class, rather than overall accuracy.

Table 1 provides accuracies by class, rather than overall accuracy. For instance, the row for the ADHD algorithm shows that of all the true ADHD test images, 99.4% were correctly classified as positives and of all the healthy test images, 78.9% were correctly classified as negative. Errors of this magnitude are to be expected when there are so few negative samples. Each model has 21,759 parameters.

Table 1. CNN models training times, sample sizes and accuracies.

Model name	Train time	Train size	Test size	Train acc. Pos.	Train acc. Neg.	Test acc. Pos.	Test acc. Neg.
ADHD	8'44 s	1,340	336	100%	100%	99.4%	78.9%
MDD	3'48 s	601	151	100%	100%	98.5%	84.2%
OCD	4'20 s	699	176	100%	100%	97.5%	100.0%

Figure 2 displays the results obtained on the training and testing sets as histograms. It is clear that the CNN separates the positives and negatives adequately.

These results validate the SPECT imaging technology for aid in diagnosis. Our algorithms use only the images, yet as shown on Fig. 2 the raw scores obtained show a clear separation between patients with the condition tested against healthy ones.

2.2 Cross-validation with Few Samples

Here we demonstrate how having very few samples (in our case for the healthy set) makes using validation of a model misleading. We show in each of our models (one for each condition we study) that validation is not straightforward. Instead this is one of the reasons why we propose the "amber zone" for soft classification, as detailed in the following section.

For each model, the total sample was divided into a training sample (80%) and a test sample (20%). In what follows, we describe the results for ten different random splits of the available data. The average accuracy for positive and
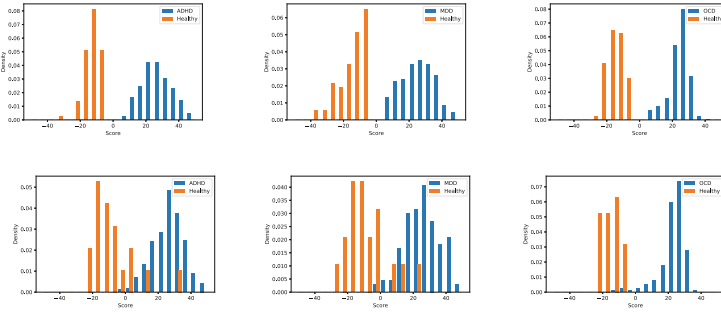
Fig. 2. Histograms of raw scores for ADHD, MDD, and OCD models. Upper row are training scores, lower row shows testing scores.

negative samples for the ADHD, MDD, and OCD models are (99.1%, 77.9%), (98.3%, 77.9%), and (99.1%, 92.1%) respectively.
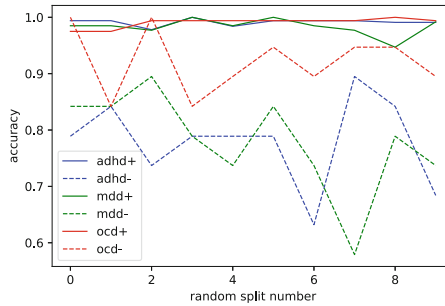
Fig. 3. Accuracy for ten randomly selected splits for models ADHD, MDD, and OCD.

Importantly, Fig. 3 shows the results per split. Clearly the mean accuracy does not provide enough information: the variance of the accuracy for negative samples (the images for the healthy) is too big for the ADHD and the MDD models and it lowers the statistical significance of the averages.

According to our medical team, there is a "diamond shape" pattern in OCD SPECT images. This pattern could explain why the OCD model has higher accuracy and smaller variance for the healthy images, than either the ADHD or MDD model. We speculate that a reason for the low accuracy values for healthy images when tested for ADHD or MDD could be that some SPECT images of healthy subjects appear to have patterns that resemble mild affliction of ADHD or MDD. Interestingly, it appears that the patterns in the healthy images are quite different from those of OCD.

2.3   The Amber Zone

"Training" a CNN means fitting the parameters, using the samples in the training set (the latter is distinct from the test set). The fitting is performed using the standard stochastic gradient or other optimization method. Denote by X the input vector (an image), and let $v(X, \theta)$ denote the result from the CNN prior to the final activation, that is, $v(X, \theta) \in \mathbb{R}$ (the raw score). Neural network classification techniques are inspired by logistic regression (a well-known statistical technique). Logistic regression **assumes** that the probability of an image being positive is the function $\sigma(v(X, \theta)) = 1/(1 + e^{-v(X,\theta)})$. Machine learning uses a sample of X's to fit the parameters needed to compute $v(X, \theta)$. Denote $\hat{\theta}$ the resulting parameter estimate.

Once training has been performed, the resulting algorithm (= machine) can be applied to produce a score for each new image that it receives. This is the algorithm that practitioners will be using when they classify new images, for which there is no diagnostic yet.

While the statistical model for the optimization assumes that the probability that an image x being positive (ADHD patient) is $\sigma(v(x, \hat{\theta}))$, this assumption may not agree with reality. The model for logistic regression is mathematically and computationally convenient in order to carry out the tuning of the algorithm's parameters in an efficient way. However, for a random individual i with SPECT scan $X(i)$, we believe it is very likely that Prob(i has ADHD) $\neq \sigma(v(X(i), \hat{\theta}))$.

To illustrate this, we show in Fig. 4 a typical cumulative distribution of the raw scores obtained, against the sigmoid function. The step function is the empirical distribution. For example, a raw score of -8 will give an estimate of 0.001 for the probability of having the condition when using the sigmoid, while the true fraction of individuals with a score of -8 or less that did have the condition is almost 20%. Calling $\sigma(v(X(i), \hat{\theta}))$ a "probability" may lead to misinformation. We believe that the analysis of the conditional histograms can provide better information than the scores alone.
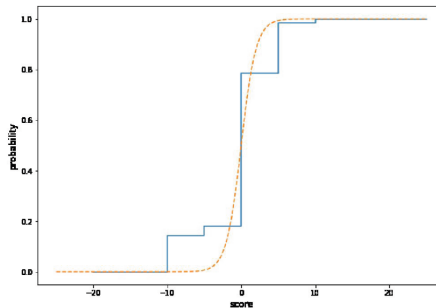
Fig. 4. Empirical probability of having a condition versus the sigmoid.

Although the model's parameter vector $\hat{\theta}$ provides the best fit for a sigmoid function, the true probability is not necessarily a sigmoid. This is the case in general for all parametric statistical estimation. Furthermore, a binary classifier must ultimately convert the raw scores into binary values, which is done using a threshold. Under the sigmoid model, the "default" threshold for the raw score $v(X, \hat{\theta})$ is chosen as 0.0. This implies that, if the model were true, the classifier is weighing equally the type I and type II errors. It is apparent from Fig. 2 that we may choose other thresholds to achieve a different balance between errors. In particular, when using these techniques for detection in order to help the medical practitioners for diagnosis, we think it is better to provide a different criterion. That is, instead of a *hard* classifier, we propose a *soft* classifier. Specifically, we suggest using boundaries $\underline{v}$, $\bar{v}$ such that for patient i,

$$\text{Prob}(i \text{ has } ADHD \,|\, v(X(i), \hat{\theta}) \leq \underline{v}) \leq \alpha_1$$
$$\text{Prob}(i \text{ has } ADHD \,|\, v(X(i), \hat{\theta}) \geq \bar{v}) \geq 1 - \alpha_2.$$

(Here α_1, α_2 are chosen levels of significance, for instance 0.05.) These two thresholds will be provided to the medical practitioner as the result of the analysis of the brain SPECT images: if the raw score $v(X, \hat{\theta}) < \underline{v}$ the algorithm will suggest "healthy", if $v(X, \hat{\theta}) > \bar{v}$ then it will suggest "ADHD". Within the "ambiguity zone" between $\underline{v}$ and $\overline{v}$, that we call the Amber Zone, the algorithm will provide an estimate of the true probability, that is: $p(v) = \text{Prob}(i \text{ has } ADHD \,|\, v(X(i), \hat{\theta}) = v)$.

In practice one would probably want to choose the empirical score distribution for the combined training and test sets, in order to obtain a better estimate of the probabilities. Smoothing techniques could also be used to turn the step function into a continuous one.

3   Conclusion and Future Work

We have shown that deep learning can help diagnose ADHD, MDD and OCD based on SPECT scans. Our results are effective in testing each of the conditions against healthy samples, especially in light of the relatively small sample sizes which were available. Typically, medical doctors arrive at diagnoses using several clinical data, while our algorithms used only the scan images of the brain. This validates the SPECT technology, as it shows that it must detect physiological characteristics of those conditions. Using our algorithms, physicians' diagnoses would potentially be more accurate, faster and more economical. Eventually this approach could lead to better predictions of possible mental illness conditions, which would in turn improve patients' outcomes.

Each of our algorithms is trained to detect a specific mental condition. In machine learning, it is customary to choose a threshold (say, 0.0) and to classify raw scores into positive/negative according to raw scores. We do use this procedure to train our algorithms but when implementing them, we express the

raw scores from training and testing as a frequency distribution (distinct for each condition tested). Then, we find where the raw score of the new scan falls within that empirical distribution. If the new raw score is smaller than a specific negative value we label the result a "clear negative" (Green Zone); if the score is larger than a specific positive value the score is labeled a "clear positive" (Red Zone); scores that fall in between are labeled "ambiguous" (Amber Zone). The practitioner is given the probability achieved by the score on the frequency distribution, meaning the higher that probability, the more likely a positive diagnostic becomes. The Green and Red Zones levels are determined by looking at Type I and II errors. We believe this is more useful in practice than a simple yes/no result, especially when a practitioner is faced with the outcomes of multiple algorithms, each detecting different conditions.

We did not apply multiple classification because we did not have enough samples. When dealing with multiple and possibly overlapping conditions, it is necessary to combine the results from the different algorithms that would be used in parallel. Using our first model presented in Sect. 2 for various conditions (here we used ADHD, MDD, or OCD), each brain scan image is assigned a raw score for each of the mental illnesses (in this case, three raw scores). We performed several experiments (not shown) mixing three conditions (healthy, ADHD and OCD), all with similar results. If all samples from various mental illnesses are mixed in the histogram, the resulting plot shows multimodal distributions, and it would be very difficult to draw meaningful conclusions. Instead, the histograms corresponding to each individual condition are plotted using different colors. It can be seen that each of the algorithms (say, one trained to detect ADHD) produces histograms that have a relatively smooth shape for each of the conditions of the subsamples (here, there would be four subsamples: healthy, ADHD only, OCD only, and patients with both ADHD and OCD). Typically the healthy inputs show empirical distributions that are skewed to the left and are well separated from the non-healthy raw scores. Those of other conditions present distinct shapes and they overlap. The overlaps are significant and further analysis of the raw scores (such as clustering techniques) may yet provide insight into the mental condition of patients. To illustrate this idea, Fig. 5 shows a scatter plot with the ADHD model scores on the x-axis and the OCD model scores on the y-axis. From the scatter plot in Fig. 5 we may already draw some ideas for future research. Looking only at the isolated conditions (all but the green dots), it is apparent that the space is divided into three clusters, corresponding to red (healthy), blue (ADHD) and orange (OCD) raw scores for the SPECT images. We performed a dozen experimental designs combining various conditions and methods for training, and all of them showed more or less the same pattern. We plan to further analyze high dimensional clustering when adding more conditions to the study. It may well turn out that clustering results from the binary machines will yield interesting new knowledge for psychiatrists.

Co-morbidity is common in mental illnesses, and it is not unusual to see mixed results for some patients, so the Amber Zone may yield insight about the patients condition. We plan to build a stochastic model to generate simulated
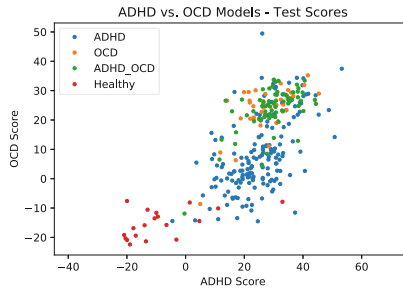
Fig. 5. Correlation between ADHD and OCD models scores.

functional images of the brain, based on its structure, to use in the pre-training phase, in order to improve the results, given extra conditions and small sample sizes.

The complexities behind diagnosis of mental illness are profound even for medical specialists. No two cases will ever be the same. This paper shows that artificial intelligence can help doctors ask meaningfully better questions in order to provide a more individualized plan for the patient. These tools will enhance the ability of medical doctors to provide an accurate diagnosis but are not meant to replace them with blind classification machines.

This study focuses on SPECT images because of the expertise of our team in this technology and the availability of data. However, our machine learning methodology can also be applied to other type of medical imaging. Future research may involve using different inputs for comparison.

References

1. Byra, M., Styczynski, G., Szmigielski, C., Kalinowski, P., Michaelowski, A., Paluszkiewicz, R., Ziarkiewicz-Wroblewska, B.: Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images (report). Int. J. Comput. Assist. Radiol. Surg. **13**(12), 1895–1903 (2018)
2. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P., Larochelle, H.: Brain tumor segmentation with deep neural networks. Med. Image Analysis **35**, 18–31 (2017). https://doi.org/10.1016/j.media.2016.05.004
3. Pelham, W., Foster, E., Robb, J.: The economic impact of attention-deficit/hyperactivity disorder in children and adolescents. J. Pediatr. Psychol. **32**(6), 711–727 (2007)

Vessel Segmentation and Stenosis Quantification from Coronary X-Ray Angiograms

Irina Andra Tache[1](✉) and Dimitrios Glotsos[2]

[1] University Politehnica of Bucharest, Bucharest, Romania
`irina.tache@acse.pub.ro`
[2] University of West Attica, Athens, Greece

Abstract. Angiography is the gold standard for diagnosis and interventional treatment of vascular pathologies, especially for stenosis, in many hospitals around the world. Still, the physicians complain of visual burdens due to its rather low spatial resolution and artefacts. Fluoroscopic angiography series from the dataset are obtained with standard clinical protocol. In the following, there is proposed an algorithm for vessel segmentation and edge detection. It is related to gradient operator applied on a pre-processed image with the Frangi's vesselness filtering for removing the equipment acquisition noises, followed by morphological operations for removing the spurs and adaptive threholding. The contour tracking along the vessel is done using Dijkstra's smallest path algorithm. The severity of the stenosis for a vessel segment can be assessed visually by a medical imagistic expert. The angiograph software can provide a graphic containing on the X axis the vessel segment length and on Y axis its corresponding cross-sectional areas. More objectively, the percentage of area stenosis can be computed.

Keywords: Vesselness filtering · Vessel pathology · Contour tracking

1 Introduction

The vascular diseases are reflected by problems in transporting the blood into the vessels, also called hemodynamic disorders. The medical imaging is an indispensable procedure in the diagnosis and treatment evaluation of these diseases.

Medical imaging techniques used in the examination of the cardiovascular diseases are: ultrasound investigations, computed tomography, magnetic resonance techniques, angiography, nuclear imaging, scintigraphy and positron emission tomography.

The X-ray angiography or conventional coronary angiography (CCA) is based on the projectional radiography and it is a clinical method for geometry assessment and filling properties of the blood vessels. They are particularly useful in detecting the aneurysms or any interruption of the blood flow through a main branch or the dilation or constriction of the large and medium vessels. Although, the small ones, which have low contrast, are difficult to be examined.

The edge detection is an important image processing step which can be included into segmentation. It is defined as the linear characteristic of the object which has at least one

neighbour outside the object. Actually, it implies the detection of the discontinuities in the grey levels.

In the scientific literature there are many methods to do vessels' segmentation. Considering that the X-ray angiograms are perturbed by noises with Poisson distributions, the algorithms suitable for on this kind of images are a few. In principal, they consider pre-processing the image before starting the image segmentation.

A review of vessels segmentation and tracking algorithms is presented in [1] and a review for coronary segmentation methods can be found in [2]. A new approach which uses classical image processing tools, such as multiscale adaptive Hessian-based enhancement method and the statistical region merging technique is detailed in [3].

The quantitative coronary arteriography is a tool used by the cardiologist to estimate the stenosis and includes the calibration, automatic contour detection and quantification of stenosis severity. In [4] they use the percent stenosis index which consists in comparing the normal regions and the ones with stenosis for in vivo data acquisition, with a catheter, using videodensity method. In [5] a system and associate methods to quantify stenosis by the percent of area stenosis which starts with the vessel centreline detection from the cardiac Computer Tomography Angiography (CTA) are presented. In [6] they use level sets for CTA segmentation and the so-called NASCET criterion for stenosis quantification which is based on diameters. Also in [7] they use the so-called cross sectional area measurement that takes into account the partial volume effect. In [8] an algorithm for stenosis identification for online 3D data set from CTA which is publicly available in [9].

Other interesting algorithms are related to blood flow estimation from CCA using densitometric measurements (the grey levels values). In [10] they assess the flow which is used for the assessment of the stenosis probability and in [11] arterial flow is computed for digital subtraction angiography.

Two commercial software packages used on the clinical X-ray angiographs are compered and analyzed in [12]. The focus is the performance of the algorithms regarding the accurate detection of stenosis on bifurcation vessels.

The aim of the study is to segment the vascular structures from angiography using simple and fast image processing algorithms and to quantify the coronary stenosis vessel on X-ray angiograms which in medical literature can be found as quantitative coronary angiography (QCA). This can be done by using an algorithm implemented in Matlab® with good performances for vessels contour detection and graphical representation of the cross-sectional areas along the vessel's length. This result can be used by a medical expert to assess the severity of the stenosis, either visually or quantitatively.

2 Methodology

2.1 Data Acquisition

The bones and the calcified plaques are strongly radio-opaque, while the low density elements are hardly visible on radiographs. Therefore, in order to make the opacification of the vessels, an iodine based substance is injected into the circulation.

On the first 3–4 acquired images there are no vessels displayed and as the contrast agent is mixing with the blood, they become to be visible on the image detector.

The image set is comprising data taken from one healthy person with non-significant vessel lesions and data from 10 patients with complete or partial stenosis on the main vessels. The data acquisition characteristics can be found in the Table 1.

Table 1. Protocol and image characteristics

Parameters	Cardiac
Injection site	Main left/right coronary
Projection types	Left and right caudal projection, right oblique anterior projection and left and right cranial projection
Spatial resolution	512 * 512
Contrast resolution	8 bits

2.2 Vessel Segmentation and Edge Detection

Segmentation algorithms for grayscale images are related to the discontinuity and similarity properties of the pixel values. For vascular network segmentation, there are filter enhancement algorithms, morphological operations, vessel tracking methods or classifier based methods. Some simple and quick algorithms for contour detection examine the magnitude of the spatial gradient, or the second order spatial derivative called Laplacian operation. There are also other algorithms such as, Sobel, Canny, Prewitt, etc. which use the spatial convolution masks for pixel by pixel passing.

The Sobel's method exploits the assumption that an edge is found in the maximal image gradient, whereas another method firstly filters the image with the Laplacian of the Gaussian filter and secondly searches the zero crossings. The Laplacian of an image J(x, y) is computed as in [13]:

$$\nabla^2 J(x, y) = \frac{\partial^2 J}{\partial x^2} + \frac{\partial^2 J}{\partial y^2} \tag{1}$$

Canny in [14] proposed a method for the image gradient computation in the neighbourhood pixels and its steps are presented in the followings. The Gaussian filtered image which is a slightly blurred version of the original image is computed. Two Sobel operators can approximate the image gradient and a searching for the local maximum admitted by the output magnitudes in the gradient direction is performed. For limiting the multiple maximum detections due to the existing noises, a hysteresis thresholding is applied for detecting only the strong edges and their low connected branches using the direction information. This algorithm is more robust to noises and can detect the low contrast edges.

An improvement of this filter for optimal edge detector is made by Deriche in [15]. Still, it has not good reviews for noise removal and localization, therefore Kumar et al. in [16] proposed a method for edge detection which implies a median filtering and histogram equalization instead of Canny algorithm.

The median filter is attended to eliminate the intensity spikes by not affecting the edges and the uniformization of the values is made by a histogram equalization which stretches and compresses the image for extracting the edges of the blood vessels. For the vesselness filtered angiograms, the Canny method performed well. An algorithm for the extraction of a vessel's segment borders is proposed:

1. The image is enhanced with Frangi's vesselness filter,
2. On the maximum magnitude image resulted from the step 1 it's applied a region growing algorithm for segmentation,
3. Some morphological operations are applied for removing the spurs and unwanted pixels,
4. On the binary image obtained at the previous step, the user draws two lines for delimitating the segment,
5. For contour detection, the Canny edge detection algorithm is applied,
6. A series of perpendicular lines on the edges are drawn in the vicinity of delimitating lines in order to find two valid intersections with the vessel borders. The closest line which respects this requirement is selected and its points are marked on the binary image,
7. The image from the step 6 is cropped for focussing the vessel segment,
8. The Dijkstra's algorithm is applied for finding the optimal path on one vessel border,
9. The other border is found at the intersection of the perpendicular lines starting from the first detected edge with the binary image obtained at the step 7,
10. Finally, the radius or the cross sectional areas of the segment can be computed.

A visual representation of the workflow is presented in the Fig. 1.

2.3 Quantitative Coronary Arteriography

The quantitative coronary arteriography with the acronym QCA is a standard used by the cardiology physician to assess the coronary artery dimensions [17]. The QCA can help the clinicians in the selection of the optimal balloon and stent. It's an analysis for single or bifurcation vessels.

After the algorithm implementation for automatic contour detection with the vesselness filtering approach presented in the previous section, the length of the vessel segment suspected to include a stenosis is detected with its corresponding cross-sectional areas. With these quantities a graphical representation of the 2D QCA is performed as in the Sect. 3.2 and the severity of the stenosis is judged by an abrupt change of areas. For a more objective perspective, the parameter which characterize the stenosis severity – the percentage of area stenosis (PAS) is introduced using the below equation:

$$PAS(\%) = \frac{A_{max} - A_{min}}{A_{max}} \cdot 100 \qquad (4)$$

Where A_{max} is the maximum cross sectional area and A_{min} is the minimal cross sectional area.
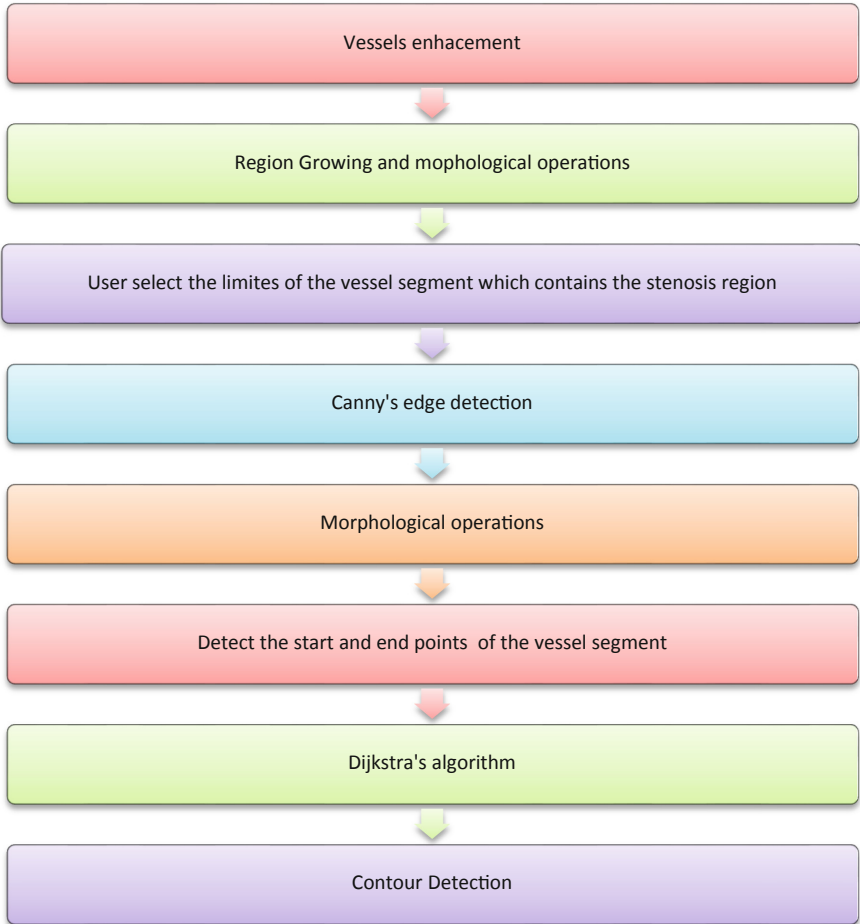
Fig. 1. The steps for automatically extraction of the vessels' contours

3   Results

3.1   Contour Detection

The proposed method is using the Frangi vesselness filtering in order to detect the pixels which are the best candidates to represent a vessel. Based on the maximum magnitude image, a graph is constructed considering the nodes as the image coordinates where the probability of vessel occurrence is high. After defining the start and end points, the shortest paths are then found by the Dijkstra's algorithm (Figs. 2 and 3).
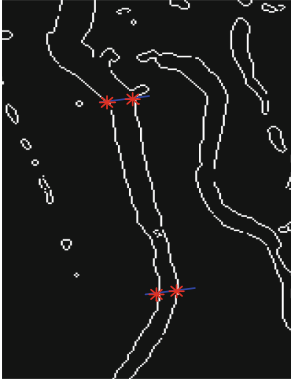
Fig. 2. Vessel's segment identification from ROIs
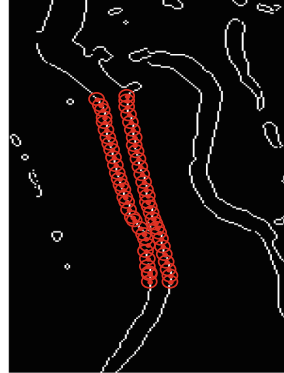
Fig. 3. Vessel's border tracking

The utility of this automatic method is specially dedicated to abnormal tubular structures which will further permit the computation of the stenosis geometrical angiographic quantification. Still, for the severe cases it can fail, due to the interruption in the vessel edge continuity.
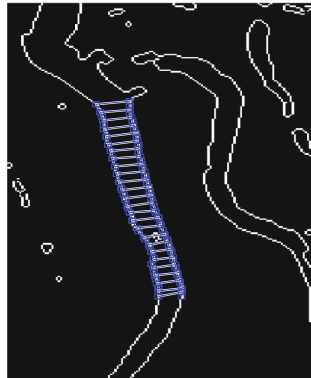
Fig. 4. Vessel's radius

3.2 Stenosis Quantification

On the results obtained in the Fig. 4 the further analysis is focussed on the small vessel narrowing. After the contour detection, a graphical representation of the cross-sectional areas along its length from a projectional view is presented in the Fig. 5(a).

In Fig. 5 it can be observed the differences between a non-significant and a severe stenosis. For validating the method, the resulted QCA graphical representation is compared with the one from the medical equipment which was previously approved by a
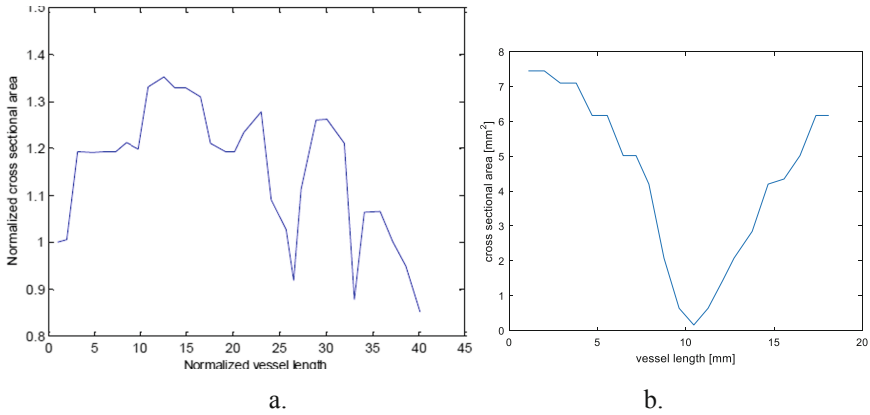
Fig. 5. The QCA graphical representation (a) for an insignificant stenosis (b) for a severe stenosis

physician. Also because the physician diagnosed the vessel obstruction from the Fig. 5(b) as complete stenosis, hence the percentage of area stenosis will expect to be almost 100% after the physician visual inspection. As can be seen from the graphic A_{max} is 7.5 mm^2 and A_{min} is 0.2 mm^2, which means that the PAS parameter is 97.33%. The results are consistent from these two reports.

4   Conclusions

The key step in improving the medical images for a quick diagnostic is the segmentation process which can be regarded as pixels' classification into two groups: the object and the background. This step is necessary for centreline and contour detection of the vessels. This target is not properly achieved if the noise is not correctly filtered in the original image.

The method works well in some cases and the vessel sizes can be computed but the performance degrades when there is background clutter in the angiogram, such as spines, catheters, and guidewires. Furthermore, vessel crossings could be incorrectly identified as vessel bifurcations.

This image processing algorithm is further used for geometrical quantification of the vessels (radius, cross-section areas and vessel's length) and for quantification the stenosis severity using the QCA tool.

Further work may include the refinement of the vessel structure extracted from a 2D angiogram by using prior shape information from a 3D coronary vessel model of the same patient. For the edge detection the gradient approach is used after the Frangi vesselness filtering followed by some morphological operations. The method is automatic and it's suitable for a wide range of vessels.

References

1. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. ACM Comput. Surv. **36**(2), 81–121 (2004)
2. Dehkordi, M.T., Sadri, S., Doosthoseini, A.: A review of coronary vessel segmentation algorithms. J. Med. Signals Sens. **1**(1), 49–54 (2011)
3. Wan, T., Shang, X., Yang, W., Chen, J., Li, D., Qin, Z.: Automated coronary artery tree segmentation in X-ray angiography using improved Hessian based enhancement and statistical region merging. Comput. Methods Programs Biomed. **157**, 179–190 (2018)
4. Wiesel, J., Grunwald, A.M., Tobiasz, C., Robin, B., Bodenheimer, M.M.: Quantitation of absolute area of a coronary arterial stenosis: experimental validation with a preparation in vivo. Circulation **74**(5), 1099–1106 (1986)
5. Zhou, J., et al.: Quantification of coronary artery Stenosis by Area Stenosis from cardiac CT angiography. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, pp. 695–698 (2015)
6. Scherl, H., et al.: Semi-automatic level-set based segmentation and stenosis quantification of the internal carotid artery in 3D CTA data sets. Med. Image Anal. **11**(1), 21–34 (2006)
7. Molloi, S., Johnson, T., Ding, H., Lipinski, J.: Accurate quantification of vessel cross-sectional area using CT angiography: a simulation study. Int. J. Cardiovasc. Imaging **33**(3), 411–419 (2016)
8. Kirişli, H.A., et al.: Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. Med. Image Anal. **17**(8), 859–876 (2013)
9. http://coronary.bigr.nl/stenoses/
10. Sarry, L., et al.: Assessment of stenosis severity using a novel method to estimate spatial and temporal variations of blood flow velocity in biplane coronarography. Phys. Med. Biol. **42**, 1549 (1997)
11. Bonnefous, O., Pereira, V.M., Ouared, R., Brina, O., Aerts, H., Hermans, R., van Nijnatten, F., Stawiaski, J., Ruijters, D.: Quantification of arterial flow using digital subtraction angiography. Med. Phys. **39**(10), 6264–6275 (2012)
12. Ishibashi, Y., Grundeken, M.J., Nakatami, S.: In vitro validation and comparison of different software packages or algorithms for coronary bifurcation analysis using calibrated phantoms: implications for clinical practice and research of bifurcation stenting. Catheter. Cardiovasc. Interv. **85**, 554–563 (2014)
13. Zhou, H., Wu, J., Zhang, J.: Digital Image Processing: Part I. Ventus Publishing ApS (2010)
14. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
15. Deriche, R.: Using Canny's criteria to derive a recursively implemented optimal edge detector. Int. J. Comput. Vis. **1**, 167–187 (1987)
16. Kumar, S.S., Amutha, R.: Edge detection of angiogram images using the classical image processing techniques. In: IEEE-International Conference on Advances In Engineering, Science and Management (ICAESM 2012), pp. 55–60 (2012)
17. Reiber, J.H.C.: Introduction to QCA, IVUS and OCT in interventional cardiology. Int. J. Cardiovasc. Imaging **27**(2), 153–154 (2011)

Improved Brain Tumor Segmentation and Diagnosis Using an SVM-Based Classifier

Krishna Ganesh$^{(\boxtimes)}$ and R. Swarnalatha

Birla Institute of Technology and Sciences Pilani, Dubai Campus, Dubai, UAE
{f20160255,swarnalatha}@dubai.bits-pilani.ac.in

Abstract. The Increased focus on the aspects in the Computer Science Field like Machine Learning and Neural Networks have brought about a change in diagnosis methods of various diseases. Machine Learning has advanced so much and provides benefits like enhanced efficiency and fast decision making which prove to be crucial in the field of medical diagnosis. One of the highly deemed research fields which happen to be greatly affected by this advance is the field of cancer diagnosis. Cancer was responsible for 18 Million cases in 2018 and was responsible for over 9 million deaths. Over the years, the cancer rate has been decreasing marginally, but the last 3 years have seen an increase by 2%. Accurate diagnostic methods need to be considered to when dealing with cancer because incorrect diagnosis can be even fatal to the patient, hence a tool is required, which can trump even the human brain at diagnosing cancer. Machine learning can prove to be this tool due to the fact that it learns along the way once you feed an input to it, so as time goes on it gets smarter and smarter eventually always exceeding the accuracy percentage of its results and thereby proving to be a useful tool for medical diagnosis. In this paper, we propose a support vector machine (SVM) classifier for brain tumor segmentation and diagnosis. 50 MRI images have been used to test this and an accuracy of 100% has been achieved in diagnosis of the tumor as benign or malignant. Kernel accuracy of 90% in each case has also been achieved.

Keywords: MRI · Image segmentation · Feature extraction · Machine learning · SVM · Kernels · Diagnosis

1 Introduction

Neoplasms or Brain Tumors are abnormal cells that form over time in a person's brain [1]. Tumors broadly can be classified into two types - the cancerous ones (malignant tumors) and the non-cancerous ones (benign). Cancerous tumors can further be sub-categorized into primary tumors which have their origin in the brain itself, or secondary Metastasized tumors which originate from other parts of the human body. The presence of a tumor can cause several problems like oedema, displacement of brain tissue, low blood flow to different parts of the brain and in most cases, slow degeneration of the brain tissue which control important functions of the human body [2].

The general symptoms of a brain tumor include:-

- Headaches with increasing frequency and severity
- Vision Impairment, speech impairment, confusion, hearing problems
- Disturbed body balance, Behavioral changes, loss of sensation in body parts
- Seizures

Primary tumors occur usually in the meninges, pineal gland, cranial nerves or in the neighborhood of these areas. The main cause of these tumors is mutation of DNA in the brain cells, which causes uncontrolled growth and eventually builds up to form a tumor. Some of the different primary tumors include Acoustic neuromas, Astrocytoma, Oligodendroglioma, Ependymoma etc. [3].

Magnetic Resonance Imaging (MRI) is the most widely used technique to detect a brain tumor. The advantage of the MRI scan is that it offers a much more well-defined contrast between soft tissues in the body as compared to a Computerized Tomography (CT) Scan. It uses a magnetic field to line up the H2 atoms of the H20 molecules in the body. The main task in diagnosis lies in segmenting tumors in each MRI Image [4]. This poses to be an uphill and time-consuming task due to factors like size variations, appearance, overlapping intensities with normal brain tissues and distortion of various structures in the brain, thereby giving phase differences in healthy tissue orientation. Segmenting tumors in MRI images is very important as it gives us the orientation and highlights the tissues which are susceptible to becoming a tumor-based tissue for later.

Machine learning is employed in techniques like computer aided diagnosis (CAD), content-based image retrieval (CBIR), Image quality assessment and brain mapping which helps in actively predicting future data from the input [5].

Mathematic Laboratory (MATLAB) is a software which efficiently integrates both, Image segmentation and machine learning due to the availability of the various toolboxes, which allows us to flexibly build a system of our liking.

2   Background

SVM is a regression prediction and classification tool which uses machine learning as a base to maximize predictive accuracy and takes care not to overfit data on a hyperplane. There are many advantages of SVM as compared to algorithms like Artificial Neural Networks (ANN) and k-Nearest Neighbors (kNN), namely [6, 7]

- Back Propagation Neural Networks (BPNN) has the drawback of calculating multiple local minimal when we are using function approximation which will affect convergence. SVM has a global solution which saves us the optimization time for the hyperparameters
- ANNs work with Empirical Risk Minimization (ERM) and SVMs work with the Structural Risk Management (SRM) principle. ERM focuses on estimating the efficiency of an algorithm, which makes it susceptible to overfitting and overtraining which results in poor performance. SRM works around these issues by balancing the computational complexity and ensures appropriate fitting of the data on the hyperplane.

- There is a possibility of extending the feature space in a kNN using kernels. Kernels are based on a set of reduction rules which remove constraints of feature space. Although kernels can overcome certain problems, kNNs have low performance with high feature space. However, SVMs work very well in high feature spaces.
- SVMs have simplified computational requirements for classifying data as compared to kNN, thereby decreasing training and deployment time.
- ANNs are 'Data-Hungry' and provide limited accuracy working with small datasets. But SVMs outmatch NNs in utilizing datasets by a large margin.
- Convolutional Neural Networks (CNN) work by convolving a set of signals which are in immediate succession to each other. When we shuffle these features however the output of the CNN changes by a large margin, which implies that order is important while using CNN. SVMs, however are unaffected by any such change.
- Computational Complexity of SVM is much lower as compared to ANNs or kNNs.

Selvanayaki K and P Kalugaselvam in their paper use meta-heuristic methods like Ant Colony Optimization (ACO), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for the segmentation of MRI Images, there's two stages in which this is carried out:-

- Stage 1 - Pre-processing and Enhancement of Images
- Stage 2 - Identifying and Segmentation of the tumor concentrated region in the image, using ACO, GA, PSO and performance measurement is studied.

Limitations: There was no provision of machine learning mentioned in this research and this helps us to automate the process and achieve higher accuracies. The algorithms mentioned here are old and outdated and would in comparison, perform very poorly when pitted against an SVM classifier [8].

The research carried out by Priya patil et al. outlines all the different image segmentation algorithms which can be employed for brain tumor detection. It talks about the various techniques which can be employed when detecting brain tumors and tells us the list of processes to be done. Techniques like Threshold segmentation, Watershed segmentation and K-means clustering were mentioned in the paper.

Limitations: Does not mention the disadvantages in each approach, and the comparison between each method does not hold true in all cases [9].

Arun Kumar Proposes a basic GUI for implementing SVM, K-Means Clustering and Principal Component Analysis (PCA) all in a minimalistic GUI. He has extensively studied on the work done in the field right from 1970, up to 2014, depicting all the changes that have been incorporated and how the progression to implementation has been done via different softwares and algorithms.

Limitations: The research only focuses on extracting a few first order statistical features from the image and ignoring the other ones. First order statistical features include:- mean, standard deviation, smoothness and skewness. Designing a highly sensitive system requires more than 4 features for pinpoint accuracy. It also doesn't mention Kernelization as the upscale feature of SVM, nor diagnoses the tumor as benign or malignant.

This is some of the background work done in the past when it comes to using SVM as a classifier for brain tumor analysis. This research proposes several improvements

and additions in comparison to the older existent methods and overcomes the drawbacks of SVM. The improvements made in the project are as follows:-

- A Pure GUI based Approach – There has been a lot of coding-based approaches in the past and the main drawbacks when it comes to coding is that (a) There is minimal reusability of code, depending on the level of expertise of the programmer, and (b) Difficult to understand when it comes to trying to decipher the code for a layman. Hence, a Graphical User Interface (GUI) based approach has been adopted on MATLAB for a much more clear, concise and user-friendly configuration.
- Better segmented image samples - Segmenting an image removes all the noise that comes with the image and helps us take better pre-op and post-op decisions based on the tumor orientation in the enhanced image. The techniques that have been used are:-

  – Median Filtering
  – Image Inversion
  – Tumor Bordering
  – Canny Edge Detection
  – Watershed Algorithm [10, 11]

- Efficient Utilization of the SVM large training set drawback - The main debatable point when it comes to ANN vs SVM is an ANN's ability to handle large datasets efficiently and reduce training time, which isn't seen in SVM. It can hinder the performance of the latter if a large training set is used. However, it has been proved that If ANN's perform great when it comes to a large training sets, SVMs perform better when it comes to small training sets. A fact which is commonly overlooked is that an SVM with a small dataset is just as powerful, if not more than an ANN with a large dataset. The SVM algorithm also gives us 85% accuracy when it comes to Kernelization (kernelization isn't advisable with an ANN and will hinder it in most cases). Therefore, we are using the main disadvantage of the SVM to reduce computation time and not compromise on the accuracy percentage
- Kernelization - Probably the greatest advantage of the SVM as compared to both the ANN and kNN. Kernels are shortcuts which help us do certain calculations much faster which would normally increase in computational complexity when we are using a higher dimensional space initially

$$K(x, y) \ = \ <f(x), f(y)>$$

This simple equation can be interpreted as 'K' being the Kernel function and 'x' and 'y' are dimensional inputs of a space 'A'. 'f' is a function which maps the variables 'x' and 'y' to a much larger feature space 'B'. What this tells us is that kernelization can help us map to an infinite feature space, so the calculations can be carried out, but with lesser expenditure of computational power. ANN has 2 disadvantages when it comes to the kernel method, namely-

– The kernel needs to work in line with the activation function of the NN. Any knowledgeable NN programmer will understand what each input feature stands for, but he will have a hard time interpreting the activation of certain units. Therefore, the main objective of designing problem specific kernels is lost.
– Back propagation of kernels is not possible due to the gradients of the expressions used.

The kNN also has problems when it comes to kernelization, which are:-

– There is not much difference when it comes to a Kernelized kNN and a kNN with normal data unless the feature maps can bring the points from the same class closer to each other.
– When we use kernels in kNN, we have the uncertainty that mapping the data to a higher dimensional space would give us a better accuracy.

Hence, when we kernelize SVM, we get much better results as compared to an ANN or kNN while using minimum data [12, 13]

- Increasing Kernel Accuracy - One of the advantages of using a kernel based SVM is that we can retrain the model until optimal accuracy has been obtained. This is one of the added advantages of the Kernelized SVM. When it comes to linear and non-linear data, SVM works in such a manner that it doesn't create a local minimum but rather a global one, thereby removing the constraint of overfitting.
- Low Computational Complexity and Less Deployment time - When non-linear classifiers are taken into consideration, they are usually defined as classifiers that transform data to a new representational space to apply better techniques. When this mapping is done, there is a possibility of increasing computational complexity. But in the case of SVM, the ability of making function specific kernels helps us to reduce the computational complexity of the function. The flexibility of using low number of feature vectors for calculations also add to the reduced complexity. Utilizing the small dataset for high accuracies is also a contributing factor in reducing the deployment time of the SVM. The training time as compared to ANNs and kNNs is drastically reduced as well.

3 Methodology

This project can be divided into 3 main sub-sections, namely:-

1. GUI initialization and Image Loading
2. Image segmentation, statistical feature extraction and tumor diagnosis
3. Kernel Accuracy calculation

- GUI Initialization and Image Loading -

- – In the opening function of the GUI, we define a matrix of ones and display it onto the axis, which is used to limit the size of the image as per the requirements. Since we have 5 Segmented image samples, we're keeping the matrix dimensions small (300×300 or 400×400)
- – There have been 2 image loading techniques that have been used, (a) - a simple push button and (b) - A Listbox. A listbox is an interactive addition, as it doesn't require you to search a folder for an image and keep reloading it. Only prerequisite is that the image name has to be added along with the extension (jpeg, png, tif) on the listbox properties editor.
- – The push button has been programmed to look for an image anywhere stored. It has also been programmed to prompt up a dialog box with an error message if no image has been selected.

- Image segmentation, Statistical feature extraction and Tumor diagnosis –

  - – Segmentation, feature extraction and tumor diagnosis have been programmed into a single push-button, which means you can get tumor diagnosis in 2 clicks
  - – Here, the Image is first returned to a variable from the axes where it was first displayed before segmentation, and then it is processed to give us all the segmented samples.
  - – From this image, we can extract 1st and 2nd order statistical features. We extract a total of 13 features, some of which are

    - Standard Deviation
    - Smoothness
    - Skewness
    - Kurtosis [9]

Once these features are extracted, they are compared with the features of the training set and then the diagnosis is displayed as whether it's a benign or malignant tumor [17].

- Kernel Accuracy Calculation –

  - Kernels have been introduced in this SVM as they can be used to perform mapping of certain calculations in high dimensional space.
  - We are evaluating 4 different Kernel Accuracies in this work.

    - Radial Basis Function (RBF) Kernel–RBF or Gaussian RBF is a function whose value is relative to the origin or any given point in space. It acts as a low bandpass filter as a tool which is used to select smooth solutions [3]
    - Linear Kernel – Is used for separating the data points of a given set by using a straight line. Linear Kernel performs very well at classifying when there are only 2 classes present. Hence the mapping of data points to a higher dimension isn't really required
    - Polynomial Kernel – The polynomial kernel is one of the most common kernels used with SVM, which denotes that vectors in a feature space are mapped over

polynomials of the original variables. It works well with non-linearly separable data

- Quadratic Kernel – Defines the decision boundary of the dataset points on the input space. Highly useful in linear separation if we use a quadratic function to map data into 2 dimensions [3]

There's a push button programmed for each kernel calculation. We are using K-holdout cross validation as compared to other validation techniques due to several reasons:-

– Ensures fully independent data. Has very low computational costs. On the contrary, K-Fold Cross Validation has very high computational costs and expensive training time as well.
– Splits the training data and testing data as compared to Resampling Validation.
– K Fold validation isn't applicable in all cases, especially when we need to train a predictive model for future, hence even in this case, Hold-out is better.

4   Results and Discussions

The trained model has been tested with 40 test data samples, 17 benign tumors and 23 Malignant tumors. We have made several observations when it comes to the overall classifying process.

Figure 1 depicts the various buttons, axes, features and values which are used in the GUI. This can be easily modified to user need and Fig. 2 shows the segmentation techniques employed in this research.
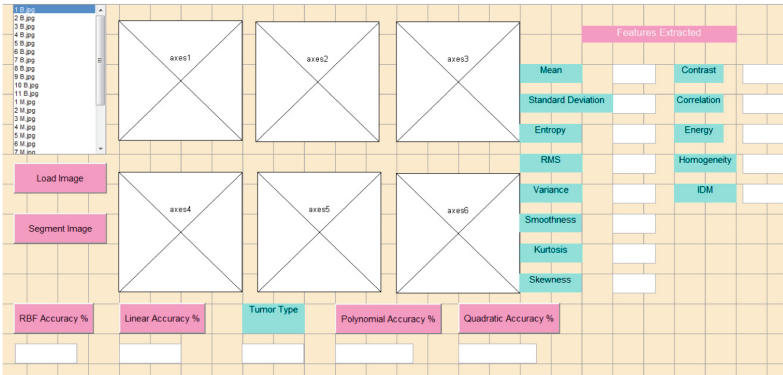
Fig. 1. GUI layout

- GUI programming in MATLAB ensures for a more clean and crisp method with easy accountability and ease of editing. With this, many techniques can be added or removed to ensure fluidity of deployment. But most of all, it ensures easy use, which is crucial in hospitals to ensure fast and accurate diagnosis.
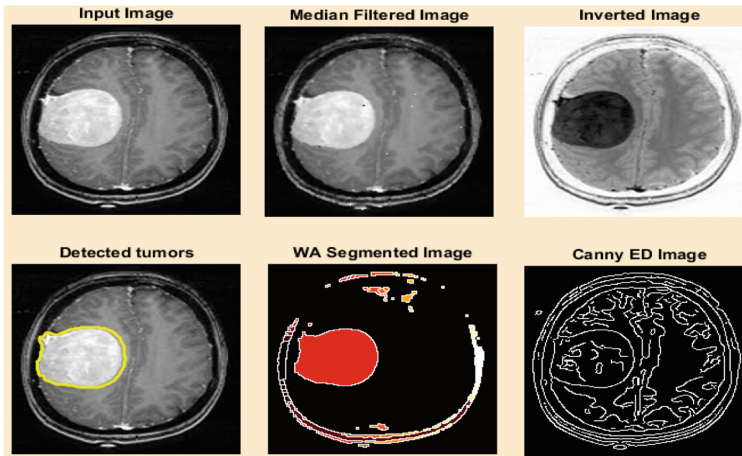
Fig. 2. Segmentation techniques used

- Precautions need to be taken while loading an image. While using a pushbutton, using the '.*' extension gives us the flexibility to accept any image format, while using a list box, care must be taken to specify the file format while using the property editor of the Listbox. Another precaution that must be taken while editing the dialog box is that ensuring the file is in the folder of the Matlab program, or an error which tells us that the file doesn't exist will pop up.
- 5 Segmentation techniques have been used to ensure that proper classification and it becomes easier to make pre-op and post op decisions. Segmentation makes the tumor more visible and allows us to view the exact region of the tumor much more easily and give us a >90% approximation of all the tumor tissue present. These techniques have advantages over the conventional ones like:-

  – Otsu Segmentation – Gray threshold level might segment entire image in certain cases
  – Laplace and Sobel Filters – They produce images of very low contrast, hence not ideal for pre-op decision making.
  – Contrast-Limited Adaptive Histogram Stretching (CLAHE) – Contrast is better, but overall dim of the image is increased. Relatively same as compared to input image [14, 15]
  – Wiener Filter – Blurs the image slightly and removes the some of the edges present [16 – 18]
  – Log and Power law Transformations – Log Transformation might make the distribution more skewed and incorrect skew values are obtained. Low Gamma value might start decreasing the contrast of the whole image.

- Gray-Level Co-occurrence matrix has been used for feature extraction as compared to individual formula-based feature extraction for 3 main reasons [19]:-

– Reduced computational complexity
– Decreased deployment time
– Increased accuracy for image classification as compared to ANN or kNN when using the GLCM with PCA [20]

- The number of iterations were fixed by trial and error, we started with 20 iterations per cycle for kernel accuracies. Observed that there was fast computation time, but there were poor accuracies in all the kernels. RBF accuracy was the lowest at 50%, and Linear accuracy was 70%. We round off the values to the highest 0 to speed up computation. As we continued increasing the number of iterations, we found out that 85–100 iterations provided optimum processing time and high accuracies
- The RBF kernel has limited programmability in this case. In theory, a fully programmed RBF Kernel will always have a higher accuracy as compared to the Linear Kernel or the Polynomial Kernel, but in this case, only 5–8 test images showed 90% accuracy. All the other kernels showed 90% accuracy in almost all cases with retraining which shows that a limited RBF kernel programmability in MATLAB is one of the drawbacks in Kernelization.
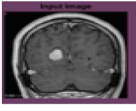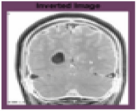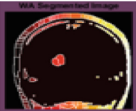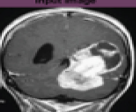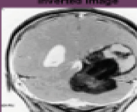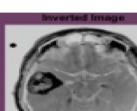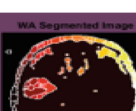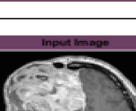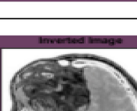
Test No.	Input Image Sample	Inverted Image	WA Segmented Image	Diagnosis	RBF Accuracy %(Max)	Linear Accuracy %(Max)	Quadratic Accuracy %(Max)	Polynomial Accuracy %(Max)
1				Benign	70%	90%	80%	90%
2				Malignant	80%	90%	90%	80%
3				Benign	70%	90%	70%	90%
4				Malignant	90%	90%	90%	90%

Fig. 3. Testing of the system and accuracies obtained in different cases of benign and malignant tumors

5   Conclusion and Future Scope

The research done in this paper proposes an improved method to classify brain tumors with increased accuracy. The modifications in this GUI based design has greatly improved the results obtained by the SVM-based classifier. To summarize the results, there has been 100% accuracy obtained in classification of tumors and >85% accuracy in kernels. We have used the Gray Level Co-occurrence Matrix (GLCM), to extract 13 statistical features which are being used to classify the tumor. This makes the model very sensitive to classification and is not susceptible to error. Kernels and GLCM have greatly reduced the computational complexity and deployment time for the system itself. The drawbacks faced by most systems is high training and deployment times, which have been overcome in this method. Most importantly, we have used a small dataset to obtain near perfect accuracy.

There are however, a few limitations when it comes to an SVM classifier, which are highlighted as follows:-

- Although SVM has the advantage of not being susceptible to overfitting, the kernels however don't have such a luxury. Although retraining is simple and can be used for higher accuracies, there are cases when overtraining a system can lead to no convergence being achieved and overfitting also takes place. Retraining it again may remove this error, but once this happens, the system tends to start overfitting a greater number of times henceforth.
- Limited RBF Kernel definition. Ideally, in any given scenario, a fully structured, defined RBF kernel outperforms all the other kernels by a large margin. The Linear Kernel is usually referred to as a 'degenerate' RBF Kernel which will never match up to its actual accuracy. But in this case, due to limited RBF Kernel programmability in Matlab, Linear accuracy exceeds RBF kernel accuracy in most cases. Therefore, we have included more than one Kernel to compensate for the partially defined RBF kernel.

Every system has its disadvantages, whether it comes to kNN, BPNN or ANN. Hence, we need to have a system with minimum disadvantages, but maximum output accuracy. We propose an Adaptive Neuro-Fuzzy Inference System (ANFIS) to overcome all the advantages of all the Neural Networks mentioned above. ANFIS has been proven to work around the disadvantages of ANNs and Fuzzy-Logic based inference systems.

References

1. Patil, P., et al.: A review paper on brain tumor segmentation and detection. Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng. **5**, 12–15 (2017)
2. Zhang, Y., Wu, L.: An MR brain images classifier via principal component analysis and kernel support vector machine. Progress Electromagn. Res. **130**, 369–388 (2012)
3. Pahwa, S., Sinwar, D.: Comparison of various kernels of support vector machine. Int. J. Res. Appl. Sci. Eng. Technol. **3**, 532–536 (2015)
4. Wernick, N., Yongyi, Y.: Machine learning in medical imaging. IEEE Signal Process. Mag. **27**, 26–38 (2010)

5. Damodharan, S., Raghavan, D.: Combining tissue segmentation and neural network for brain tumor detection. Int. Arab J. Inf. Technol. **12**(15), 42–52 (2015)
6. Vani, N., et al.: Brain tumor classification using support vector machine. Int. Res. J. Eng. Technol. **4**, 1724–1729 (2017)
7. Fabelo, H., et al.: SVM optimization for brain tumor identification using infrared spectroscopic samples. Sensors **18**, 4487 (2018)
8. Richika, K.A.: A novel approach for brain tumor detection using support vector machine, K-Means and PCA algorithm. Int. J. Comput. Sci. Mob. Comput. **4**, 457–474 (2015)
9. Kumar, V., Gupta, P.: Importance of statistical measures in digital image processing. Int. J. Emerg. Technol. Adv. Eng. **2**, 56–62 (2012)
10. Khalifa, A., Frigui, H.: Multiple instance fuzzy inference neural networks. IEEE Explore (5), 1–28 (2016)
11. Maiti, I., Chakraborty, M.: A new method for brain tumor segmentation based on watershed and edge detection algorithms in HSV color model. In: Computing and Communication Systems (NCCCS) National Conference, pp. 1–5 (2012)
12. Sharma, B., et al.: Review paper on brain tumor detection using pattern recognition techniques. Int. J. Recent Res. Asp. **3**, 151–156 (2016)
13. Bray, F., et al.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**, 394–424 (2018)
14. Selvanayaki, K., Kalugasalam, P.: Intelligent brain tumor tissue segmentation from Magnetic Resonance Image (MRI) using meta heuristic algorithm. J. Global Res. Comput. Sci. **4**, 13–20 (2013)
15. Verma, K., et al.: Image processing techniques for the enhancement of brain tumor patterns. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **2**, 1611–1615 (2013)
16. Selkar, R., et al.: Review on Detection and segmentation of brain tumor using watershed and thresholding algorithm. IORD J. Sci. Technol. **1**, 11–14 (2014)
17. Shidnal, S.: Texture feature extraction of crop field images using GLCM approach. IJSEAT **2**, 1006–1011 (2014)
18. Motoyoshi, I., et al.: Image statistics and the perception of surface qualities. Nature **447**, 206–209 (2007)
19. Kumar, S., et al.: Performance comparison of median and wiener filter in image de-noising. Int. J. Comput. Appl. **12**(4), 27–31 (2010)
20. Lixin, G., et al.: Design of SVM based on radial basis function Neural Networks pre-partition. In: ICSP 2014 Proceedings, pp. 1480–1483. IEEE (2014)

3D-Reconstruction and Semantic Segmentation of Cystoscopic Images

M. Negassi[1,2(✉)], U. Parupalli[1,2], R. Suarez-Ibarrola[3], A. Schmitt[1,2], S. Hein[3], A. Miernik[3], and A. Reiterer[1,2]

[1] Department of Sustainable Systems Engineering INATECH, University of Freiburg, Freiburg, Germany
misgana.negassi@ipm.fraunhofer.de

[2] Fraunhofer Institute for Physical Measurement Techniques IPM, Emmy-Noether-Strasse 2, 79108 Freiburg, Germany

[3] Department of Urology, Faculty of Medicine, University of Freiburg – Medical Centre, Hugstetter Str 55, 79106 Freiburg, Germany

Abstract. Bladder cancer (BCa) is the fourth most common cancer and the eighth most common cause of cancer-related mortality in men. Although roughly 75% of patients are diagnosed with non-muscle invasive bladder cancer (NMIBC), recurrence rates are high even at a low stage and grade. Transurethral resection (TURB) is required for establishing the pathological diagnosis and clinical staging of patients. In daily clinical practice, however, conventional tumor documentation after TURB lacks accuracy, posing a major limitation for patient follow-up and surveillance. Novel technologies to facilitate data documentation and interpretation processes are imperative to maximize patients' outcomes. As part of the RaVeNNA-4pi initiative, our contribution is twofold: first, we propose a bladder 3D-reconstruction method using Structure-from-Motion (SfM). Second, we propose deep convolutional neural networks (DCNN) for cystoscopic image segmentation to improve the interpretation of cystoscopic findings and localization of tumors. 3D reconstruction of endoscopic images assists physicians in navigating the bladder and monitoring successive resections. Nevertheless, this process is challenging due to an endoscope's narrow field of view (FoV), illumination conditions and the bladder's highly dynamic structure. So far in our project, the SfM approach has been tested on a bladder phantom, demonstrating that the processing sequence permits a 3D reconstruction. Subsequently, we will test our approach on bladder images from patients generated in real-time with a rigid cystoscope. In recent years, deep learning (DL) has enabled significant progress in medical image analysis. Accurate localization of structures such as tumors is of particular interest in processing medical images. In this work, we apply a DCNN for multi-class semantic segmentation of cystoscopic images. Moreover, we introduce a new training dataset for evaluating state-of-the-art DL models on cystoscopic images. Our results show that on average a 0.67 Dice score coefficient (DSC) can be achieved.

Keywords: Cystoscopy · Bladder cancer · 3D-reconstruction · Structure from motion (SfM) · Deep learning · Neural networks · Semantic segmentation

1   Introduction

Cystoscopy is one of the most important diagnostic tools employed by a urologist and the current standard of care for the diagnosis and surveillance of BCa. The standard approach is white light cystoscopy (WLC), which has undergone significant technological innovations for improved image quality. Despite these advances, WLC has numerous limitations that affect its accuracy to detect BCa. First, the inability to accurately identify small or flat lesions, including carcinoma *in situ* (CIS), which is missed in up to 20% of cases [1]. Second, the difficulty in distinguishing benign reactive lesions from malignancy, particularly in those with prior transurethral resection (TUR), infections or intravesical BCG therapy. To address these limitations, complementary endoscopic technologies have been developed to improve the detection rate, and include fluorescence cystoscopy (FC), narrow band imaging (NBI), confocal laser endomicroscopy (CLE) and optical coherence tomography (OCT) [2]. Although these methods can improve detection and/or accuracy, they are invasive, expensive, and require a human professional for interpretation which can often be ambiguous.

Artificial intelligence (AI) is reshaping the field of medicine as a key component of computer-aided diagnosis (CAD), treatment selection and patient follow-up. The advent of AI-assisted endoscopy, by training deep learning (DL) algorithms with large image and video databases, has shown promising results, providing CAD software with high sensitivity and specificity for lesion identification and diagnosis [3]. AI-assisted endoscopy can radically change medical practice by aiding physicians in identifying potential areas of malignancy. In the field of urology, cystoscopy may benefit from this approach. However, the applicability of CAD in cystoscopy has not been extensively investigated.

Furthermore, the European Association of Urology guidelines strongly recommend describing all macroscopic tumor features, such as site, size, number and appearance, and mucosal abnormalities during cystoscopy using a 2D bladder diagram [4]. Numerous studies, however, have described written reports and bladder diagrams as suboptimal and suggest they are not only a major limitation for patient follow-up and treatment, but may also play a role in the high recurrence rates [5, 6]. The current documentation approach lacks anatomical accuracy due to the inter- and intra-urologist variability related to the diagram and interpretation of macroscopic findings.

To tackle the challenges stated above, the goal of the RaVeNNA-4pi initiative is to develop a standardized and accurate 3D documentation method, along with image segmentation, by exploiting technological advancements with the goal of contributing to the reduction of BCa recurrence and progression rates. The first part of Ravenna-4pi aims to optimize the documentation of findings with a 3D reconstruction of the bladder from the patients' cystoscopic videos that are collected from cystourethroscopy procedures. In this work, a classical SfM pipeline is adopted to reconstruct the bladder in 3D. As proof of principle, the pipeline was tested on a bladder phantom. A rigid cystoscope with a 30° lens under white light was used to record the bladder phantom. Individual frames are then extracted from the bladder phantom video and point correspondences between the frames are established by feature matching and detecting algorithms.

The second part of RaVeNNa-4pi focuses on DL-based multi-class segmentation of cystoscopic images. Semantic segmentation allows the precise segmentation of bladder

walls and localization of classes of interest in an image. To this end, a novel cystoscopic image dataset was created and annotated by trained investigators. Video frames of interest are extracted and manually annotated using label editor software developed at the Fraunhofer Institute for Physical Measurement Techniques IPM. The annotation was conducted by two trained physicians and subsequently confirmed by a board-certified urologist. The RaVeNNA-4pi cystoscopy dataset currently consists of 4200 annotated images extracted from 62 videos. A DCNN based on U-Net network architecture was trained on this dataset with a performance of 0.67 dice score coefficient (DSC). Furthermore, the automated detection of bladder structures based on cystoscopic images also supports tumor documentation and visualization in the 3D reconstructed model by projecting 2D semantic information to a 3D reconstructed bladder model.

2   3D Reconstruction

The endoscopic images of biological structures are complex and vary from regular scenes. The endoscope's small field of view, varying illumination, large distortions, bladder's dynamic nature, and non-lambertian surfaces makes the 3D reconstruction process extremely challenging. The aim of this work is to 3D reconstruct the human bladder with monocular rigid endoscopic images. The image-based 3D reconstruction follows the standard flow of the SfM pipeline. The pipeline is tested on a synthetic bladder, assuming that the model is static. The 3D reconstruction pipeline consists of mainly two steps: Endoscope calibration and structure from motion (SfM).

2.1   Endoscope Calibration

The endoscope's fish-eyed lenses induce non-linear distortions in the images. For 3D reconstruction, it is crucial to correct the endoscopic images with a calibration routine. For this purpose, it is necessary to know the model-based parameters that define the camera. Cystoscopies in general, use an oblique-viewing rigid endoscope, usually with a 12° or 30° viewing angle. In this work, a standard rigid endoscope with a 70° field of view was calibrated to fit the images for 3D reconstruction. In an oblique viewing endoscope, the field of view is increased by rotating the scope's cylinder. This results in highly blurred images. Therefore, the endoscope was calibrated to obtain the camera's intrinsic parameters such as projection center, distortion parameters and focal length. A previously captured checkerboard pattern and a fish-eye camera was used in the calibration process [7]. Through an OpenCV camera calibration routine described by Zhang [8], the intrinsic camera parameters were computed by capturing the images of the checkerboard pattern from various view angles. From this routine, all the required camera intrinsics were obtained with an accuracy of 0.10 mm mean re-projection error.

2.2   Structure-from-Motion

In Computer Vision, there are several 3D reconstruction approaches such as, photometric stereo, shape from shading, shape-from-silhouette, and shape-from-structured light [9]. These approaches assume uniform illumination, require rich-feature structures and

correspondences and costly experimental setups. In this work, the bladder 3D model was reconstructed using the structure-from-motion algorithm proposed by Wu et al. and Widya et al. [7]. The SfM term is understood as generating structure using the movement of the optical sensor. It is a cost-effective photogrammetric approach [10]. SfM technique differs from other conventional photogrammetric approaches by solving the camera's geometry, position and orientation without specifying any a priori known 3D positions [11]. Rather, these are solved simultaneously with a bundle adjustment technique that accesses a database of features that are extracted from the image pairs.

Frame Processing. The endoscope's optical system has a video processing frame rate of 50–60 Hz, with a resolution of 1920×1080 Pixel. At first, the image frames are extracted from each video. The extracted images are then further undistorted to remove the radial and tangential distortions by estimating the camera's intrinsic parameters, motion parameters and the 3D coordinates of the scene in the real world [6].

Surface Reconstruction. The initial processing step in the SfM process chain is to identify the features in individual frames as shown in Fig. 1. For this purpose, a popularly used Scale-Invariant Feature Transform (SIFT) algorithm is used. This algorithm detects and describes the features in distinctive frames and also determines the correspondences between these individual feature points [12]. The SIFT detector recognizes the interesting feature points using the Difference of Gaussian (DoG) function. Each feature point is assigned with one or more orientations depending on the direction of the local image gradient. The obtained feature points are subjected to scale and orientation invariant. The uniqueness of each feature point is realized by a SIFT descriptor that is obtained by combining a high-dimensional vector that represents image gradients within a local region of the image [13]. A distance ratio of 0.8 as proposed by Lowe [12] is used to eliminate the false matches. Though Lowe's ratio already eliminates a large number of false matches, it is imperative to correct the matches by discarding the outliers. In our approach, RANSAC is used to filter the outliers. Finally, tracks are established linking the feature points over a set of frames. The tracks that have feature points in a minimum of three frames are taken for reconstruction and the ones that do not fall into this criteria are omitted automatically. Feature tracking helps to ignore the transitory features of the dynamic scenes such as moving air bubbles, blood clots and stone particles in the bladder.

Now that the relative pose between the camera pairs is known, each corresponding feature points can be triangulated into a 3D point. A 3D point is the intersection of two projection rays that pass through the camera's center and the feature point associated with the corresponding frame. Due to measurement inaccuracies of the corresponding key points, the rays usually do not overlap. The 3D point should be chosen such that they minimize the error. In 3D reconstruction algorithms, triangulation minimizes the re-projection error between observed and predicted feature points, by varying camera parameters and triangulated 3D points. Then to improve the inaccuracies in feature detection, a complex, non-linear minimization of a suitable cost function is used. In the last step of the SfM pipeline, bundle adjustment is performed on previously constructed cameras to optimize camera parameters and feature points' 3D coordinates [14].
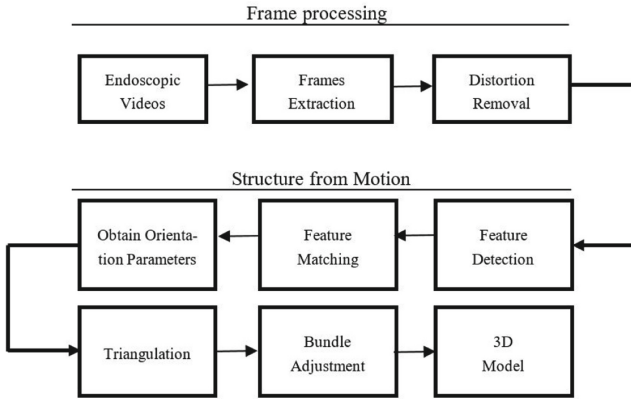
Fig. 1. 3D reconstruction pipeline

2.3 Current Work and Results

In the current work, the SfM pipeline was tested on a bladder phantom using a rigid endoscope 27005BA by Karl Storz SE & Co. KG with 30° view angle and a 30 cm in length. The endoscope was connected to a camera H3-Z FI TH102 from Karl Storz to record the phantom model. An experimental setup was built by Karl Storz. The setup contains a pivot point level with a ball joint to which the endoscope is fixed to enable the movement of the endoscope, inertial and mouse sensor board and a video recording pattern. The endoscope was coupled to an inertial sensor, a mouse sensor, and an electromagnetic tracking. The inertial sensor delivers the quaternions and the endoscope's rotation angle while the mouse sensors record the horizontal and vertical translation of the endoscope. The electromagnetic tracking has two sensors embedded in it. A static sensor was attached to the feature plate and served as a reference sensor while the other is fixed on the endoscope's tip, which changes the position with the movement of the endoscope. The distance from the feature plate and the endoscope is given by the difference between the two sensor's values in Millimeter.

A semi-circular phantom was made of synthetic plastic with a 45 cm longitudinal diameter. To obtain a rich feature texture in the synthetic bladder, numbered pieces of paper were pasted inside. The endoscope was introduced through the open end of the phantom. To ensure the bladder's maximum coverage, the recording of the phantom was taken through the metallic template plate that is inserted before the feature plate as shown in Fig. 2. A freehand spiral scan was likewise performed. The scans were performed with constant slow speed as possible to maintain the frame. To allow constant illumination while recording, the distance to the feature plate and the endoscope tip was maintained.

After the video is recorded, 1694 frames were extracted from the recorded video. All 1694 images were used for reconstruction and 165,325 feature points were used for triangulation. For reconstruction, image pairs that contain a minimum of 8 inlier matches with their appearance in at least three images were used. The obtained 3D points are further visualized as point cloud in cloud comparison.

(a)   (b)

(c)   (d)
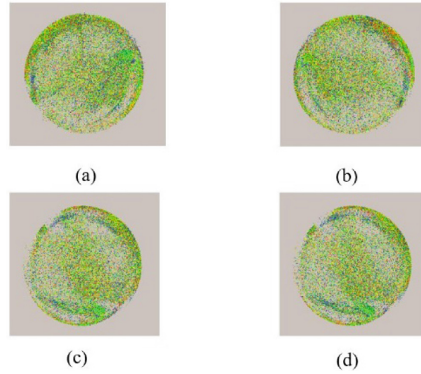
Fig. 2. Reconstruction results of a bladder phantom. (a) and (b) show the sparse point cloud from the top and bottom view. (c) and (d) show the sparse point cloud from the left and right side view.

Figure 2 shows the bladder phantom's reconstruction results from the top, bottom, left and right side view of the point cloud. The results show, that the chosen SfM pipeline enables a fully automated 3D reconstruction of the bladder phantom. Moreover, the tested SfM pipeline was applied to clinical bladder dataset obtained from the University of Freiburg – Medical Centre. The surface reconstruction with SfM works, however it fails in the case of images with moving particles like air bubbles, blood clots and stone particles. In future work, the aim is to overcome these challenges during the reconstruction to obtain a dense point cloud. Moreover, additional information from the inertial sensor will be applied in the next step to improve the 3D reconstruction.

3   Deep Learning

In recent years, AI has played a major role in computer aided diagnosis to assist physicians in improving diagnoses and offering patients effective and personalized treatment [15]. A considerable amount of progress in the area of medical image analysis can be attributed to the application of neural networks, also known as artificial neural networks [16].

3.1   Feed-Forward Neural Networks

The design of neural networks is based on a simplified model of biological neural systems. Neural networks are made up of artificial neurons that are organized in the form of input, hidden and output layers. Each layer consists of neurons that are connected with each other in a directed acylic graph, with no connections between neurons within the same layer. A single neuron consists of inputs, an activation function and outputs. The input neurons are determined by the number of pixels in an image, and the output, for instance, in a classification task by the number of classes. The connections between neurons have learnable weights. The input and weights interact multiplicatively and their sum is passed through a non-linear activation function. Common choices of activation functions are rectified linear unit (ReLU), hyperbolic tangent and sigmoid function.

Training a neural network in essense is optimizing the network parameters namely the learnable weights and biases (a constant input) via backpropagation method. Backpropagation uses stochastic gradient descent to optimize the weights by evaluating the gradient of a loss function.

Convolutional Neural Networks (CNNs) are neural networks that are well suited for data with grid-like structures. CNNs can learn complex hierarchical data feature respresentation. They are designed to exploit the spatial connectivity of input data and are translation invariant. The sparse connections between the neurons reduces the amount of parameters that need to be optimized and enables CNNs to model complex functions and train deep architectures effciently. CNNs have been successfully used in various image analysis tasks such as image and video classification, image captioning and image segmentation. This success is driven by the availability of data and computation resources (GPUs).

In this work, we focus on image segmentation, which is one of the main applications of CNNs. Image segmentation assigns each pixel in an image a probability value, that the pixel belongs to a certain class i.e. per-pixel image classification. Fully convolutional neural networks (FCN) [16] are CNN invariants that are often employed for this purpose. These networks have an encoder-decoder architecture and can be trained end-to-end. The encoder part has similar structure to a classification network and successively produces feature maps from input data. The decoder part upsamples these feature maps to high resolution segmentation maps to achieve a per-pixel classification of input images.

Convolutional Neural Networks (CNNs) have been shown to be valuable in the analysis of various urological findings. DL-based automated tumor lesion segmentation – without the need for level set segmentation – on 2D CT slices and 3D CT volumes was recently studied by Hadjiiski et al. [17]. Furthermore, Dolz et al. [18] showed that a fast and precise segmentation of bladder walls and tumors can be obtained on T2-weighted MRI images with deep fully CNN. Kouznetsova et al. studied the classification of bladder cancer stages [19].

3.2   RaVeNNA 4pi: Semantic Segmentation

As part of the RaVeNNA-4pi project, we focus on multi-class segmentation of cystoscopic images with CNN assisstance. Semantic segmentation assigns each pixel in an image a semantic class of interest. This assists in achieving object recognition and understanding of scenes, that is beneficial for cystoscopic image diagnostic analysis. To this end, a novel dataset of 4200 images was collected from 62 cystoscopic videos. The videos used for this RaVeNNA-dataset were acquired from procedures performed with a rigid cystoscope and a 30° lens under white light. The extracted images were manually annotated by trained physicians using label editor software and subsequently confirmed by a board-certified urologist. For RaVeNNA multi-classification, the classes of interest are: (1) papillary tumor, (2) carcinoma in-situ suspicion, (3) resected tumor, (4) bladder diverticulum, (5) bladder stone, and (6) left and right ureteric orifices.

Our method is based on U-Net [15], a successful fully convolutional network (FCN) for medical image segmentation. Precise segmentation of images with fine structures

requires a combination of semantic information and spatial context. The encoder of the U-Net accumulates semantic information and passes it to the decoder. The encoder additionally passes spatial context information in the form of higher resolution feature maps through skip connections to the decoder. The decoder combines these two sources of information to precisely segment fine structures in an image.

For training purposes, we use 70% of the data and the rest for evaluation and testing. Due to the highly imbalanced nature of the data, the segmentation results are biased towards majority classes. To address this problem, we use weighted cross-entropy loss during training. We use Adam optimizer with initial learning rate of 3×10^{-4}. To improve the robustness of the network, we apply data augmentation methods by randomly applying a sequence of data augmentation methods. To this end, we apply rotation with $(-45°, 45°)$, zoom (0.2 factor) and translation (0.1%). We report results of 0.67 average dice score coefficient (DSC).

The final model is trained for 100 epochs with batch size of 2. The implementation was based on Keras DL framework with Tensorflow backend. All experiments were carried out on a system equipped with Nvidia GeForce GTX 1080 GPU with 11 GB memory. Results are visualized in the form of segmentation overlay over example images as shown in Fig. 3. The study shows that deep CNNs can be successfully used for image segmentation of cystoscopic images.
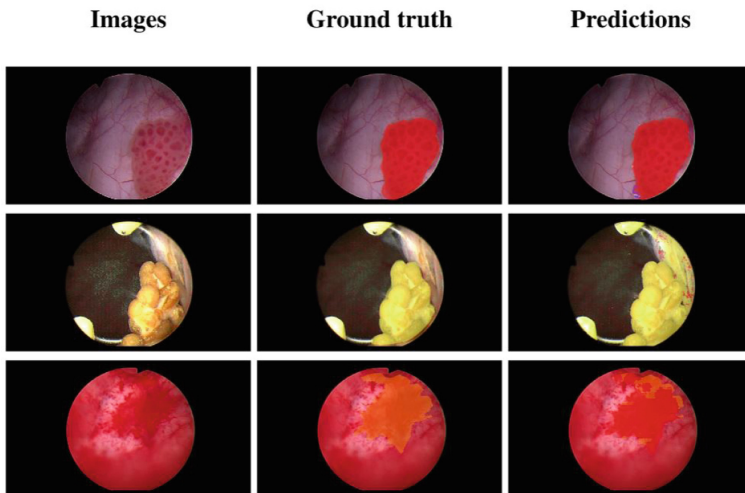
Fig. 3. Visual comparison of the performance of the deep convolutional neural network on RaVeNNA-4pi dataset. The example images are of papillary tumor, bladderstone and ca in-situ suspicion in the rows respectively. The second column are the manual annotations of experts and the last column are the automatic segmentation of a neural network.

4   Conclusion and Outlook

In this study, we demonstrated that DL can be applied in cystoscopic images analysis. Moreover, we apply U-Net for multi-class segmentation of a novel cystoscopic dataset,

introduced in this work. We demonstrate that data augmentation and weighted loss can help combat the class imbalance problem. In future work, we will focus segmentation of 3D-reconstructed bladder model with DL assistance.

The presented work also demonstrates our current developments in bladder phantom 3D reconstruction from endoscopic images. The frames are extracted from the video streams and are undistorted by a calibration routine. With undistorted images as input, the SfM pipeline begins with feature detection and matching. The surface approximation is carried out by triangulation and bundle adjustment. The results show that the SfM pipeline can be adopted to 3D reconstruct the bladder. The SfM algorithm is then applied on the clinical dataset. The algorithm performs well when bladder images are free from visual obstructions but fails when moving particles interfere with the endoscope. In future work, the aim is to overcome the challenge posed by moving particles and occlusions and also incorporate sensor information during the reconstruction process to obtain a dense point cloud.

Conflict of Interest. The authors declare that they have no conflict of interest.

References

1. Fradet, Y., Grossman, H.B., Gomella, L., et al.: A comparison of hexaminolevulinate fluorescence cystoscopy and white light cystoscopy for the detection of carcinoma in situ in patients with bladder cancer: a phase III, multicenter study. J. Urol. **178**(1), 68–73 (2007). https://doi.org/10.1016/j.juro.2007.03.028. discussion 73

2. Soubra, A., Risk, M.C.: Diagnostics techniques in nonmuscle invasive bladder cancer. Indian J. Urol. **31**(4), 283–288 (2015). https://doi.org/10.4103/0970-1591.166449

3. Rees, C.J., Koo, S.: Artificial intelligence - upping the game in gastrointestinal endoscopy? Nat. Rev. Gastroenterol. Hepatol. **16**(10), 584–585 (2019). https://doi.org/10.1038/s41575-019-0178-y

4. Babjuk, M., Burger, M., Compérat, E., Gontero, P., Mostafid, A.H., Palou, J., van Rhijn, B.W.G., Rouprêt, M., Shariat, S.F., Sylvester, R., Zigeuner, R.: EAU Guidelines edn. presented at the EAU Annual Congress Barcelona (2019)

5. Lurie, K.L., Angst, R., Zlatev, D.V., et al.: 3D reconstruction of cystoscopy videos for comprehensive bladder records. Biomed. Opt. Express **8**(4), 2106–2123 (2017). https://doi.org/10.1364/BOE.8.002106

6. Péntek, Q., Hein, S., Miernik, A., et al.: Image-based 3D surface approximation of the bladder using structure-from-motion for enhanced cystoscopy based on phantom data. Biomed. Tech. (Berl) **63**(4), 461–466 (2018). https://doi.org/10.1515/bmt-2016-0185

7. Widya, A.R., Monno, Y., Imahori, K. et al.: 3D Reconstruction of whole stomach from endoscope video using structure-from-motion. CoRR abs/1905.12988 (2019)
8. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000). https://doi.org/10.1109/34.888718
9. Péntek, Q.: 3D-Rekonstruktion menschlicher Harnblasen anhand Videoaufnahmen eines monokularen Endoskops (2016)
10. Hanna, S.: Structure from Motion (SfM) (2019). http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson8-2/SfM.html. Accessed 12 Nov 2019
11. Nyimbili, P., Demirel, H., Seker, D., Erden, T.: Structure from motion (SfM) - approaches and applications (2016)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)
14. Triggs, B., Mclauchlan, P., Hartley, R. et al.: Bundle adjustment – a modern synthesis. In: International Workshop on Vision Algorithms, pp. 298–372, September 2000
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks For Biomedical Image Segmentation. Springer, Cham (2015)
16. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation (2015). https://arxiv.org/pdf/1411.4038
17. Ma, X., Hadjiiski, L., Wei, J. et al.: 2D and 3D bladder segmentation using U-Net-based deep-learning. In: International Society for Optics and Photonics, p. 109500Y (2019)
18. Dolz, J., Xu, X., Rony, J. et al.: Multi-region segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks (2018). https://arxiv.org/pdf/1805.10720
19. Kouznetsova, V.L., Kim, E., Romm, E.L., et al.: Recognition of early and late stages of bladder cancer using metabolites and machine learning. Metabolomics **15**(7), 94 (2019). https://doi.org/10.1007/s11306-019-1555-9

A Biomedical Survey on Osteoporosis Classification Techniques

Zahra Amiri[1], Fatemeh Tavakoli[2], Vala Mehryar Alviri[3],
and Morteza Modarresi Asem[3(✉)] (iD)

[1] Department of Molecular and Cellular Biochemistry, Tehran Medical Sciences Branch,
Islamic Azad University, Tehran, Iran
zramiri97@gmail.com

[2] Rajaei Cardiovascular, Medical and Research, Iran University of Medical Sciences, Tehran,
Iran
f.tavakoli95@gmail.com

[3] Department of Clinical Biomedical Engineering, Tehran Medical Sciences Branch, Islamic
Azad University, Tehran, Iran
mehryarvala@gmail.com, Modaresi.bme@gmail.com

Abstract. Osteoporosis is a severe malady and its primary diagnosis is highly significant now. The appropriate diagnosis of osteoporosis leads to more appropriate management in terms of prevention and sufficient pharmaco-logic or surgical treatment. In this study, the researchers focused on the primary biomedical Osteoporosis classification techniques. Such techniques are categorized into four groups including Radiographic Techniques, Bio-Markers Classification, Invasive Techniques, and Biosensors Classification. Authors in this study attempted to explain the direction of future studies in the field of Osteoporosis diagnosis by presenting an accurate classification of Osteoporosis diagnostic techniques. Finally the role of stress and bone displacement in osteoporosis is simulated.

Keywords: Osteoporosis · Medical imaging · Bone health · Computer classification · Biochemistry biomarkers · Biomedical assessment

1 Introduction

Bone is an integral organ that has the power to constantly restructure the lives of people. Bone health relies on the composition of the cells in the body in the restoration process [1]. Osteoporosis is the most prevalent metabolic bone condition recognized for a simple bone tissue degradation, which contributes to a cure. In addition, the medical classification is to be accepted for the occurrence of a certain illness, depending on the identification of signs and examination of specific features. Osteoporosis, though, is challenging to predict since it moves quietly into the body with limited indications of back pain. Such early signs relate to age, and therefore the mystery individual is only identified after a bone fracture. For an accurate diagnosis of osteoporosis, the radiographic methods are required as they provide accurate data in the diagnosis of the fatal illness. Finding out

the outcomes and direct intervention strategies, stratifying the subjects of clinical trials, tracking medication responses and, most significantly, in customized health care are also helpful biomarkers for the identification of patients in the clinical course earlier for a disorder. While providing good knowledge to direct the clinical decision-making, such biomarkers are related in a significant way to the legalization and/or pathogenic use of such therapeutic/interventional approaches [2, 3].

Osteoporosis is a subtle condition that can escalate to a terrifying injury of hip and split, and causes a disorder that is linked to morbidity in the aged, and four times higher in the disabled. As a critical problem of elders' public health, therefore, the heavy burden imposed on society can be predicted and early diagnosed. Biosensor is a platform for early Osteoporosis detection. Essentially, biosensors are devices where the biological detector is combined with a detection array with a transducer. Biosensors explore "easy-to-use" tools to aid with early diagnosis and infectious management. The approach to effectively handling multiple diseases is early diagnosis. To order to claim a target molecule, biosensors use certain properties of the physical and biological materials that impact signal transductions. Fast response and high sensitivity are the essential blessings of biosensors. In addition, the essential advantage for the purpose of care devices reminiscent of biosensors would come with the integration of nano-materials, micro fluidics, automatic samplers, and transduction devices on one chip. Biosensors are used as new analytical tools for checking the drugs [4]. Eventually, the authors evaluate the role of stress and bone displacement in osteoporosis and shows that the cooperating bone material and structure of bone confers strength on a skeleton, and the material or structural components of skeletal tissues should not be measured on macroscopic scales, either morphological or as an appropriate surrogate measure, in the isolation and/or development of bone strength, or as mechanical competence.

1.1 Motivation

Todays, Radiological approaches test for osteoporosis. Due to the disadvantages such as excessive intake of radiation, the cost of radiological techniques and the need for the device to specialize, alternative methods are required. The objective assessment of radiography, biomarkers, biosensors and intrusive methods plays a role in bone mineralization and can be a good alternative for detection of osteoporosis and, in fact, in therapies follow-up [26]. There are many research regarding osteoporosis diagnosis classifications. Over the last 30 years certain modern diagnostic methods have been established while those technicians have their limits and trusted details for first level of the disease is not obtained. Thus, the precise categorization of such methods and the evaluation of their weaknesses or strengths can help researchers or physicians to choose a clear pathway to research or treat this kind of disease. The fast development in the area of osteoporosis diagnosis methods over the past decades, both in the research and product development level, is mainly due to: (i) developments in miniaturization technologies, (ii) the application of novel bio-recognition techniques, (iii) novel nano-materials and nanostructured devices, (iv) better interaction between life scientists and engineering/physical scientists. Thus, this study aimed to divide various kinds of osteoporosis and their applications in detecting this disease.

This study is organized as follows. Section 2, presents a synopsis of Related Works. Section 3 presents Overview of Medical Assessment of Osteoporosis. In Sect. 4, Classification of Osteoporosis Detection Techniques is expounded. Sections 5 is devoted to Bone Turnover Markers (BTMs). In Sect. 6 Proposed Simulations and Experimental Results is presented in detail Ultimately, Sect. 7 presents Conclusion and Future Works.

2   Related Works

Wen et al. [5], developed a micro-manufactured strain gauge to be used in bones. In the analysis, an electro-chemical test for the control of bone tension during bone reformation showed the film gage inserted in a Polydimethylexane membrane (PDMS) that displays more specific strain sensation than those commercially available. An extremely pliant compound, real-time sensor was proposed by the authors at [6] according to two interlocked sequences of platinum-coated polymer nano-fibers backed by PDMS layers with high-aspect ratio. The variations in the electrical resistance were reported as a feature of the strain for three loads; friction, shear and torsion. This integrated strain sensor is considerably sensitive and repeatable demonstrating its potential for real time monitoring Implant medical equipment implementations.

Fresving et al. [7] extended the Bragg fiber optic grafting sensor to calculate the bone pressure by performing both acrylic and human femoral diaphysis experiments. The optical fibers are very narrow and bones can easily be adhered to using approved polymethylmethacrylate (PMMA) for bonding with the Food and Drug Administration (FDA) and left in a patient. The sensor-to-load linearity shows that Bragg grafting fibers are ideal for in vitro and in vivo functional calculation of strain on bones. The Acoustic Emission (AE) operation of human femoral bone fracture was measured by the [8] researches. The analysis showed that AE behavior can be used to assess which cracking begins even sooner than a macro-fracture through piezoelectric detectors on the morphological traits layer. Yael et al. [9] provided the initial phases of a screening system for fractured and damaged bones. Mentioned solution is based on a time tracking of secondary spikes transmitted backward when the limb is illuminated with a laser and when the pressure injection is delivered regularly by a loudspeaker. Initial studies have shown the efficacy of the experimental bone testing system. The average motion of the bone is proportional to the shift of the spinal pattern induced by the changes in the speaker. The paper [10], used for early diagnosis of bone loss, used the architecture and extension of a portable Internet for sensing.

The Type I collagen (CTx-I) biomarker was calculated as the marker for bone resorption in serum samples. Urinary CTx-I is a balanced interaction with urinary creatinine in order to determine urine movement shifts. CTx-I and creatinine from the same aliquot should therefore be measured. The CTx-I concentration should be adjusted with creatinine levels by means of Eq. 1 once the creatinine concentrations of the urine sample are established:

$$Corrected\ CTx - Ivalue\left(\frac{ng}{mmol}\right) = \frac{1000 \times CTx - I\left(\frac{\mu g}{L}\right)}{Creatinine\ level\left(\frac{mmol}{L}\right)} \tag{1}$$

A molecular manufacturing technology can be used to synthesize different memory locations for creatinine. Creatinine is a polar compound molecule, and strongly hydrophobic monomers like methacrylic acid (MAA) might assume a better relationship. CTx-I concentrations of 0.09 ng/mL could be measured by the point of care unit. The range 0.1–2.5 ng/mL displayed outstanding longitudinal activity that represented the usual bone loss diagnostic comparison ranges [10].

Authors in [11] used immune-sensing biosensor production to diagnose the type-1 collagen bone turnover C-terminal telopeptide marker. Dithiodipropionic acid was mounted on a gold electrode by a Self-Assembled Monolayer (SAM). Streptavidin and biotinylated C-terminal telopeptide antibody were then paired with the self-assembled monolayer. In order to describe every stage of the SAM/streptavidin/biotinyled antibody binding, electrochemical impedance tests were carried out. Sirivisoot et al. [12], created the electrochemical sensor by MWCNTs-Ti electrode in order to feel the forming of bones through the extracellular elements of osteoblasts which display excellent electrochemical properties. In addition, the electric signals provided by the MWCNTs-Ti electrode biosensing can be viewed later in as bone disease detection information.

Paper [13] proposed that a sensor measuring CTX was designed for early diagnosis of bone loss, so that Electrochemical Impedance Spectroscopy (EIS) measurements can be made. Streptavidin agarose on the sensor region is immobilized and an interdigital planar sensor was used to select the antigen-antibody solution. The practical sensor surface was used to detect the target molecule in several concentration levels of CTX-I. The analysis with the assistance of the respective equivalent circuit operating sensor technology utilizing Complex Non-linear Least Square (CNLS) received initial results. The statistics were motivated to create a reliable system of promoting and existing structures support. In addition, researchers in [14] have proposed a CAOD methodology that tests Bone Mineral Density (BMD) efficiently and produces an osteoporosis study based on the DXA exam. DXA photographs using a non-local mean filter, computer pixel mark random forests respectively, are categorized into the CAOD model and are placed in regions of interest with greater accuracy. The random forest pixels categorize the soft or bone tissue of a pixel and derive contours from the binary picture in order to identify areas of significance and measure BMD from the pixel of bone and soft tissue. The statistical analysis was used to test precision and accuracy of BMD measures by way of standard deviation and correlation coefficients. The CAOD model found that the standard deviation from manual tests on three different individual data sets was 0.0029 while the standard deviation from manual assessments is reported as 0.1199.

3 Medical Assessment of Osteoporosis

3.1 Background

Almost all of the patients stay anonymous and untreated, especially bad patients 60 and older, the rise in Osteoporosis breaks is characterized by a weakening of their bone mass and bone microstructure in the mid of a grain effect on the incapacity and mortality disorder. Bone is a device that is metabolically complex and capable of recovering the entire life. Amid crest body mass, bone is experiencing continuous remodeling by bone resorption in the central multi-cell unit of bone dubbed the 'bonus remodeling device'

amid a structure successively. Specific biomolecules are recognized as bone turnover markers (BTM) produced in the bloodstream by bone absorption and organization [1]. For bone turnover evaluation, biochemical markers include enzymes and non-enzymatic peptides from both cell and non-cell bone compartments. BMT identified by metabolic process in which they are presented:

- Bone formation markers
- Bone resorption markers

Remodeling has a critical impact on mechanical function and the power of the bone directly. In the ordinary statues to provide stability and balance in skeletal metabolism, it should be tightly coupled to function of bones to each other [3]. When bone formation and resorption balance is impaired, osteoporosis happens and in this problem, the process of bone resorption is more than formation. Figure 1 indicates the both normal bone and bone with osteoporosis [18].
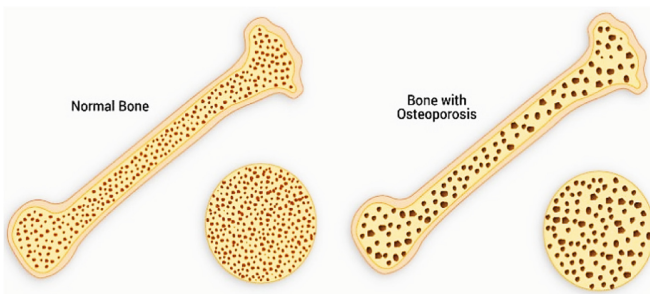
Fig. 1. Displaying of both normal and osteoporosis bone [18].

To generate bone loss there are two major pathophysiological processes. The key effects of the imbalance in steroid hormones involve the fibrous tissue, which may be due to weakness and improper use of cortical adrenal steroids, and to pathological pathologies over time. Bone loss would grow by 70 years. Table 1 contains four different categories of osteoporosis of reduced bone mass depends on the calculation of bone mineral density by Dual Energy X-ray Absorptiometry (DXA) [18].

Table 1. Osteoporosis and low bone mass based on bone mineral density measurement [18].

Category	Bone mass measurement
Normal	T-Score ≥ -1
Osteopenia	T-Score <-1 and >-2.5 SD
Osteoporosis	T-Score $\leq$ to -2.5 SD
Severe osteoporosis	T-Score $\leq$ to -2.5 in the presence of fracture

Abbreviations: SD; Standard Deviation, T-Score; Adult Population of the Same Gender.

4 Classification of Osteoporosis Diagnosis Techniques

Several methods exist for the diagnosis of Osteoporosis as shown in Fig. 2.

4.1 Radiographic Techniques

The method used to determine the bones in a side cortical thickness (metacarpals), which is used for more than 30 years, focuses on tests of bones. The methodology measuring Bone Mineral Density (BMD) is an effective method to osteoporosis treatment. Reduces the amount of tissue, which contributes to weak and fracture-prone bones. For BMD research, multiple X-rays are accessible. For example; the main BMD research is still double-energy X-RAY Absorption (DXA). The bone density within a certain area of the bone is calculated by double energy absorption (gr/cm^2). Such technology provides the benefits of higher precision, reduced ionizing exposure to radiation, a shorter scan period and proximity to cortical and trabecular bone mass at annexed and axial locations. The restriction, involves staff costs, a requirement for training X-Ray engineers and a lack of portability. Furthermore, DXA spine scans may imply that osteophytes, aortic calcifications and arthritic degenerative adjustments have been shown to be increasing spinal BMD.

The origin of osteoporosis is not about mass nor bone health. The protection of a bone means that it requires sufficient strength to resist accidental fractures triggered by voluntary loads. Therefore, osteoporosis treatment would be both bone strength and muscle strength biomechanical. It includes two types of problems: (a) the correct measurement of bone structure, structural design and power; and (b) correlating the measures with specific muscle strength indicators. Since uniform densitometry is not useful for determining the endurance or bone mass, it should be used in other, better cross-sectional bone structure studies such as those given by the QCT, PCT, MRI, or related methods. The benefit of pQCT is the capacity of the algorithm to take into account all "Body Energy" variables, content and design influences, and to provide data on muscle transverse segments. In the long term, X-rays are particularly poorly accurate and reliable and have been greatly replaced by new technologies [14].

1- CT Scan Imaging: ACT bone density scan (CT) is a highly precise diagnostic technique to assess the risk of future fractures. CT Imaging techniques like QCT and μCT
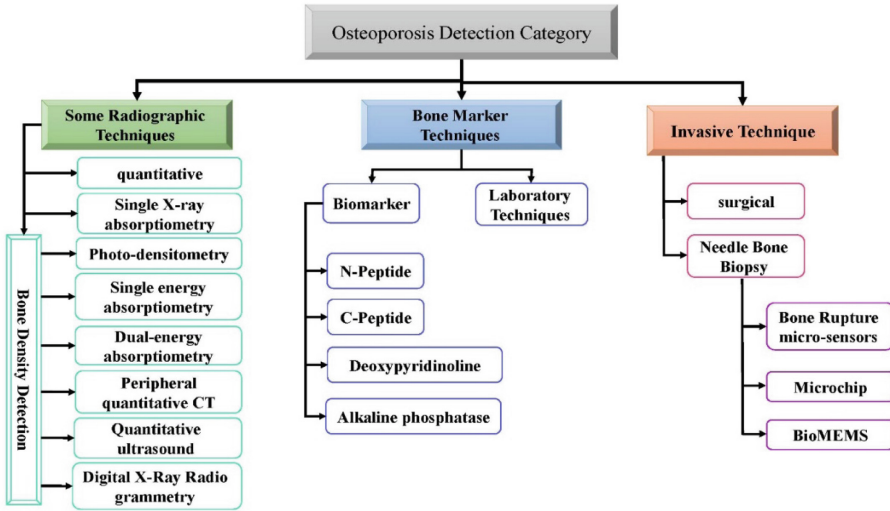
Fig. 2. Categorization of Osteoporosis Detection Techniques.

are more usable and practical than MRI devices. More information can be found. The Bone Mineral Density (BMD) evaluation in the radius and the tibia, and measurement of trabecular structure of the forearm, is available with devoted peripheral CT scanner. The initial focus of QCT was on trabecular BMD measurements in transverse CT single slices at the middle and forearm lumbar levels. A modern QCT discovery system is still used to evaluate BMD. For fact, the imaging of specimens is usually performed solely on micro-CT (μCT) scanners for the study of bone structure.

The BMD study of the trabecular and cortical compartments is one benefit of QCT relative to DXA, which has already been suggested for the original 2D single cut applications. While the trabecular bone is metabolized and can serve as an early indicator of treatment progress, cortical bone, especially, the hip, is significant in assessing the likelihood of fractures. Figure 3, demonstrates the CT Imaging for distal radius bone osteoporosis and the Three-Dimensional (3D) results of diagnosis [21].

2- Single-Energy Absorptiometry: In peripheral locations like the wrist and heel, this measured bone mineral. A radioactive energy source (typically iodine125) was used by single photon absorptiometry for estimating the amount of bone mineral at the peripheral measurement site. Over these past years, X-ray Single-Energy (SXA) has substituted SPA with peripheral skeleton (heel and wrist) measurements because of its increased reproductivity and ease of use. SXA prohibits nuclear energy sources from being collected and disposed of, includes exposure in the bath portion, and can also calculate bone mass at the periphery bones such as forearm bones and legs bones [14].

3- Dual-Energy Absorptiometry: Checks of bone density provide painless, non-invasive and health checks. The dual-energy absorptions were designed to determine the amount of bone that could not be calculated with a mono-energy system in areas of the backbone (lumbar spine, hip and whole body). The most widely employed technique
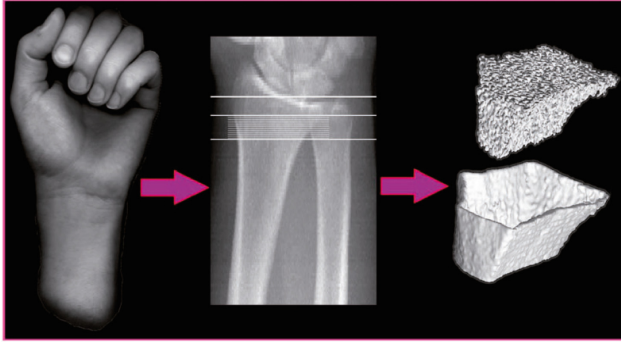
Fig. 3. 3D Medical Imaging of the distal radius using the μCT system, with images showing the region of the distal radius imaged (center) and the resulting 3D trabecular and cortical envelopes (right) [21].

at these sites is now dual-energy X-ray absorptiometry (DXA), which can also measure bone and wrist with high precision and very low radiation exposure [14].

4- Photo-densitometry: Radiographic Absorptiometry (RA) has historically utilized regular X-ray hand photographs and a graded aluminum connected to the distal forearms. A video camera records the radiological picture of the hand and the wrist and quantifies the gray levels of the hand sample, which is contrasted with gray levels of the bone mineral density norm (BMD). It calculates the cortical thickness of the bones. Photo-densitometry consists of an aluminum-step-wedge comparison of the optical density of bones X-ray. Although this method was cost effective and easily accessible, it was poorly reproducible. Such failures were minimized by computer-assisted approaches and several commercial applications were produced in recent years. While radiation absorption meter is typically less reliable than DEXA, this is promising to test cases of osteoporosis as a cost-effective method. In addition, research is necessary to evaluate the utility of fracture prediction and therapy tracking [13].

5- Peripheral Bone Densitometry: Such devices have the benefit of being less costly, compact instruments, acceptable precision and minimal radiation exposure to a wide number of single-energy radiation detection equipment, peripheral DXA and peripheral CT. The use of clinical ultrasound for osteoporosis monitoring and risk management has been increased. The ultrasound densitometer accounts for bone elasticity and stiffness in distant locations using the speed of sound and distance measures of ultrasonic absorption. The advantages of this technique are low costs and the absence of ionization exposure [14].

6- Peripheral Quantitative CT (pQCT): The QCT tests the stiffness of the trabecular vertebral tissue, the spongy bone in the middle of the backbone. The most popular methods for the assessment of BMD are dual X-ray absorption (DXA) and quantitative computed tomography (QCT). DXA measures BMD, particularly trabecular and cortical bones, in two dimensions, with results that show the areal density (gram per centimeter in square). QCT enables volumetric trabecular bone density to be determined

by calculating the calcium hydroxyapatite in milligrams per cubic centimeter without superposed cortical bone or other tissues. QCT can prevent DXA-associated DMD over-estimation from spinal, AAC and other sclerosis lesions, such as the islands of bone. Although QCT can be more prone to osteoporosis diagnosis in a greater population, it must be confirmed. PQCT systems are QCT instruments that were modified for remote measures like the hand [14].

7- Ultrasonic Measurement: Bone assessment at a variety of skeleton locations, including the foot, forearm, wrist, and lower leg. The heel form is mainly made up of trabecular bone that matches the vertebral structure. Broadband ultrasound weakness and sonic intensity (or ultrasound level) at the heel are the most commonly studied techniques. Although, such methods do not contain ionizing radiation and can provide any data on the fundamental connections between bone developments and bone mass, their use is very intriguing. For the reasons pointed out below, these methods cannot be used for osteoporosis research, but do include loans for the assessment of elderly people's break-downs. Figure 4 is a visual Wireless Implantable Passive Strain Sensor (WIPSS) for ultrasound control of device bone deformity. The processor of the sensor was focused on hydro-mechanical results [21].
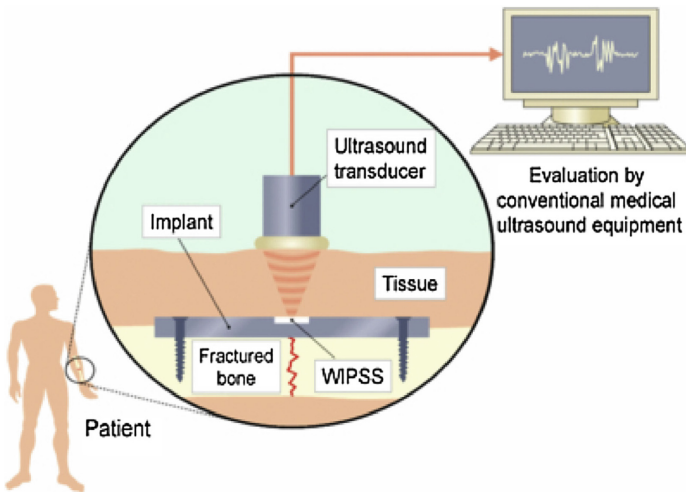
Fig. 4. Ultrasonic measurement bone mass [21].

8- Digital X-ray Radiogrammetry (DXR): Recently, this old method has become more of a concern in computer technology. The Pronosco X-posure method measures a forearm bone mass by using a single, simple radiograph of the hand and wrist to determine the cortical thickness of the bone. Of cortical porosity and striation, the BMD approximation defined as DXR-BMD is updated. The results revealed that this method is very applicable in its potential for inclusion in patients with low bone weight of the spine and/or hip or osteoporosis disorders, at least as well as other methods of marginal bone assessment [19].

9- Double Photon Absorptiometry: A photon beam containing two different energy concentrations is the definition of double photon absorption (DPA). The soft tissue retains one energy gain and the other the bone. The component of the soft tissue can be subtracted mathematically to identify BMD [19].

10- Neutron Activation Analysis: Aslow neutron limb from a generator that is used to transform the soft tissue into a thermal neutron bombards a limb. The nucleus of calcium ion absorbs a warm neutron. The core is radioactive. Nucleus decay emanates photon which can be measured by a Geiger counter, gives an insight into bone mass and reduces osteoporosis [19].

11- Spectroscopic Techniques: Visible spectroscopy (UV-Vis) Optical techniques including Fourier Transform Infrared Spectroscopy (FTIR) are used to detect bone markers which emphasize the noninvasive methods for early diagnoses of osteoporosis. Two types, doctors and healthy people checked blood plasma samples. The two optical approaches showed strong differences in the spectrum; for example, the sensitivity difference between the two classes for OP individuals. Variations in the spectrum of the two vials permit faster and economic oppression of future patients with a bone densitometer checked optical strategies in the hospitals. This is an almost simple, efficient and successful new procedure [19].

12- MR Imaging: Progresses in 3-T imaging and enhanced belt modeling software and hardware have allowed substantially improved the imagery of trabecular bone architecture. Calibration and confirmation experiments have shown that trabecular framework-derived RM measurements agree with in vitro studies in the sense of histology, micro CT and biomechanical power. MR imaging is particularly attractive in experimental and clinical studies due to the lack of radiation. Several methods have however been specifically designed for distal radius, tibia and calcaneus peripheral imaging. Initial studies focused on the proximal femur, but this location is complex, and trabecular bone architecture visualization remains limited. Spatial resolution with images from MR also includes the spectrum of trabecular dimensions (in-plane resolution, 0.15–0.3 mm^2; section thickness, 0.3–1 mm), which results in considerable volume affect [20].

4.2   Biochemistry Bio-Markers Classification

Through bone remodeling processes bone biomarkers are released, including forming, resorption and controlling agent. The scientific review of osteoporosis diagnosis in the last decade has been very attentive to these bone biomarks. The clinical uses of biomarkers for bone pathology are merged with bone mineral density analysis with detailed diagnostic knowledge. For example, complete alkaline phosphatase (ALP), bone-specific alkaline phosphatase (BALP), (OC), procollagen typology-1 N-terminal-propeptide (P1NP) and procollagen type-1 C-terminal-propeptide (P1CP) are biomarkers of bone forming. Bone Resorption biomarkers include: hyproxyproline (HYP), hydroxylysine (HYL), deoxypyridinoline (DPD), pyridinoline (PYD), bone sialoprotein (BSP), osteopontine, tartrate resistant phosphatase 5b acid (TRAP5b), carboxy-terminal cross-linked

telopeptide of type 1 collagen (CTX-1). The bone turnover regulators are NF-kB lig-and (RANKL), dickkopf-1 (DDK-1), osteoprotegerin (OPG), sclerostin receptor acti-vator. Both biomarkers are helpful when the BMD DXA test fails to provide enough information to make a prediction early assessment of osteoporosis [17, 26].

It indicates the application of DXA and bone biomarker tests with a high potential for enhancing an early identification of individuals with a high risk of osteoporosis. Since these biochemical biomolecules are the most widely employed biomolecules for the measurement of bone turnover, the biomarker biosensors for the osteoporosis treatment process can be used to track and to predict medication effectiveness [26]. BTMs showed promising results in not only tracking medication success, but also anticipating the use of pharmaceutical drugs by high-risk groups. Proteins extracted from osteocytes were therefore confirmed to be successful in controlling bone growth at the injury site for injury patients with devices [17].

1- Cell: Advancing bone tissue implantable electrochemical biosensors presents a the-oretical average for in-situ therapeutic diagnosis owing to their capacity for tracking exercise by transfusing a range of chemical and physical markers with a view to provid-ing specific substances. Protein recognition from specific forms of cell approval in situ may function as a guide for certain cellular reactions or bone regeneration. Such biosen-sors have used advantages and disadvantages. The biosensor's diagnostic limit is partly dictated by the normal conditions under which the cell will remain alive for a long time when the physical and chemical parameter of the atmosphere need to be regu-lated. Nonetheless, the cell-based biosensor's key limitation is the durability of the cell according to different conditions such as sterilization, service life, biocompatibility etc. [17].

2- Calcium: Scientists also recognized that light isotopes of calcium are incorporated into the bone from either the blood through bone development while they are emitted into the bloodstream as the bones weaken and this makes the bone loss screening dependent on calcium isotropy. In fact, it is of valuable importance for the diagnosis of bone illnesses such as osteoporosis if a calcium sensing mechanism is defined. A new, hybrid microchip sensor has been designed for this purpose. This microchip sensor marks a new approach to calcium ion measurement throughout an important physiological range; it offers a regenerative scheme for different estimates, and is least affected by conflicting ions inside the trial than most other methods [17].

3- Enzyme: In biosensor, a quartz crystal microbalance (QCM) sensor was used to assess the level or activity of tartrate-resistant acid phosphatase (TRAP) in the blood. The biosensors have taken into account osteoclast function by calculating the distraction situation of a cantilever beam using an optical discovery plate and antibody for TRAP, TRAP 5a or TRAP 5b. Since the breakthrough has rendered the experiment much more robust for a longer time, it stays one of the most successful ways to determine antigen rates in a dynamic study for the sample. Directly, most BTM prediction approaches are measured in terms of the time-consuming, expensive, and requires by qualified professionals, according to enzyme based immune sorbent. The marginally poor quality statement of other elements that hinder medical bone use of BTMs [1].

4- Gene: MiRNAs are a category of small, genetically expressed, non-coding RNA molecules. MiRNAs play a significant role in bone regeneration and are particularly important for biomarkers and future treatment targets in osteoporosis, the latest studies have found [17].

5- Alkaline Phosphatase: For clinics and hospitals, the most commonly employed bone marker are Alkaline Phosphatase (ALP). ALP has physiologic substrates that separate inorganic phosphatases by organic phosphates, improve the output of calcium phosphatases and permit mineralization. For proper bone mineralization, ALP is necessary. Bone ALP (BALP) comprises around 50% Serum ALP and serum lives from 24–48 h. Since the half-life may vary in cardiac rhythm, the maximum rates in the afternoon and during the night may be fairly large. It is not established the specific metabolic route. ALP is calculated using p-nitro phenyl phosphate as substratum with the assistance of a spectrophotometer. The ALP can be isolated by electrophoresis of the bone and liver while this process is time consuming and tests are semi-quantitative. Two antibodies with minor affinity variations to isoforms can be used to calculate BALP concentration [14].

6- Laboratory Techniques: Numerous initial tests for the detection of bone mass loss are valid. Blood and urine test results may be subject to a number of laboratory tests. The most regular blood tests determine [14]:

- Blood calcium levels
- Blood vitamin D levels
- Thyroid function
- Parathyroid hormone levels
- Estradiol levels to measure estrogen (in women)
- Follicle Stimulating Hormone (FSH) test to establish menopause status
- Testosterone levels (in men)
- Osteocalcin levels to measure bone formation.

4.3   Invasive Techniques

The procedure of penetration entails many hazards such as pinching the scalp. The instrument has a small probability that the bone is broken or that one of the muscles, arteries and kidneys in the area of the biopsy site is damaged. Another procedure may be required to treat the issue if problems arise. The bone will become weakened, osteomyelitis, or not healed properly after a bone biopsy. In rare cases, a fragment may become unstable and crack in the bone from which the biopsy sample was taken later such a treatment is useful except in extreme cases.

1- Needle Bone Biopsy: This is not very popular bone density measurement procedure, but is selective and is better used as a research method for evaluating bone disease care schemes. Double tetracycline marking for evaluating appositional bone growth combines the best scientific use of bone biopsy to remove osteomalacia. Weeks apart for the doses of Tetracycline, the bone biopsy is wrapped in a plastic paste, thinly sliced, and tested under fluorescent light, and reveals tetracycline lines (which auto fluoresce) [14].

2- Bone Fracture Diagnosis Micro-sensor: A micro-pump is the type of osteoporosis treatment or examination. To order to move the piston to two ways, the micro-pump is constructed using the electromagnetic theory. To transfer the fluid with the mixing tool a closed loop system is used. A form of pump is useful to collect blood or to take medication. There, the oscillating micro-pump is used to test the mechanical response of the osteoporosis cell. An implantable MEMS bone sensor focused on telemetry has been suggested by another study [14], the bone tension assessment capacity by way of the wireless RF system. The tension of the bone is identified using the incorporated piezo-resistive pressure gauges with polysilicon and the CMOS module.

For in vitro or in vivo diagnosis of osteoporosis, another version of a micro-fabricated strain gauge system is being used. These forms of micro-sensors provide an analysis of the scattered stress data on bone surfaces to track bone structural completeness. The wireless and implantable membranes of this form are combined in a compact membrane. An assay with this micro-strain gage for the study of osteoporotic bone was performed to [15]. A piezoelectric BioMEMS bone sensor was used. To order to detect tension in the body, an attempt has been made to create piezoelectric bone sensors. Another bone surface tension calculation micro-scale sensor has been addressed. It sensor form is used to research osteoporosis systemic effects. Development and simulation use thin-film metal strain gauges finite element design method ANSYS. PDMS encapsulated metal films is tested for electrical interconnection. The PDMS membrane was developed to simplify embedding projects. Basic processing processes such as salinization, PDMS planning, photolithography, PDMS metallization. Through tests, instruments such as mechanical testing, electrochemical testing and stick testing were refined and described and a new design was addressed [15].

3- BioMEMS Chip: Amodern BioMEMS device for diagnosing Osteoproteogerin (OPG), focused on golden nanoparticles. A biochip is used to accurately assess the bone disorders that are directly linked to osteoporosis treatment and prognosis. Fluorescence caused by laser (LIF) micro-scale is used to imagine blend capabilities. The treatment of the BioMEMS chip is focused on efficient immunoassay. On the AuNPs embedded conducting polymer, the OPG monoclonal antibody is immobilized with covalent bonding with a carboxylic acid group. Amperometrically, the catalytic reduction is monitoring at -0.4 V compared with Ag/AgCl. For 2 ng/ml diagnostic limit, the linear aperture is 2 to 24 ng/ml [11, 14].

4- BioMEMS-Based Sensors: BioMEMS, utilizing electrochemical microfluidic immunoassay (Heineman et al. 2001), assist the heterogeneous immunoassay in the blood sample study. In this work, there are two definitions in immunoassay. First of all, it incorporates the reaction and diagnostic chamber within the electromagnetic field, according to the analog micro-capillary immune reactor. The device for alkaline phosphatase testing is both MEMS based. From the point of care unit, a separate MEMS micro valve has been suggested with PDMS diaphragm and two tubes with a thermos pneumatic valve for the integrated silicon blood control scheme. The MEMS technology will blood test device to affordable rates [14]. The long-stroke micro-valve has been produced by two thermal-pneumatic tube actuators.

4.4 Osteoporosis Biosensors Classification

The assessment of bone mineral density (BMD) and Dual-Energy X-ray (DEX) was widely used in osteoporosis diagnostics. Such processes, though, have certain limitations. Engineering investments and bone loss management take at least two or three years. It shows how important and appropriate it is to test bone metabolism instantaneously. Recently, advancement in bio metabolism research can be used to measure the bone metabolism. Both the cellular and trabecular bone morphology, two characteristics considered to be important for determining overall bone strength, could not be sufficient to determine the intensity alone of BMD. Although biosensors are monitoring devices that can efficiently evaluate low-cost materials comparison to another process, biosensors to use with osteoporosis can be produced at a low cost, quick, accurate and high-precision pace [26].

A state-of-the-art data was collected using a self-assembled monolayers (SAM) method for the reliable, responsive and functional calculation of octane concentration [26]. Conventional devices of the EIS involve immune-sensor procedures that use immobilized antibodies to evaluate protein-protein relations in complex biological situations or identify smaller molecules such as antibiotics [26]. Recent improvements in bone biosensors have allowed biomechanical value assessments to be performed as the metabolic location of the bone [16]. Although the use of these methods is extremely sensitive and selective, the high cost of materials and the labor engaged is extremely restrictive. Preventing early stage bone loss depends greatly on bone responsiveness to structural signals and is critical for the identification, preservation and regeneration of bone mass. In the analysis of physiological consequences of osteoporosis, evaluating the in-vivo bone surface tension is essential. Therefore, many biosensors were proposed in this portion. Bone biosensors can be graded as follows [10]:

1- Electromechanical Biosensors: Stress is an essential part of determining the bone shape in this procedure, including that it is more evident in cases of overburden negligence [1]. The significant benefit of encouraging bone sensors to determine the bone development and consistency of the human body is that they demonstrate what's occurring in the skeleton. Preventing advanced stage bone collapse relies immensely on the effect the bone has on structural signals, which are critical in the identification, preservation and retrieval of bone mass. Evaluating in-vivo bone plate tension will be important to take the side effects of osteoporosis into account. A variety of electromechanical biosensors were therefore introduced. Electrochemical biosensors typically calculate the oxidation-and decrease-reactive current generated. This generated current may be associated with either the accumulation or the level of output/production of the existing electro-active material. The resultant electrical signal is connected with the aim and analyst cycle and proportionate to the research stage [17].

2- Electrical Biosensors: Most of the research was based on using electrical sensors to convert a mechanical movement into an electrical signal made up of pressure gauges, piezoelectric sensors and piezo-resistance sensors, so that minor functional deformities can be detected. This consists of compact frequencies and compares damping ratios

calculated from Impedance Function (FRF) to test structural and specific structural char-
acteristics of the bones. As shown in Fig. 5, it was claimed that the technique could be
used for monitoring the heating process of bones after surgery [3].
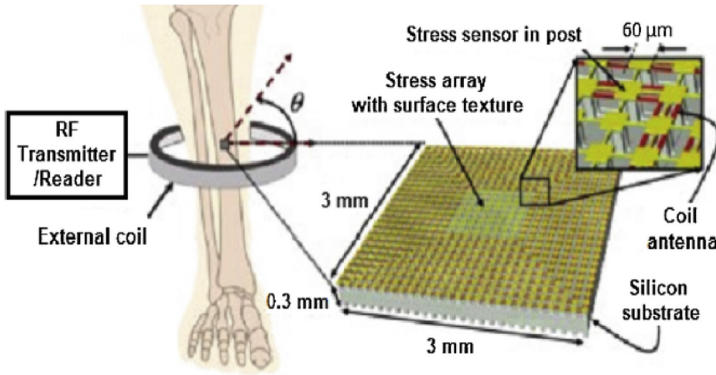
Fig. 5. Application of piezoelectric biosensor in osteoporosis diagnosis [3].

3- Photometric Biosensors: They have the benefit and don't cause thrombosis of being
inert chemically. These are also light adaptable and resistant to EMI/RFI. With the biosen-
sor adjustment to the bone strain calculation, the initiation of osteoporosis can be more
accurately differentiated. According to the micro bend, excess communication failure
from the Optical fiber is induced by anti-bend micro bend sensors to detect/measure
displacement, weight, pressure, temperature etc. [17].

4- Acoustic Biosensors: The non-invasive method used to track the likelihood of bone
breaks could be acoustic emissions (AE). The AE parameter values are aligned with
the assisted set, with the degree of bone damage. The amount of reported exercises.
This technique will discern the extent of damage to the bones by osteoporosis in three
steps: (i) using a skin-near-bone acoustic sensor, (ii) using the tension to the bone in an
AE instance, (iii) detecting AE with an acoustic sensor to provide a sound signal and
evaluate the sound signal range to determine the level of damage to osteoporosis of the
bone [17].

5   Bone Turnover Markers (BTMs)

The Bone Metabolism molecular markers found by bone renovation are also a therapeutic
tool for osteoporosis treatment, which is named Bone Turnover Markers (BTMs) and is
time-consuming, costly and needs of skilled specialists. There have been recent studies
in which identification BTM biosensors have been established. Creative advances have
greatly enhanced the efficiency of initiatives producing strong, rapid and non-invasive
biosensors with improved impacts and precision that can be predictive of advancement

towards point of care by unavailable, real-time monitoring of health [20]. BTMs may, however, be helpful in measuring bone turnover by determining bone cell count and behavior. BTMs may predict the likelihood of osteoporosis or of getting fractures or possible treatment reaction, particularly with regard to medical usefulness. Biochemical bone turnover markers were therefore tested for three main aims [22]:

- Anabolic or anti-resorption medication control of osteoporosis cases
- The potential loss of bone and the possibility of osteoporosis
- Better breaks risk assessment of various cases

5.1 Advantages of Using BTMs

DXA is the commonly known, non-invasive and approved BMD tool for the diagnosis and likely risk-based detection of persons with osteoporosis. On the other hand, although BMD improvements are deemed a legitimate end of the osteoporosis clinical monitoring, the use of DXA in early steps cannot be evaluated for BMD adjustments and fracture risk reduction. DXA is also not able to assess bones and its tests are impaired by age change and bone disorders such as osteoarthritis [23]. However, BMD alone does not diagnose additional osteoporosis cases, which is triggered by some medical problems influencing the disorder or medications that may conflict with the accomplishment of optimal bone mass or contribute to bone loss. In these cases, the treatment process may include the root cause of the disease, which is important for recognizing body biology through testing and recasting processes for order to predict the likelihood of fracturing and track the bone reaction to care, according to additional tests like Bone Turnover Marker. However, despite the penetration and repeatability of bone biopsy in clinical research, and the scarcity of specialist labs indicates that bone histomorphometry has gained momentum care as [24]:

- Blood and urine tests are easily taken
- They are relatively particular for bone recovery or structure
- They supply BMD with additional information
- Modifications in BTMs arise until BMD.

BTMs are capable of identifying patients with anti-resorption medications that profit more. Research showed that BTM-values for women are 75–80% at a risk of accelerated bone loss in at least two standard deviations [24].

5.2 Disadvantage of Using BTMs

The new recommendations addressed multiple roles in the treatment and control of osteoporosis with BTM, which reported different views about the effectiveness of BTM. Nevertheless, in clinical research, the regular use of BTMs is not advised, as BTM principles do not depend solely on bone retention, but on other factors such as age or gender. In fact, the relationship among BTMs and the bone loss level depends on the precision of the calculating methods greatly. Analytical variation, further addressed here, is a major limitation to the daily use of BTMs in clinical research [25].

6 Proposed Simulations and Experimental Results

6.1 Bone Stress Properties in Osteoporosis

The function of stress and bone motions in osteoporosis is described in this section and the implications of the bone shape for its impact on the mechanical strength of the bone are addressed. Equation (2) to Eq. (4) define the 3-dimensional stress-strain (ε) in terms of displacement (u) [27]:

$$\varepsilon_{xy} = \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) \tag{2}$$

$$\varepsilon_{xz} = \frac{1}{2}\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right) \tag{3}$$

$$\varepsilon_{xy} = \frac{1}{2}\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_y}{\partial y}\right) \tag{4}$$

Here, the authors simulate the bone structure as a cylinder with a fixed constraint terminal side and an x-axis boundary load on side surfaces. The simulations and analysis through COMSOL are driven based on specific properties given in Table 2.

Table 2. The utilized values for current simulations.

Parameter	C_p	ρ	k	E	ν
Description	Heat capacity at constant pressure	Density	Thermal conductivity	Young's modulus	Poisson's ratio
Values	$1313 \frac{J}{KgK}$	$1908 \frac{Kg}{m^3}$	$0.32 \frac{W}{mK}$	20.7 GPa	0.29

Figure 6 illustrates and compares the stress and displacement of a cylindrical bone with an applied external force to side surfaces. Based on Fig. 6(a), the stress through a cylindrical bone is not monotone, and in osteoporosis cases the fracture possibility on red pars is more. Also, Fig. 6(b) and Fig. 6(c) shows the displacement affected by applying stress. Cortical bone is deeper in the body skeletal network and can handle the overall tension rather than the trabecular bone, but also is more delicate. Trabecular bone in vitro can cope with the tension of 30%. With strains of only two percent cortical bone fails. Although, the cortical bone's biomechanical activity is rather consistent, the trabecular bone displays a large ranging strength and rigidity. This variance relies mostly on the obvious density of the trabecular bone. In addition to its obvious density, the strength of the cortical and trabecular bone and its anisotropic microstructure depend on the direction of processing. Bone will typically tolerate higher output charges than strain charges and smaller compression charges than frictional charges. Mechanical processing reactions vary greatly among cortical and trabecular bones. For example, cortical bone exhibits a minor carriage ability when stacked with compressive and compression loads beyond its

elastic deformation scope. The muscles engage in bone mass, content and form to assess the mechanical nature and loading response of bones to a defined operating condition. In fact, the association between loading severity and stress results in a degree of tiredness of muscles and physiological weakness that ultimately develops in damage or in use of in the lack of adequate rest and regeneration. The typically strong correlation between severity and recurrence indicates the origin of a cycle of high volumes, medium repeat to low magnitude and high replication loads before structural damage.

Table 3, shows the normalized applied stress of random points based on Fig. 6(a). Based on this table, Mechanical strain stimulates bone forming, regrowth and deterioration methods in the sort of respiration, pressure and gravitational perturbations as a consequence of implementing mechanical stress. The force (outer strength) that causes strain is present in the bone. Implemented forces in general create stresses of different intensities that yield stresses of different sizes. In some sections the pain is the greatest, and in instances of osteoporosis these segments are the most fragile bone tissues. The equilibrium of endothelial dysfunction and development is unfavorable also in sensitive parts, with aging and the environment of osteoporosis. Bone loss is a result of dysfunction and premature bone re-modelling in aging and osteoporosis cases.
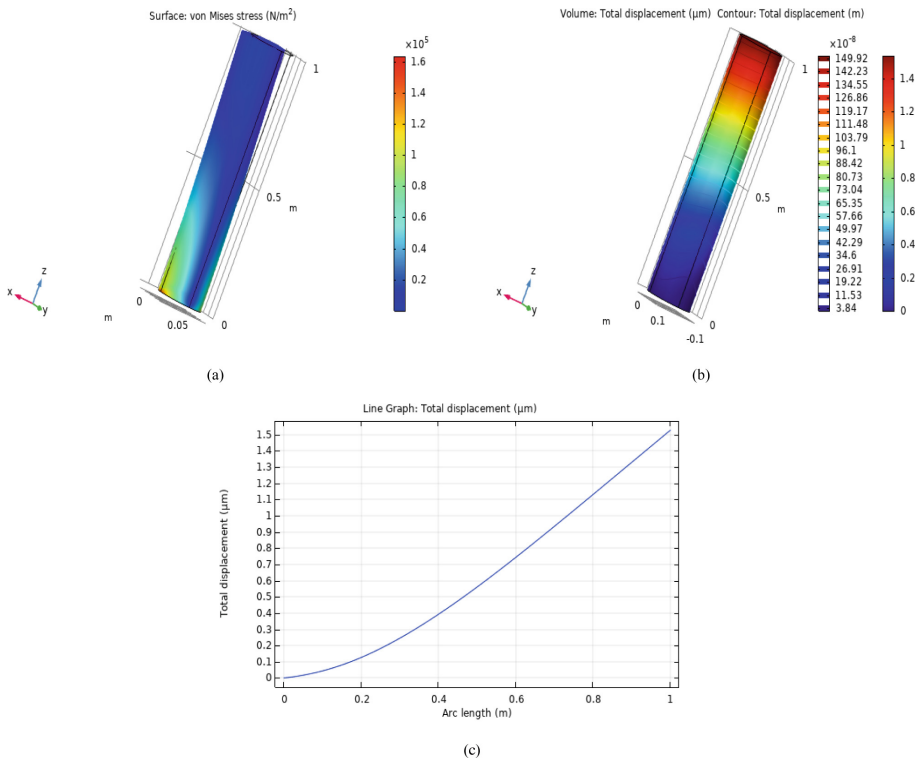
Fig. 6. The proposed simulation of mechanical properties of the cylindrical bone (a) Stress pattern, (b) Contour plot of total displacement, and (c) Line plot of total displacement.

Table 3. Proposed normalized applied stress of random points based on Fig. 6(a).

Random point	X-coordinate	Y-coordinate	Z-coordinate	Normalized applied stress
1	0.006	0.095	0.820	0.248
2	0.097	0.044	0.322	0.490
3	0.101	0.014	0.156	0.838
4	0.101	0.004	0.105	0.944
5	−0.079	−0.016	0.00	0.577

7   Conclusion and Future Works

Osteoporosis, which is regarded as a "Silent Killer", becomes the dominant social dis-order of bone deformation. Osteoporosis is a condition that reduces the consistency, amount and cracking of the bones. To order to identify the burden of living bones or to recognize specific proteins (biomarkers or molecules containing signals within or beyond bone cells), an early diagnosis of pathology may be very useful, to addition to the existing medical pathology procedure. This research aimed at providing input into emerging state of the art developments in the category of osteoporosis therapies, underlining the need to support new approaches to transfer this important research into clinical research. Bone densitometry and other noninvasive imaging techniques that are commonly used for the quantitative and qualitative evaluation of bone teach us little regarding improvements in the structure of bone content and biomechanical architecture during the healing process. Most methods in osteoporosis studies and their operating cri-teria are accessible. Nevertheless, such methods have certain limitations. It's very costly to check such machines, early diagnosis isn't straightforward, they are intrusive tests, magnifies mistakes are likely, exactness cannot be accomplished, the testing period is large, harmful radiation can trigger body complications, home care cannot be performed and the instruments are not portable. Those sophisticated devices must in fact be oper-ated by qualified people. While calculating Bone Mineral Density (BMD) is the most precise diagnostic tool in osteoporosis, it cannot benefit and it is quite expensive in early diagnosis.

As the biomedical engineering science and mechanics, motors, electronics, informa-tion technology, pharmaceutical production and medicines become a major part of many sectors, medical diagnosis can quickly move into the future as physicists, engineers and physicians will create intelligent biosensors for different conditions inside the body. Osteoporotic injuries can be avoided through the use of pharmacotherapy. Nonetheless, it is important to implement such therapies with clear guidelines because expensive and side-effects, such as atypical sub-trochanteric fractures and alendronate, are correlated with these therapies. Ideally, biomarkers for assessment should be accessible. In order to control the beneficial effect, such as evaluation of reactions and non-response to these treatments, diagnostic methods should also be used. The restrictions of calculation of bone mass/density in these two sectors have caused new imaging biomarkers that focus on bone quality to expand. In order to diagnose common osteoporotic brakes, the nor-mal methods of imaging must be used in relation to those devoted to bone mass and

quality assessment because they impact therapeutic suggestions and can prevent future brakes. The first finding is critical for the human race with osteoporosis as a subtle and deadly disease. Current, compact, cheap diagnostic devices for point-of-care testing are therefore expected to be built. Finally, the importance of osteoporosis, stress and bone displacement is simulated and it has been shown that the bone components and bone framework collaboratively grant strength on the skeleton without being regarded as the appropriate replacement measure for isolation by morphological characteristics. The material and structural elements of skeletal tissue should be quantified, quantified, and, where attainable, examined by clinicians, researchers and physical therapists who wish to check or monitor or develop bones and the strength and mechanical integrity.

In future, it would be very important for authors to develop and extend inexpensive and reliable bone imaging to analyze fragile fractures with all the accessible imaging modalities in view of the good profit for earlier detection, diagnosis and treatment of diseases related to pathology.

References

1. Khashayar, P., et al.: Bone biosensors: knowing the present and predicting the future. J. Microengineering (2016)
2. Shetty, S., et al.: Bone turnover markers: emerging tool in the management of osteoporosis. J. Metab. **20**, 846 (2016)
3. Afsarimanesh, N., et al.: Smart sensing system for the prognostic monitoring of bone health. Sensors **16**(7), 976 (2016)
4. Bhalinge, P., et al., "Biosensors: nanotools of detection-a review. Res. J. (2016)
5. Wen, Y.-H., et al.: Mechanically robust micro-fabricated strain gauges for use on bones (2016)
6. Pang, C., et al.: A flexible and highly sensitive strain-gauge sensor using reversible interlocking of nanofibres. Nat. Mater. **11**, 795–801 (2012)
7. Fresvig, T., et al.: Fiber optic Bragg grating sensors: an alternative method to strain gauges for measuring deformation in bone. Med. Eng. Phys. **30**, 104–108 (2008)
8. Aggelis, D., et al.: Fracture of human femur tissue monitored by acoustic emission sensors. Sensors **15**, 5803–5819 (2015)
9. Bishitz, Y., et al.: Noncontact optical sensor for bone fracture diagnostics. Biomed. Opt. Express **6**, 651–657 (2015)
10. Afsarimanesh, N., et al.: Smart sensing system for early detection of bone loss: current status and future possibilities. J. Sens. Actuator Netw. **7**(1), 10 (2018)
11. Yun, Y.-H., et al.: A label-free electronic biosensor for detection of bone turnover markers. Sensors **9**(10), 7957–7969 (2009)
12. Sirivisoot, S., et al.: Multiwalled carbon nanotubes enhance electrochemical properties of titanium to determine in situ bone formation. Nanotechnology **19**, 295101 (2008)
13. Afsarimanesh, N., et al.: Sensors and instrumentation towards early detection of osteoporosis (2018)
14. Singh, K., et al.: Osteoporosis: new biomedical engineering aspects. J. Mech. Sci. Technol. **20**, 2265–2283 (2006)
15. Yang, G.Y., et al.: Design of microfabricated strain gauge array to monitor bone deformation, in vitro and in vivo. In: Proceedings of the 4th IEEE Symposium on Bioinformatics and Bioengineering (2004)
16. Afsarimanesh, N., et al.: Sensing technologies for monitoring of bone-health: a review. Sens. Actuators A Phys. **274**, 165–178 (2018)

17. Liu, L., et al.: In situ sensor advancements for osteoporosis prevention, diagnosis, and treatment. Curr. Osteoporos. Rep. **14**, 386–395 (2016)
18. Alejandro, P., et al.: A review of osteoporosis in the older adult: an update. Rheum. Dis. Clin. **44**(3), 437–451 (2018)
19. Sartoris, D.J.: Osteoporosis Diagnosis and treatment. Marcel Dekker Inc., New York (1996). ISBN 9780824795078
20. Sheu, A., et al.: Diagnostic tests: bone mineral density: testing for osteoporosis. Aust. Prescr. **39**, 35 (2016)
21. Langton, C.M., et al.: The measurement of broadband ultrasonic attenuation in cancellous bone. Eng. Med. **13**(2), 89–91 (1984)
22. Lenora, J., et al.: Use of bone turnover markers in osteoporosis. Clin. Rev. Bone Miner. Metab. **8**(1), 1–14 (2010)
23. Cefalu, C.A.: Is bone mineral density predictive of fracture risk reduction? Curr. Med. Res. Opin. **20**(3), 341–349 (2004)
24. Harper, K.D., et al.: Secondary osteoporosis. Diagnostic considerations. Endocrinol. Metab. Clin. North Am. **27**(2), 325–348 (1998)
25. Wheater, G., et al.: The clinical utility of bone marker measurements in osteoporosis. J. Transl. Med. **11**, 201 (2013)
26. Inal Kabalaa, S., Yagarb, H., Ozcan, H.M.: A new biosensor for osteoporosis detection. J. Prep. Biochem. Biotechnol. **49**(5), 511–520 (2019)
27. Hills, D.A., et al.: Introduction to Fracture Mechanics. Solution of Crack Problems, vol. 44, pp. 1–27. Springer (1996)

Segment Medical Image Using U-Net Combining Recurrent Residuals and Attention

Yuan Wang[1], Zhiyou He[2], Peizhen Xie[1], Canqun Yang[1], Yu Zhang[3,4,5],
Fangfang Li[3,4,5], Xiang Chen[3,4,5], Kai Lu[1], Tao Li[1], Jiao Zhou[1], and Ke Zuo[1(✉)]

[1] National University of Defense Technology, Changsha 410073, Hunan, China
zuoke@nudt.edu.cn
[2] The Department of Burns and Reconstructive Surgery, Xiangya Hospital, Central South
University, Changsha 410008, Hunan, China
[3] The Department of Dermatology, Xiangya Hospital, Central Source University,
Changsha 410008, Hunan, China
[4] Hunan Key Laboratory of Skin Cancer and Psoriasis, Changsha, Hunan, China
[5] Hunan Engineering Research Center of Skin Health and Disease, Changsha, Hunan, China

Abstract. Medical image segmentation is the key to decide the issue of medical
images in clinical practice that can provide a reliable basis. The development
of medical image segmentation technology not only affects the development of
other related technologies in medical image processing, such as visualization 3D
reconstruction, but in the analysis of biomedical images also occupies an extremely
important position. With the application of deep learning algorithms in medical
image segmentation, medical image segmentation technology has made significant
progress. In this paper, we discuss the segmentation method of 2D medical images
about U-net variant network. Use the U-net combing recurrent residual model and
attention model to segmented the image can get better result.

Keywords: Segmentation · Medical image · U-net · Residual · Attention

1 Introduction

Biomedical imaging has become an indispensable part of disease diagnosis and treat-
ment, and is becoming increasingly important. Cone-beam CT, 3D Ultrasound Imaging,
Single Photon Radiation Tomography (SPECT) and other medical imaging technolo-
gies have been widely used in clinical examination, diagnosis, treatment and decision-
making. Image segmentation is the specific image into a plurality of areas with unique
properties and make the object of interest and technical process [1]. Medical image seg-
mentation is to divide 2D or 3D medical images into different regions and extract the
regions of interest. Medical image segmentation technology is an indispensable method
for extracting quantitative information of special tissues, and it is widely used in vari-
ous occasions, such as quantitative analysis and diagnosis of tissue volume, location of

---

Y. Wang—Fist Author.

diseased tissue, learning of anatomical structure, treatment planning, functional imaging data local correction and computer-guided surgery [2]. In fact, many researchers have proposed many medical image segmentation methods, but because medical images are inherently more complicated than natural images, there are still many segmentation issues waiting to be solved in medicine.

Since 2006, deep learning has achieved outstanding achievements in various fields in computer vision tasks that use natural images as the analysis and processing object, but the deep learning successful application still has a lot of challenges in medical image analysis. On the one hand, it is manifested that natural images are easier to obtain than medical images, and they have very large-scale learning and training sample data sets. But it is very difficult to collect and obtain large-scale medical data. On the other hand, natural images not only have high spatial resolution and large contrast, but also have many natural visual features that can be easily identified. However, medical images have only a certain kind of special signal strength value, and usually have a relatively low signal-to-noise ratio. These increasing the difficulty of medical image segmentation.

Medical image segmentation plays an important role in image processing and analysis. It is also a core component of other advanced medical image analysis and interpretation systems. For medical image segmentation, it can be divided into three methods: manual segmentation, semi-automatic segmentation, and fully automatic segmentation [3]. Manual segmentation refers to the delineation of the boundaries of tissues directly on the original image by experienced clinicians, or the drawing of regions of interest on a computer through an image editor. This method has high accuracy but is too laborious and time-consuming; The semi-automatic segmentation technology combines the powerful computer data processing, algorithm analysis and storage capabilities with the knowledge and experience of medical experts to complete image segmentation through human-computer interaction; Automatic segmentation means that the image is completely segmented by computer without manual intervention, mainly by using fuzzy theory and neural network. With the wide application of deep learning in various fields, the segmentation methods developed on the traditional segmentation algorithm have achieved remarkable achievements in the image segmentation area, and its segmentation accuracy has exceeded that of the traditional segmentation algorithm.

2   Related Work

In this section, we will briefly introduce the development of deep learning, the segmentation of deep learning in medical images, and the selection of models after comprehensive comparison.

2.1   Deep Learning

Deep learning is an important part of machine learning, which evolved on the basis of artificial neural networks. From the 1940s to the 1960s, with the implementation of biological theory and the first model perceptron [4], a single neuron can be trained. But a single-layer perceptron cannot solve the problem of linear inseparability. Because of this model not being extended to multi-layer neural networks, people's research on neural

networks has stagnated for nearly 20 years. Until 1986, the BP algorithm [5] and its application in neural networks perfectly solved the non-linear classification problem, and the artificial neural network once again attracted widespread attention. However, due to the limited level of computer hardware at that time, the development of the BP algorithm was greatly restricted, so the artificial neural network has once again entered the bottleneck period. In 2006, Hinton and his student Salakhutdinov [6] formally introduced the concept of deep learning, which caused a great response in the academic circle, and deep learning once again returned to people's vision. In 2012, in the well-known ImageNet image recognition competition, AlexNet—a deep learning model [7], fundamentally solved the problem of vanishing gradients, and used GPUs to greatly improve the running speed. With the continuous improvement of computer capabilities and the continuous increase of available data, deep learning has continuously made major breakthroughs. Now both in academia and industry are doing research on deep learning. Deep learning has been widely used in speech recognition [8, 9], Image retrieval [10–12], natural language processing [13, 14], face recognition [15–17] and other challenges, achieved extraordinary results.

2.2   Medical Segmentation Based on Deep Learning

Medical image segmentation is a key task to identify the internal voxel and external contour of the target area of interest in the image. The segmentation of deep learning in medical images can be divided into two categories, one of which is the segmentation of organs and tissues. The purpose is to quantify the clinical parameters related to the volume or shape of the gaps, so as to further study the organs and tissues. Kumar et al. used block-based CNN for nuclear segmentation of H&E staining of pathological images [18]; In the MICCAI 2018 digital pathology challenge for brain tumors, the total survival rate was predicted by segmentation of brain tumors from structural magnetic resonance imaging and extraction of fiber bundle tracking characteristics [19]; Qaiser et al. used CNN to extract image block features and construct continuous homology distribution based on topological features for H&E staining of pathological images for automatic tumor segmentation [20]. The other is the segmentation of the lesion and tumor, which is the accurate segmentation of the lesion or tumor before the treatment of the lesion and tumor, to ensure the killing of tumor cells while protecting the normal organs during the treatment. For example, Yu et al. combined the characteristics of residual network and full convolutional network to construct the Fully Convolutional Residual Network (FCRN) to automatically segment melanoma in dermoscopy, and won the second place in the ISBI2016 challenge [21].

2.3   Segmentation Research Based on U-Net

In 2015, Ronneberger et al. [22] proposed the U-net network structure for the first time. U-net is a semantic segmentation network developed on the basis of FCN, which performs well on medical segmentation with sparse datasets. After the U-net network presented, outstanding performance in the field of medical image segmentation. It caught many studies using U-net network architecture to do medical image research, and suggest improvements on the basis of U-net network structure. Cicek et al. [23] proposed 3D

structure on basis of U-net, which used continuous 2D slice sequences of 3D images to achieve segmentation. Milletari et al. [24] proposed V-net, a 3D deformation structure of U-net. The V-net structure designed loss function based on the image characteristics, and uses 3D convolution to check the image for convolution, reducing the channel dimension through the convolution kernel of $1 \times 1 \times 1$. Drozdzal et al. [25] proposed that in the U-net long hop connection structure and short hop connection structure can be used. In the task of lesion segmentation, deep learning algorithm needs to complete multiple tasks such as target recognition organ segmentation and tissue segmentation, so the global and local information of the image should be combined in the segmentation process to achieve the accurate segmentation of the lesion. Both Kamnitsas et al. [26] and Ghafoorian et al. [27] make a point that the use of multi-scale convolution to extract global and local image information. In addition, Wang et al. [28] proposed the use of U-net network to segment the wound image, and then the trained SVM classifier was used to classify the segmented wound image to determine whether the wound was infected. Finally, the GP regression algorithm was used to predict the wound healing time. Brosch et al. [29] using the U-net network on the white matter lesions on brain MRI segmentation, and added a jump connection structure between the first layer convolution and the last layer deconvolution of U-net, which made the network structure obtain good segmentation results even with less training data.

3   Method

3.1   U-Net Module

U-net is often used for image segmentation tasks. It performs well in medical segmentation and can use very few training images to obtain very accurate segmentation results. U-net is developed from the network architecture of FCN, and many feature channels
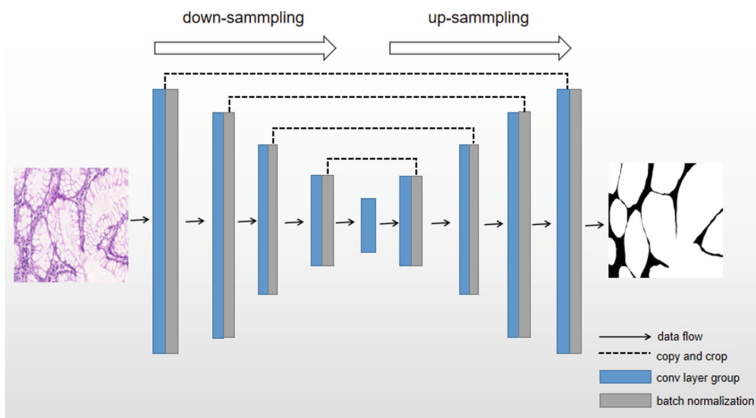
Fig. 1. U-net structure that use to segmentation the medical image. Input the medical pictures into the u-net network, extract the feature information through the down-sampling step by step, and integrate the feature information directly with the up-sampling to achieve the purpose of segmentation.

are added to allows more details of the original image is spread in layers in different resolutions. There is no full connect layer in this network, and the whole process uses valid convolution to ensure that the segmentation results are obtained based on no missing context features. In addition, enhanced the ability to predict edge portions of the input image, the lost context information is extrapolated by mirroring the input image. Finally, background between important information is given higher weight. The U-net module is shown in Fig. 1. U-net network can be seen simply as the down-sampling, through different degrees of convolution, to learn the deep characteristics. Then the size of the original image is recovered by up-sampling, which is realized by deconvolution. Finally, the softmax function is used to activate the pixels, so that the pixels become a probability map.

3.2   Recurrent Residuals Module

Residual networks show excellent performance in computer vision problems. Residual neural network constructs cross-layer superposition between convolutional layers through shortcut, which provides more abundant features for deep network and makes the overall structure of network more optimized and diversified. Recurrent network has a specific memory pattern. Its input comes from multiple layers. Not only the input information of the current layer, but also includes the information the network sensed at the previous moment. The decision of the recursive network in the previous time step will affect its decision in the subsequent time step. So the recurrent network has two inputs from the present and the recent past. The recurrent residuals module is shown in Fig. 2. We can see the difference between this and a traditional module.
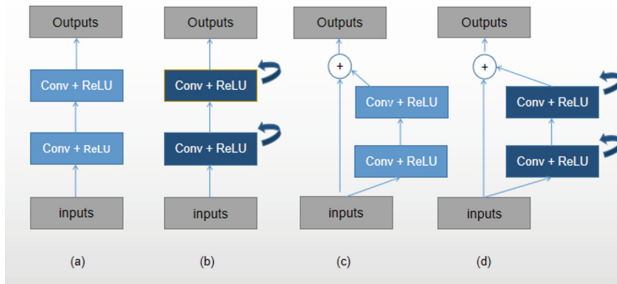
Fig. 2. Different variant of convolutional and recurrent convolutional modules applied in deep networks. (a) Forward convolutional module, (b) Recurrent convolutional module, (c) Residual convolutional module, and (d) Recurrent Residual convolutional module.

3.3   Attention Units

The Attention mechanism mimics the internal process of biological observation behavior and is a mechanism to align the internal experience with the external sensation to increase the observation accuracy of some areas. After scanning it through the global information

obtained need to focus on the target area, we put more attention to this area resources to get the details of an object of interest, and to inhibit other useless information. According to different application scenarios, attention can be divided into spatial attention and temporal attention, while spatial attention is used for image processing. The Attention mechanism is initially referred to in the Seq2seq model to break the restriction of the original code model on the fixed vector [30]. The principle of the attention mechanism is to compare the degree of matching between the input sequence and the output vector. The higher the matching degree is, the higher the relative score of Attention concentration is, and the matching weight obtained by Attention calculation is only limited to the current sequence pair. The attention module in this paper is used in the skip connection. The attention units is shown in Fig. 3. The Attention module is used to process the feature graph of the same layer of the lower sampling and the next layer of the upper sampling layer, and then connect the feature graph.
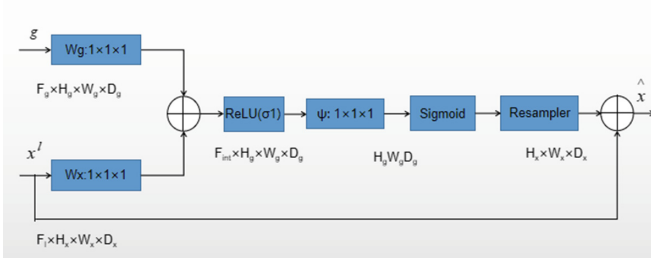
Fig. 3. Attention module structure that use to segmentation the medical image. g is the feature of the same layer of down-sampling layer, and x is the feature of the upper layer of the up-sampling layer. After the obtained feature graph is activated by the ReLU activation function, the convolution operation of $1 \times 1 \times 1$ is carried out. Finally, sigmoid is activated and the final attention result is obtained by resampler.

4   Experiment

4.1   Implementation Details

This experiment was conducted under the Pytorch 1.3.0 Python 2.8.0 cuda 4.5.8 in the Linux system. The process framework of this paper is shown in Fig. 4. We randomly make them into training sets, test sets, and validation sets. Through the training of the U-net, the segmentation of image will be finally obtained, and then get the correct rate of the test set. In this experiment, we combined U-net with recurrent residuals module and attention model to carry out experimental verification. Experiments show that the combination of recurrent residuals module and attention module will improve the segmentation accuracy with less data sets.

4.2   Evaluation Metric

The evaluation criteria of this experiment were using accuracy, Recall (SE), Precision (PC), F1-score, Jaccard similarity (JS) and Dice similarity coefficient (DC). We used the
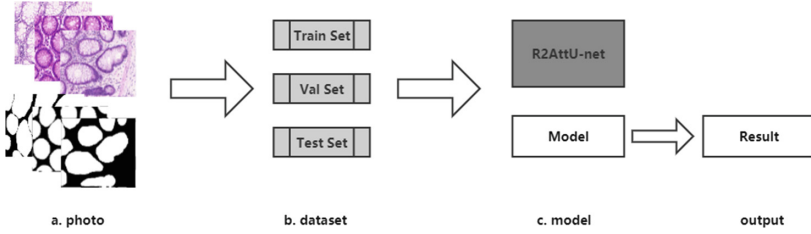
Fig. 4. Overview of the framework. (a) photo: choose pictures that need segmentation and make the grout truth to relevant photos (b) dataset: the photos are divided into training sets, verification sets and test sets; (c) training: use the dataset to training the network to adjust the parameters and save the model after training; (d) output: get the result of segmentation.

TP, FN, FP, TN, GT, SR respectively to represent the true positive, false negative, false positive, true negative, ground truth, segmentation result. The criteria can be defined as:

$$accuarcy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$SE = \frac{TP}{TP + FN} \tag{2}$$

$$PC = \frac{TP}{TP + FP} \tag{3}$$

$$F1 = 2 \times \frac{PC \times SE}{PC + SE} \tag{4}$$

$$JS = \frac{|GT \cap SR|}{|GT \cup SR|} \tag{5}$$

$$DC = \frac{2 \times |GT \cap SR|}{|GT \cup SR|} \tag{6}$$

Accuracy is the ratio of the number of records used for correct segmentation to the total of records, and F1_score measures P, R at the same time. JS was to measures the similarity between two sets. The larger the test result, the more similar the two sets are, or the more the two sets overlap. The DC is very similar to JS that used to measure the similarity of two sets, but the calculation methods are difficult.

4.3 Result

In this experiment, we segmented the medical image based on different U-net. We performed segmentation on the cell dataset, and the experimental results are shown in Table 1. The experiment shows the U-net network is suitable for processing medical image tasks. Form Table 1 we observed that the ordinary U-net network also obtain JS 0.7 about the segmentation, and R2AttU-net can obtain the JS is 0.8. The JS on the basis of original increased by 10%, and all other measures were raised by about 5%. Select

a picture of the data concentration and the image segmentation results of each network are selected to synthesize in Fig. 5. It can be intuitively observed from the Fig. 5 that for adding different modules, the performance of U-net network will be different, and the best performance is the U-net which adds recurrent residuals module and attention model at the same time.

Table 1. The result of cell segmentation

Heading level	U-net	R2U-net	AttU-net	R2AttU-net
Accuracy	0.81	0.85	0.84	0.87
SE	0.87	0.85	0.88	0.92
PC	0.79	0.84	0.82	0.84
F1-score	0.81	0.83	0.83	0.86
JS	0.70	0.74	0.73	0.80
DC	0.81	0.83	0.83	0.86

(a).image   (b).GT   (c).U-net_SR   (d).R2U-net_SR   (e).AttU-net_SR   (f).R2AttU-net_SR
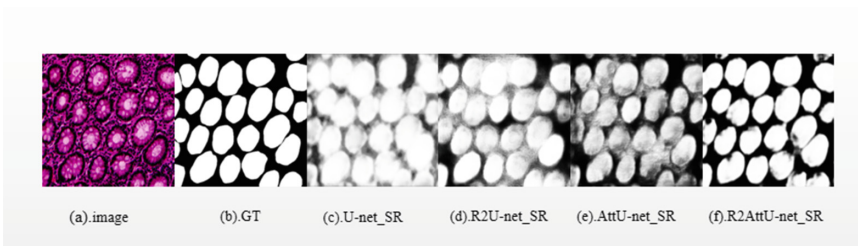
Fig. 5. Results of segmentation. (a) image, the original image of segmentation; (b) GT, and the ground truth for the image, (c) (d) (e) (f) are respectively the segmentation results of U-net, R2U-net_SR, AttU-net_SR and R2AttU-net_SR.

5   Discussion and Conclusions

Deep learning algorithm has important theoretical significance and practical value in medical image processing. There is still much room for development of medical image segmentation through deep learning. Although the research time is not long, it has made some achievements with the continuous improvement of medical image quality and the application of deep learning in medical images has grown rapidly. For pathological pictures, morphologic differences mean differences in the type and severity of the disease. Segmentation is the basis of discrimination, and it is also the most time-consuming and laborious place for doctors to diagnose. Deep learning used in the medical image not only improves efficiency and accuracy, but also can distinguish features that are difficult

for doctors to distinguish from morphology, and even discover new unknown forms, providing strong technical support for discovering new morphological markers.

This paper uses a variant of U-net network to segment medical images which contain recursive residuals module and attention module. Firstly, the data was preprocessed, and then the optimized u-net was used to segmentation. Finally, the segmentation results were compared through relevant evaluation criteria. This method has a certain universality, which proves to some extent that u-net is suitable for medical image segmentation, and can be used to build network model quickly. However, this method is not perfect and it has some shortcoming: during the training process, the network model memory footprint is too large; The experimental results are easily influenced by the number of training sets. Therefore, the neural network will continue to be optimized in the future, and the image deformation method will be used to increasing the number of training set for the model to improve the segmentation accuracy within an acceptable range.

References

1. Haralick, R.M., Shapiro, L.G.: Image segmentation techniques. Comput. Vis. Graph. Image Process. **29**(1), 100–132 (1985)
2. Sharma, N., Aggarwal, L.M.: Automated medical image segmentation techniques. J. Med. Phys./Assoc. Med. Phys. India **35**(1), 3 (2010)
3. Yushkevich, P.A., Gao, Y., Gerig, G.: ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3342–3345. IEEE (2016)
4. Stephen, I.: Perceptron-based learning algorithms. IEEE Trans. Neural Netw. **50**(2), 179 (1990)
5. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986)
6. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
8. Hinton, G., Deng, L., Yu, D., et al.: Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Process. Mag. **29**, 82–97 (2012)
9. Han, K., Yu, D., Tashev, I.: Speech emotion recognition using deep neural network and extreme learning machine. In: Fifteenth Annual Conference of the International Speech Communication Association (2014)
10. Lin, K., Yang, H.F., Hsiao, J.H., et al.: Deep learning of binary hash codes for fast image retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 27–35 (2015)
11. Gordo, A., Almazán, J., Revaud, J., et al.: Deep image retrieval: learning global representations for image search. In: European Conference on Computer Vision, pp. 241–257. Springer (2016)
12. Liu, H., Wang, R., Shan, S., et al.: Deep supervised hashing for fast image retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2064–2072 (2016)
13. Sarikaya, R., Hinton, G.E., Deoras, A.: Application of deep belief networks for natural language understanding. IEEE/ACM Trans. Audio Speech Lang. Process. (TASLP), **22**(4), 778–784 (2014)

14. Young, T., Hazarika, D., Poria, S., et al.: Recent trends in deep learning based natural language processing. IEEE Comput. Intell. Mag. **13**(3), 55–75 (2018)
15. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: BMVC, vol. 1, no. 3, p. 6 (2015)
16. Wen, Y., Zhang, K., Li, Z., et al.: A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision, pp. 499–515. Springer, Cham (2016)
17. Liu, W., Wen, Y., Yu, Z., et al.: SphereFace: deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 212–220 (2017)
18. Kumar, N., Verma, R., Sharma, S., et al.: A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans. Med. Imaging **36**(7), 1550–1560 (2017)
19. Kao, P.Y., Ngo, T., Zhang, A., et al.: Brain tumor segmentation and tractographic feature extraction from structural MR images for overall survival prediction. In: International MICCAI Brainlesion Workshop, pp. 128–141. Springer, Cham (2018)
20. Qaiser, T., Sirinukunwattana, K., Nakane, K., et al.: Persistent homology for fast tumor segmentation in whole slide histology images. Procedia Comput. Sci. **90**, 119–124 (2016)
21. Yu, L., Chen, H., Dou, Q., et al.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging **36**(4), 994–1004 (2016)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer, Cham (2015)
23. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., et al.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 424–432. Springer, Cham (2016)
24. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
25. Drozdzal, M., Vorontsov, E., Chartrand, G., et al.: The importance of skip connections in biomedical image segmentation. In: Deep Learning and Data Labeling for Medical Applications, pp. 179–187. Springer, Cham (2016)
26. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2016)
27. Ghafoorian, M., Karssemeijer, N., Heskes, T., et al.: Non-uniform patch sampling with deep convolutional neural networks for white matter hyperintensity segmentation. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1414–1417. IEEE (2016)
28. Wang, C., Yan, X., Smith, M., et al.: A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2415–2418 IEEE (2015)
29. Brosch, T., Tang, L.Y.W., Yoo, Y., et al.: Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Trans. Med. Imaging **35**(5), 1229–1239 (2016)
30. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in NIPS (2014)

A New Importance-Performance Analysis by Size-Insensitive Integrity-Based Fuzzy C-Means

Shou-Hsiung Cheng(✉)

Graduate Institute of Services and Technology Management, Chienkuo Technology University, Changhua, Taiwan
shcheng@ctu.edu.tw

Abstract. In conventional Importance-Performance Analysis, lots of researchers assume that a given set of attributes is spread across four quadrants. The researcher uses the vertical axis and the horizontal axis on the drawing to optimally divide the quadrants. From the perspective of machine learning, a new Importance-Performance Analysis by Size-Insensitive Integrity-based Fuzzy C-Means is proposed. From the empirical results, the new IPA by Size-Insensitive Integrity-based Fuzzy C-Means is plain and simple, successful, and easy to know.

Keywords: Size-Insensitive Integrity-based Fuzzy C-Means · Importance-Performance Analysis · Tourist friendly destination

1   Introduction

Importance–Performance Analysis (IPA) presumes that a given set of attributes is spread across four quadrants, "Quadrant I (Concentrate Here)", "Quadrant II (Keep Up the Good Work)", "Quadrant III (Low Priority)" and "Quadrant IV (Possible Overkill)". Nevertheless, there are quite a few disadvantages in conventional IPA. Lots of researchers made an endeavor to conquer these shortcomings mentioned above by means of revising traditional IPA [2]. The conventional IPA problem is transformed into a cluster analysis problem, and an IPA based on the fuzzy C-Means clustering is proposed. In order to refine the approach in the conventional IPA, the replacement of the drastic membership of an attribute to a set by a membership degree between 0 and 1 to each of these sets is proposed, that is a smoother transition between sets in the paper [1]. However, there are disadvantages that fuzzy C-Means clustering is not only easily affected by the amount of data in each cluster, but can only be applied to perfectly cut spherical data clusters [3].

In this paper, a new IPA by Size-Insensitive Integrity-based Fuzzy C-Means is proposed. To illustrate the applicability of this new method, traditional IPA [6] and IPA by Size-Insensitive Integrity-based Fuzzy C-Means are applied to dig into the importance and performance of various friendly tourism elements in friendly tourism destinations. With empirical research, all aspects of activities and product spaces, Sun Moon Lake Scenic Spots (SMLSS) are regarding as friendly tourism destination. The tourist destination which is friendly has long been considered a traveler-based concept. The tourists

are thought of as the very spectators at the tourist-friendly destination. IPA [6], the researchers presume that a given set of attributes is spread across four groups combined with the degree of performance and the importance of attributes. It is useful to allocate limited resources in practical applications to further enhance the friendliness of the tourism industry.

The rest of the paper in question is organized as following statements: Sect. 2 talks about the methodology, whereas Sect. 3 offers the short discussion and results. Section 4 shows the conclusions and futuristic practical implement.

2 Methodology

2.1 Conventional IPA Method

Conventional IPA presumes that a given set of attributes is spread across four Quadrant, "Quadrant I", "Quadrant II", "Quadrant III" and "Quadrant IV". However, when determining the intersection module of vertical axis and the horizontal axis, how to determine the intersection is mostly determined by investigator subjectively.

The general IPA algorithm is as follows:

Step 1   Determine the service attributes of the assessed unit through expert interviews.
Step 2   Satisfaction P_{ij} and importance I_{ij} of each evaluation attribute are measured through a questionnaire. P_{ij} stands for the degree of satisfaction of the ith respondent with respect to the jth assessment attribute. I_{ij} represents the degree of importance of the ith respondent to the jth assessment attribute.
Step 3   Calculate the average satisfaction degree of the jth evaluation attribute $\overline{P}_J$ and importance of the jth evaluation attribute $\overline{I}_J$.

$$\overline{P}_J = \frac{1}{n} \sum_{i=1}^{n} P_{ij} \tag{1}$$

$$\overline{I}_J = \frac{1}{n} \sum_{i=1}^{n} I_{ij} \tag{2}$$

Step 4   Compute and count the center coordinates of the IPA matrix $(\overline{P}, \overline{I})$. $\overline{P}$ is the center of the axis of the matrix. $\overline{I}$ is the center of the ordinate of the matrix.

$$\overline{P} = \frac{1}{k} \sum_{i=1}^{k} \overline{P}_J \tag{3}$$

$$\overline{I} = \frac{1}{k} \sum_{i=1}^{k} \overline{I}_J \tag{4}$$

Step 4   The degree of performance is considered horizontal and the degree of importance is considered the vertical coordinate of the matrix.
Step 5   On the basis of its coordinates $(\overline{P}_J, \overline{I}_J)$ each evaluation attribute is distributed on the IPA matrix.

However, a different set of methods may produce considerable differences in the classification results [7]. Conventional IPA can result in ambiguities or contradictions in the classification results. Vaske et al. [8] believed that use of the wrong way may cause segmentation error analysis conclusion and influence decision-making quality.

2.2 Importance-Performance Analysis by Size-Insensitive Integrity-Based Fuzzy C-Means (IPASIIBFCM)

Size-insensitive integrity-based Fuzzy C-Means (siibFCM) is proposed by [4]. Its purpose is to solve the problem that the number of each group of the FCM algorithm must be similar. SiibFCM can solve the problem of unbalanced cluster size and different clusters. SiibFCM uses both the viewpoint of the cluster and the viewpoint of the data points. SiibFCM uses the above two viewpoints, i.e., the proportion of the cluster size and the distribution of the data points inside the cluster, to adjust the fuzzy cutting matrix through the above two viewpoints. A new IPA by Size-Insensitive Integrity-based Fuzzy C-Means (IPASIIBFCM) is proposed. The algorithm of IPASIIBFCM is presented as following statement:

Step 1: Fixed original condition

Let $l = 0$; $x^j = \left(p^j, w^j\right) \in R^2$, $j \in \{1, 2, \ldots, n\}$, $i \in \{1, 2, 3, 4\}$, $j \in \{1, 2, \ldots, n\}$, and put $q_{ij}^{-1} = 0$.

Step 2: Calculate

$$\left(L^1\right)_0 = \left(\left(L_1^1\right)_0, \left(L_2^1\right)_0\right) = \left(\max_{1 \leq j \leq n} p^j, \max_{1 \leq j \leq n} w^j\right), \tag{5}$$

$$\left(L^2\right)_0 = \left(\left(L_1^2\right)_0, \left(L_2^2\right)_0\right) = \left(\max_{1 \leq j \leq n} p^j, \max_{1 \leq j \leq n} w^j\right), \tag{6}$$

$$\left(L^3\right)_0 = \left(\left(L_1^3\right)_0, \left(L_2^3\right)_0\right) = \left(\max_{1 \leq j \leq n} p^j, \max_{1 \leq j \leq n} w^j\right), \tag{7}$$

$$\left(L^4\right)_0 = \left(\left(L_1^4\right)_0, \left(L_2^4\right)_0\right) = \left(\max_{1 \leq j \leq n} p^j, \max_{1 \leq j \leq n} w^j\right), \tag{8}$$

Step 3: Calculate

$$\left(d_{1j}\right)_l = d^2\left(x^j, \left(L^i\right)_l\right) = \left(p^j - \left(L_1^i\right)_l\right)^2 + \left(w^j - \left(L_2^i\right)_l\right)^2, \tag{9}$$

for $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$

Step 4: Calculate

If

$$\left(d_{1j}\right)_l = 0 \vee \left(d_{2j}\right)_l = 0 \vee \left(d_{3j}\right)_l = 0 \vee \left(d_{4j}\right)_l = 0 \text{ then } q_{ij}^l = 1 \tag{10}$$

Else,

$$q_{ij}^l = \frac{w_{ij}}{\dfrac{(d_{ij})_l}{(d_{1j})_l} + \dfrac{(d_{ij})_l}{(d_{2j})_l} + \dfrac{(d_{ij})_l}{(d_{3j})_l} + \dfrac{(d_{ij})_l}{(d_{4j})_l}}, \tag{11}$$

for $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$

where

$$W_{ij} = F_i \cdot P_{ij}^* = \frac{1 - N_i/N}{\max\limits_{m=1,2,3,4}(1 - N_m/N)} e^{(1-I_i^*)P_{ij}}, \tag{12}$$

$$I_i^* = \frac{I_i - \min\limits_{i=1,2,3,4}(I_m)}{\max\limits_{i=1,2,3,4}(I_m) - \min\limits_{i=1,2,3,4}(I_m)}, \tag{13}$$

where I_i is the integrity of a cluster,

$$I_i = \frac{1}{2}(C_i + P_i), \tag{14}$$

where C_i is the compactness of a cluster,

$$C_i = 1 - \sqrt{\frac{1}{|A_i|} \sum_{j \in A_i} \left(\left(\sqrt{d_{ji}} \right)_i - \mu_i \right)^2}, \tag{15}$$

where P_i is the purity of a cluster,

$$P_i = \frac{1}{|A_i|} \sum_{j \in A_i} P_{ij}, \tag{16}$$

with

$$P_{ij} = \frac{abs\left(\sqrt{d_{ji}} - \sqrt{d_{jk}} \right)}{\sqrt{d_{hi}}}, \tag{17}$$

which the kth cluster is the nearest to the ith cluster.

$$\mu_i = \frac{1}{|A_i|} \sum_{j \in A_i} \left(\sqrt{d_{ji}} \right)_i, \tag{18}$$

μ_i is the average distance between each data point x_j in cluster A_i and the center of A_i

Step 5: Update the centers of all clusters

$$\left(L^i \right)_l = \left(\left(L_1^i \right)_l, \left(L_2^i \right)_l \right) = \left(\frac{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1} \right)^2 p^j}{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1} \right)^2}, \frac{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1} \right)^2 w^j}{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1} \right)^2} \right), \tag{19}$$

Step 6: Calculate

$$\left(d_{1j} \right)_l = d^2 \left(x^j, \left(L^i \right)_l \right) = \left(p^j - \left(L_1^i \right)_l \right)^2 + \left(w^j - \left(L_2^i \right)_l \right)^2, \tag{20}$$

for $i = 1, 2, 3, 4, j = 1, 2, \ldots, n$

Step 7: Calculate

If

$$\left(d_{1j}\right)_l = 0 \vee \left(d_{2j}\right)_l = 0 \vee \left(d_{3j}\right)_l = 0 \vee \left(d_{4j}\right)_l = 0 \text{ then } q_{ij}^l = 1$$

Else,

$$q_{ij}^l = \frac{W_{ij}}{\dfrac{(d_{ij})_l}{(d_{1j})_l} + \dfrac{(d_{ij})_l}{(d_{2j})_l} + \dfrac{(d_{ij})_l}{(d_{3j})_l} + \dfrac{(d_{ij})_l}{(d_{4j})_l}}, \tag{21}$$

For $i = 1, 2, 3, 4, j = 1, 2, \ldots, n$

Step 8: Calculate

If

$$\left| q_{ij}^l - q_{ij}^{l-1} \right| < 10^\varepsilon, i = 1, 2, 3, 4, j = 1, 2, \ldots, n \tag{22}$$

then $P = \{A_1, A_2, A_3, A_4\}$, where $A_i\left(x^j\right) = q_{ij}^l := q_{ij}$, stop

Else,

let $l := l + 1$

Step 9: Calculate

$$\left(L^i\right)_l = \left(\left(L_1^i\right)_l, \left(L_2^i\right)_l\right) = \left(\frac{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1}\right)^2 p^j}{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1}\right)^2}, \frac{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1}\right)^2 w^j}{\sum\limits_{j=1}^{n} \left(q_{ij}^{l-1}\right)^2} \right), \tag{23}$$

For $i = 1, 2, 3, 4$, and go to Step 5.

3 Results and Analysis

3.1 Samples Collection

The purpose of this study is to understand the satisfaction and importance of foreign tourists outing and hiking in the SMLSS of Taiwan. Thus, the major objects of this research would target at foreign tourists. The test date is from July 1, 2019 to September 30, 2019, and the interview time starts from 9:00 am to 5:00 pm. Since it is unlikely to know the daily arrival time of tourists in advance, it is not possible to use random sampling methods to obtain samples. Therefore, samples must be received after a convenient face-to-face interview between the tester and the interviewee.

A total of 100 valid samples were collected. IPA presumes that a given set of attributes is spread across four quadrants, "Quadrant I (Concentrate Here)", "Quadrant II (Keep Up the Good Work)", "Quadrant III (Low Priority)" and "Quadrant IV (Possible Overkill)".

While these properties are located in a "Quadrant I" area, the travelers would think these properties to be very vital but they feel it a bit of not satisfying or too normal. It means that when the properties fall in this area of "Quadrant II", the tourist value and performance of these properties will be relatively high, so it is wiser for them to

continue to maintain their performance attributes. In addition, if the properties fall within the "Quadrant III" area, it means that travelers do not pay much attention to the properties. In other words, the company does not provide these properties to travelers, or very limited resources invested by government. However, if the property belongs to the "Quadrant IV" area, it indicates high performance, but these properties were not enough attention for the tourists. This shows that the government has invested too many resources here, but the properties of the region have not affected the satisfaction of travelers.

3.2 Traditional IPA Analysis

3.2.1 Quadrant I

If these are placed in the "Quadrant I" area by using conventional IPA attributes, then the value and performance of these attribute travelers will be relatively high. Accordingly, it is sensible that they should keep on maintaining their attributes of performance. In this study, there are five items in quadrant KW, i.e., item 2. Provision of guidelines system, item 3. Provide recreational facilities, item 20. Landscape, item 22. Hospitality shown by the local community, and item 24. Acceptance of foreign languages by local communities.

3.2.2 Quadrant II

When the property drops in this area of "Quadrant II" with the use of traditional IPA, it shows that the travelers regard these properties as very vital but they feel it too common or unsatisfactory. There are five items in quadrant CH in this study, i.e., item 4. Number of attractions, item 8. Levels of products and services, item 11. Participate in activities while on vacation, item 19. Diversity of terrain, item 21. Cleanliness and damage to nature, and item 7. Security.

3.2.3 Quadrant III

When using traditional IPA to drop the property into the "Quadrant III" area, this means that tourists do not pay too much focus on the properties of these services. In other words, these companies have very limited resources, or the enterprises don't offer these properties for travelers. In this study, there are five items in quadrant LP, i.e., item 10. Enjoying food, item 12. Watch famous events internationally, item 14. Increase the scope of your life experience, and item 15. Get a hands-on entertainment experience.

3.2.4 Quadrant IV

When the property drops in this area of "Quadrant IV" with the use of conventional IPA, it shows that, despite the high performance, travellers pay insufficient attention to these attributes. The results show that travellers have very little satisfaction with the service attributes of the area. However, the government invests relatively too much resources here. In this study, there are five items in quadrant PO, i.e., item 9. Uniqueness of culture, item 18. Comfortable, item 16. Suitability to provide entertainment, and item 25. Information services.

However, there are two items on the horizontal axis, i.e., item 6. Convenience of access and item 17. Quality of facilities for activities. There are four items on the vertical axis, i.e., item 1. Quality of the transportation system, item 5. Different types of accommodation, item 13. Look at some events that reflect the nature of the destination, and item 23. Use of information centres. These values of the service attributes are exactly on the cross axis, so that these service attributes cannot be distinguished to what quadrant which should be assigned.

3.3   IPASIIBFCM Analysis

3.3.1   Quadrant I

There are eight items in quadrant "Quadrant I", i.e., item 2. Offer of guidelines system, item 3. Provide recreational facilities, item 20. Landscape, item 22. Hospitality as is shown by the local residence, item 24. Acceptance of foreign languages by local communities, item 6. Convenience of access, item 1. Quality of the transportation system, item 13. Look at some events that reflect the nature of the destination, item 23. Use of information centres.

3.3.2   Quadrant II

There are seven items in "Quadrant II", i.e., item 4. Number of attractions, item 8. Cost Value Levels of products and services, item 11. Participate in activities while on vacation. item 19. Diversity of terrain, item 21. Cleanliness and damage to nature, item 7. Security, and item 5. Various types of housing.

3.3.3   Quadrant III

There are four items in "Quadrant III", i.e., item 10. Enjoying food, item 12. Watch famous events internationally, item 14. Increase the scope of your life experience, and item 15. Get a hands-on entertainment experience.

3.3.4   Quadrant IV

There are four items in "Quadrant IV", i.e., item 9. Uniqueness of culture, item 18. Comfortable, item 16. Suitability to provide entertainment, item 25. Information services, and item 17. Quality of facilities for activities.

4   Conclusion

The results show that it is easy, simplified, and effective to put to use the IPA by Size-Insensitive Integrity-based Fuzzy C-Means. The results also show that Size-Insensitive IPA based on Fuzzy C-Means can overcome the problem of attribute classification on the cross axis that cannot be classified in conventional IPA.

References

1. Ban, O.I., Ban, A.I., Tuse, D.A.: Importance-performance analysis by fuzzy C-Means algorithm. Expert Syst. Appl. **50**, 9–16 (2016)
2. Caber, M., Albayrak, T., Matzler, K.: Classification of the destination attributes in the content of competitiveness by revised importance–performance analysis. J. Vacat. Mark. **18**, 43–56 (2012)
3. Cheng, S.H.: Appling importance-performance analysis by fuzzy C-means to evaluate tourist friendly destination. In: Proceedings of the 2018 International Conference on Machine Learning and Cybernetic, pp. 480–484, July 2018
4. Gath, I., Smolyak Iskoz, A., van Cutsem, B.: Data induced metric and fuzzy clustering of non-convex patterns of arbitrary shape. Pattern Recognit. Lett. **18**, 541–553 (1997)
5. Lin, P.L., Huang, P.W., Kuo, C.H., Lai, Y.H.: A size-insensitive integrity-based fuzzy c-means method for data clustering. Pattern Recognit. **47**, 2042–2056 (2014)
6. Martilla, J.A., James, J.C.: Importance–performance analysis. J. Mark. **41**, 77–79 (1977)
7. Oh, H.: Revisiting importance–performance analysis. Tour. Manag. **22**(6), 617–627 (2001)
8. Vaske, J.J., Beaman, J., Stanley, R., Grenier, M.: Importance-performance and segmentation: where do we go from here. J. Travel. Tour. Mark. **5**(3), 225–240 (1996)

Gingivitis Identification via GLCM and Artificial Neural Network

Yihao Chen[1,2] and Xianqing Chen[1(✉)]

[1] Department of Electrical Engineering, College of Engineering, Zhejiang Normal University, Jinhua 321004, Zhejiang, China
cxq@zjnu.edu.cn
[2] School of Informatics, University of Leicester, Leicester LE1 7RH, UK

Abstract. Gingivitis is a common oral disease. The diagnosis process of gingivitis disease is usually based on the experience of the dentist and previous medical records. In order to diagnose gingivitis more efficiently and accurately, we proposed a gingivitis recognition program based on Gray-Level Co-Occurrence Matrix (GLCM), Artificial Neural Network (ANN) and Genetic Algorithms (GA). We obtained 180 oral images from Nanjing Stomatological Hospital through two professional medical cameras, 90 of which belong to the gingivitis class, and the rest of which are from the healthy class. We combined GLCM and Artificial Neural Network to identify gingivitis by K-fold Cross-Validation (CV), in our experiment, we utilized the 10-fold Cross-Validation algorithm. We used six evaluation indicators to objectively evaluate the classification performance, which are sensitivity, specificity, precision, accuracy, F1, and MCC. Compared with Contrast Limited Adaptive Histogram Equalization (CLAHE) and GLCM plus Extreme Learning Machine (ELM), the identification performance of our algorithm is better than them.

Keywords: Gingivitis identification · Gray-Level Co-occurrence Matrix (GLCM) · Artificial Neural Network (ANN) · Genetic Algorithms (GA) · K-fold Cross-Validation (CV)

1 Introduction

The tooth is an important organ in the human digestive system. With the growth of people, the tooth plays a different and indispensable role because it is necessary to chew up the food and promote digestion and absorption [1–3]. There are a large number of primary and secondary tooth diseases at all ages, such as gingivitis [1], periodontitis [4], periapical periodontitis [5] and etc. Gingivitis is a major dental disease which is caused by the gingiva pathological changes. It divides into acute gingivitis and chronic gingivitis. The main clinical manifestation always shows bright-red or dark-red free gingival and gingival nipples, swelling of gum tissue, soft and fragile gums. Generally, plaque buildup and long-term effects of harmful substances cause damage to the gum tissue and promote the severity of gingivitis symptoms. Removing tooth calculus and

controlling the plaque are the main measures to cure it and brushing teeth is an effective way to prevent it.

The diagnosis is mainly based on the main clinical manifestations, obvious plaque, calculus accumulation, and other plaque retention factors near the gingival margin. However, the manifestation of gingivitis and early periodontitis are quite similar because they are both caused by the buildup of the plaque and the main manifestations occur around the gum tissue. Whether there is a periodontal pocket and whether the alveolar bone is absorbed are the keys to distinguish the two diseases. Some blood diseases can also cause symptoms of bleeding gums, such as hemophilia and leukemia. Therefore, it is necessary to detect the main features of gingivitis. In clinical diagnosis, medical images or computed tomography images should be utilized to help doctors to diagnose.

Machine learning technique has been developed a lot and applied in many fields, such as computer vision, medical image analysis, speech recognition, categories classification. In recent years, the medical image analysis based on machine learning and artificial intelligence is one of the most popular and effective topics. Due to the development of the image processing algorithms and machine learning technique, the performance of the projects in medical imaging field appears a significant process. Brown [6] combined Gray-Level Co-occurrence Matrix (GLCM) and Extreme Learning Machine (ELM). Li [7] added a preprocessing technique using Contrast-Limited Adaptive Histogram Equalization (CLAHE).

In this study, we proposed an automatical gingivitis identification system with high performance to solve the binary classification problem. Gray-Level Co-occurrence Matrix is a traditional and efficient algorithm focusing the texture features in the images. Artificial Neural Network is a common method which can effectively classify the different classes. According to our dataset is not big enough, some deep learning methods can not be utilized, such as Convolutional neural network (CNN). And the aim of this project is to improve classification accuracy and calculation efficiency.

2 Dataset

In this project, we obtained our dataset from Nanjing Stomatological hospital, Nanjing City, Jiangsu Province, China. We selected some gingivitis patients whose teeth and gums have different severity levels as our experimental subjects from this hospital. In order to acquire the gingivitis images, we photographed patients' different inflammatory areas of the oral cavities through two professional dental digital single-lens reflexes which are named A and B. The images include inflamed parts and healthy parts. Three categories of the fields of vision are pointed which are mesial, distal and middle, and the focal length of the shot lens is from 51 to 200 mm. Eventually, we sifted the images we shot and selected 180 images randomly, in which 90 are gingivitis and 90 are healthy gum images, as shown in Fig. 1.

<table>
<tr><td>(a) Gingivitis</td><td>(b) Healthy</td></tr>
</table>
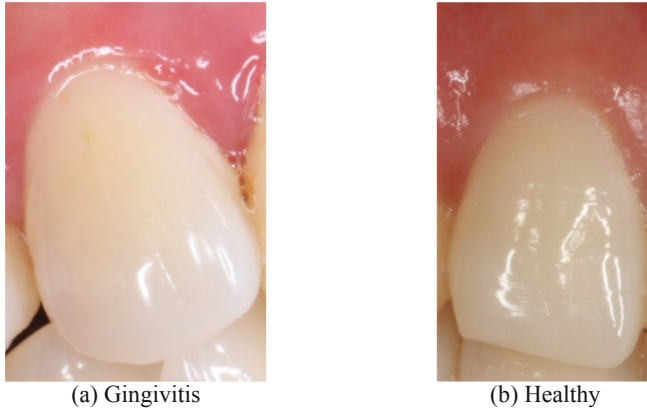
Fig. 1. Two samples of the gingivitis and the healthy class.

Due to some inevitable causes of man-made and shooting limitations, we adjusted the size of the images and the gray level to highlight the key objects. Eventually, we obtained a set of images, The length and width of which are in the range of 701 ± 461 and 1338 ± 973 pixels, respectively. The image pixel's gray value is set as 8 bit.

3 Methodology

With the development of machine learning and computer vision, the application of medical image analysis is increasingly extensive. We proposed a system combined Gray-Level Co-occurrence Matrix (GLCM) algorithm and Artificial Neural Network (ANN) on our dataset. The results are obtained through 10-fold cross-validation

3.1 Gray-Level Co-occurrence Matrix

Gray-Level Co-occurrence Matrix (GLCM) algorithms are sufficient algorithms that can identify and extract the texture feature of an image. It represents the relationship between neighbor pixels by calculating the gray-level value of every pixel. Compared with other methods to extract features of an image, such as wavelet transform and Fourier transform, GLCM can represent the spatial relations in texture features because it focuses more on directions and local features.

There are mainly four steps to obtain the features: grayscale the image, determine the gray value, calculate the Gray-Level Co-occurrence Matrix in different directions, and calculate the features of the matrixes. Since we obtained the RGB-three-channel images, it is significant to utilize GLCM algorithm to extract the features because it can focus more on the texture feature and ignore the influence in color and light of the images.

There are two main parameters of the GLCM algorithm, which are d and θ. d represents the direct distance between two neighbor pixels, and θ represents the angle of calculation. Since every pixel is surrounded by other eight pixels, normally θ equals $45°$, $90°$, $135°$ and $180°$. Therefore, the image can be scanned in four directions.

G, the value of the gray level, needs to be determined from 0 to G-1. In order to reduce the amount of calculation and enhance efficiency, G should be compressed to an appreciate range. Normally, it would be reduced from 255 to 16 or 8. Histogram equalization is a commonly used method to compress the gray level and guaranteed not to affect texture features and image information.

Suppose f (x, y) is an objective image, and the size of it is M * N. A window, the size of which can be determined, slides over the matrix and covers every part of the matrix in one direction. The relative frequencies can be represented as P (i, j), which satisfies the following equation:

$$P(i, j) = \#\{(x_1, y_1), (x_2, y_2) \in M \times N | f(x_1, y_1) = i, f(x_2, y_2) = j\}, \quad (1)$$

where #(x) represents a set of elements.

GLCM can represent local texture features in different directions of an image by the difference in gray values between directly adjacent pixels in a specific direction. Besides, if there are too many complicated features in the four directions, the features in the four directions can be averaged and finally a matrix containing comprehensive features can be obtained.

3.2   Artificial Neural Network

Artificial Neural Network (ANN) is a complex network structure [8–12] formed by the interconnection of many processing units (neurons). ANN is an abstraction, simplification, and simulation of the structure of the human brain. It is applied in the machine learning field, such as speech identification and image processing [13–17]. There are two main networks, which are single layer neural networks and multi-layer neural networks respectively. In single layer neural networks structure, there are only Input layer and Output layer. And in multi-layer neural networks, the Hidden layer is added into the networks, which calculates and transfers the data and inputs processed data to later layers.

The role of the hidden layer is to perform non-linear calculations and to deepen network depth and strengthen network computing capabilities. Every neuron is associated with a specific output function, which is an activation function. The connection between every two neurons shows a weighted value for signals passing through the connection, namely weight (ω).

A layer perceptron consists of one or more hidden layers [18–24]. It should be trained to determine the parameters in the networks, such as an appropriate weight and bias. The normal structure is shown in Fig. 2. x_i represents the input element, y_i represents the output of the networks, and ω_i represents the weighted value.

Three popular activation functions are as following:

1. Sigmoid:

$$f(x) = \frac{1}{1+e^{-x}}. \quad (2)$$

2. TanH:

$$f(x) = \frac{sinX}{cosX} = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (3)$$
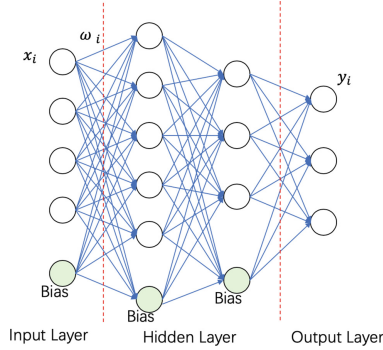
Fig. 2. The structure of a sample neural network.

3. ReLU:

$$f(x) = \max(0, x).\tag{4}$$

In this study, we used sigmoid function [25–29] since it is the most popular function in shallow neural networks.

3.3 Genetic Algorithm

Genetic algorithm is a computational model that simulates the biological evolutionary processes of natural selection and genetics, which is a method of searching for the optimal solution [30]. The purpose of genetic algorithm is to find one or some individuals with specific features, which is named chromosome, in a population. The chromosome is a set of one or several genes and it determines the external appearance of the individual.

The original population is artificially encoded to produce the primary population. In this generation of population, excellent individuals are selected according to the principle of natural selection, which is approximate solutions. By continuing this process, the optimal solution to the problem can be obtained. In this process, genetic operators of natural genetics can be used for crossover and mutation. The flow of the genetic algorithm is shown in Fig. 3.

Several common encoding methods in genetic algorithms include binary encoding, floating point encoding, and character encoding. Particularly, binary encoding is the most commonly used encoding. Besides, there are three basic genetic operators, which are selection, crossover, and mutation. Selection shows a process to select advantaged individuals and eliminate disadvantaged individuals. Crossover means that substituting and reorganizing parts of the structure of two previous generation individuals to produce the next generation of individuals. Mutation will change some of genetic value of individuals.

After each evolution round, we can calculate the fitness of each chromosome to determine the probability that a particular chromosome would be selected.

$$P_k = \frac{F_k}{\sum_{i=1}^{m} F_i},\tag{5}$$
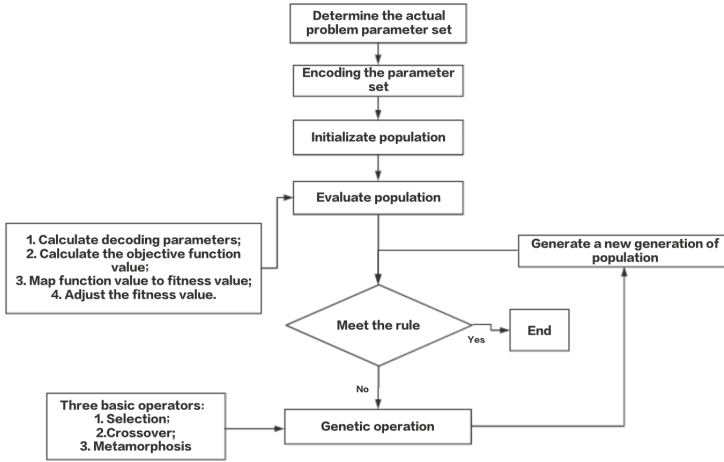
Fig. 3. The flow chart of genetic algorithm.

where P_k represents the selection probability of the chromosome k, F_i represents the fitness of the chromosome i, and m represents the number of chromosomes.

If the fitness of the best individual equals to the given threshold, or the best individual fitness and population fitness no longer increase, or the number of iterations reaches the preset number, the algorithm terminates.

3.4 10-Fold Cross-Validation

K-fold Cross-Validation (CV) is a popular testing method. The aim of it is to test the accuracy of the algorithm. Due to the disadvantages of a small dataset, K-fold Cross-Validation is utilized to increase the calculation times and guarantee accuracy. Normally, 10-fold Cross-Validation is a commonly used method.

The process is that the dataset should be evenly divided into 10 subsets. For every validation round, each subset should be regarded as a test set and the rest as a training set, as shown in Fig. 4. The cross-validation process should repeat 10 times, and the average accuracy is more reliable.
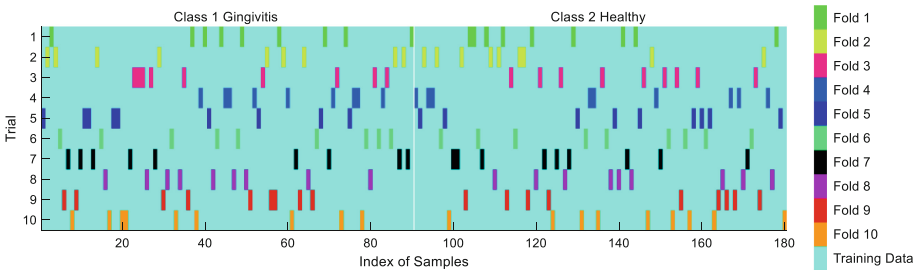
Fig. 4. Index of Cross Validation

There are some performance evaluation criterions in classification field, which are sensitivity, specificity, precision, accuracy, F1, Matthews correlation coefficient (MCC), and etc. There are four parameters in binary classification, which are True Positive (Positive samples predicted to be positive), True Negative (Negative samples predicted to be negative), False Positive (Positive samples predicted to be negative), False Negative (Negative samples predicted to be positive). Sensitivity means true positive rate, Specificity means true negative rate, Precision means the correct positive prediction rate in the positive prediction class, and Accuracy means the correct prediction rate in all samples. The formulas show in following:

$$Sensitivity = TPR = \frac{TP}{TP+FN} \cdot \tag{6}$$

$$Specificity = TNR = \frac{TN}{TN+FP} \cdot \tag{7}$$

$$Precision = \frac{TP}{TN+TP} \cdot \tag{8}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \cdot \tag{9}$$

And F1 means weighted harmonic mean of recall and precision, and MCC is a correlation coefficient describing the actual classification and predicted classification. The fomulas are as showing in following:

$$F1 = \frac{2*TP}{2*TP+FN+FP} \cdot \tag{10}$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \cdot \tag{11}$$

4 Experiment Results and Discussions

4.1 Statistical Results

This program was developed based on GLCM and Artificial neural network by utilizing the valuable dataset which was obtained from Nanjing Stomatological Hospital. In order to reduce the amount of calculation and maintain enough features in the image, we reduced the gray level value from 16 bit to 8 bit for both training set and testing set. GLCM method can reduce the effect of color, and instead promote to pay more attention to texture features.

In our program, we regard the gingivitis class as the positive class and the healthy class as the negative class. Therefore, True Positive represents that the gingivitis samples are classified to be gingivitis, and False Negative represents that the healthy samples are classified to be health.

Due to the small size of our dataset, we proposed to apply the 10-fold cross-validation algorithm in our program. To make the results less accidental, we ran the program 10 times and take the average of the 10 results. Also we introduced another indicator standard deviation (SD) to show the reliability of the results. SD can reflect the degree of dispersion of the results. The results are as shown in Table 1. The mean sensitivity, specificity, precision, accuracy, F1 and MCC are 75.11%, 75.78%, 75.91%, 75.44%, 75.18% and 51.28% respectively.

Table 1. Results of our proposed method.

Run	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
1	74.44	78.89	78.64	76.67	76.21	53.76
2	71.11	76.67	75.12	73.89	72.80	48.10
3	74.44	82.22	80.93	78.33	77.43	56.99
4	72.22	78.89	77.73	75.56	74.60	51.52
5	73.33	75.56	75.09	74.44	74.07	49.02
6	78.89	73.33	74.55	76.11	76.25	52.78
7	77.78	74.44	75.44	76.11	76.55	52.31
8	81.11	71.11	73.97	76.11	76.78	53.26
9	77.78	68.89	71.98	73.33	74.63	46.98
10	70.00	77.78	75.62	73.89	72.49	48.10
Mean + SD	75.11 ± 3.45	75.78 ± 3.78	75.91 ± 2.42	75.44 ± 1.47	75.18 ± 1.63	51.28 ± 3.00

4.2   Training Algorithm Comparison

We proposed two methods based on back propagation (BP) [31] and simulated annealing (SA) [32], instead of genetic algorithm. And the results are as shown in Table 2 and Table 3. We used the same indicators to record the results of the two algorithms.

Table 2. Results of the methods based on BP algorithm

BP [31]	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
1	64.44	68.89	68.03	66.67	66.01	33.56
2	68.89	68.89	69.22	68.89	68.85	38.00
3	66.67	66.67	66.83	66.67	66.71	33.37
4	73.33	77.78	76.67	75.56	74.90	51.22
5	66.67	60.00	62.83	63.33	64.60	26.79
6	61.11	71.11	68.40	66.11	64.35	32.59
7	67.78	70.00	69.13	68.89	68.20	38.01
8	60.00	71.11	68.52	65.56	63.66	31.69
9	68.89	61.11	63.53	65.00	65.96	30.31
10	70.00	72.22	71.44	71.11	70.45	42.56
Mean + SD	66.78 ± 3.83	68.78 ± 4.96	68.46 ± 3.69	67.78 ± 3.34	67.37 ± 3.22	35.81 ± 6.65

We compared the results of genetic algorithm with BP and SA, as shown in Table 4. From this table, each indicator of our approach is better than BP and SA, and the value of standard deviation is smaller than them, which means the classification performance of our genetic algorithm is more stable and reliable.

Table 3. Results of the methods based on SA algorithm

SA [32]	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
1	55.56	63.33	60.56	59.44	57.88	19.00
2	70.00	57.78	62.57	63.89	65.90	28.14
3	63.33	65.56	64.75	64.44	63.96	28.94
4	64.44	55.56	59.11	60.00	61.64	20.11
5	66.67	64.44	65.33	65.56	65.96	31.14
6	66.67	58.89	62.24	62.78	64.28	25.65
7	55.56	66.67	62.50	61.11	58.82	22.36
8	62.22	56.67	59.40	59.44	60.56	18.97
9	64.44	57.78	60.95	61.11	62.53	22.32
10	63.33	68.89	67.25	66.11	65.14	32.36
Mean + SD	63.22 ± 4.37	61.56 ± 4.51	62.47 ± 2.51	62.39 ± 2.38	62.67 ± 2.74	24.90 ± 4.77

Table 4. Comparison of different training algorithms

Approach	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
BP [31]	66.78 ± 3.83	68.78 ± 4.96	68.46 ± 3.69	67.78 ± 3.34	67.37 ± 3.22	35.81 ± 6.65
SA [32]	63.22 ± 4.37	61.56 ± 4.51	62.47 ± 2.51	62.39 ± 2.38	62.67 ± 2.74	24.90 ± 4.77
GA (Ours)	75.11 ± 3.45	75.78 ± 3.78	75.91 ± 2.42	75.44 ± 1.47	75.18 ± 1.63	51.28 ± 3.00

4.3 Comparison to State-of-the-Art Approaches

There were two new methods were applied in the same field, which are GLCM+Extreme Learning Machine (ELM) and Contrast Limited Adaptive Histogram Equalization (CLAHE). We can obtain a set of better results compared with the two algorithms. The rate of each indicator is higher than the previous methods, as shown in Table 5.

Table 5. Comparison with SOTA

Approach	Sensitivity	Specificity	Precision	Accuracy
GLCM + ELM [6]	72	70	70.59	71
CLAHE [7]	75	73	74	74
GLCM + GNN (Ours)	75.11 ± 3.45	75.78 ± 3.78	75.91 ± 2.42	75.44 ± 1.47

5   Conclusions

The aim of our project is to develop a gingivitis identification system with high performance. The dataset is built by our group from Nanjing Stomatological hospital, including 180 images. Besides, we combined Gray-Level Co-occurrence Matrix, Artificial Neural Network and Genetic Algorithm to solve this binary classification problem. The results showed that this method can successfully identify the gingivitis and healthy patients. Moreover, we compared our project with some new methods, and our research shows a reliable classification performance.

However, the accuracy of each indicator is not high enough, which means there are some mistakes or gaps in the experiment. In the future, we will try to collect more gingivitis images and utilize some data augmentation methods to build a valid dataset. And we will also improve and develop the methods and attempt to apply other methods to obtain a better performance.

References

1. Lang, N.P., et al.: Gingivitis as a risk factor in periodontal disease & NBSP. J. Clin. Periodontol. **36**(s10), 3–8 (2009)
2. Molnar, S.: Human tooth wear, tooth function and cultural variability. Am. J. Phys. Anthropol. **34**, 175–179 (1971)
3. Lee, J.W., et al.: Fracture in teeth: a diagnostic for inferring bite force and tooth function. Biol. Rev. **86**(4), 959–974 (2011)
4. Tonetti, M.S., et al.: Treatment of periodontitis and endothelial function. J. Vasc. Surg. **45**(6), 1286–1287 (2007)
5. Noiri, Y., et al.: Participation of bacterial biofilms in refractory and chronic periapical periodontitis. J. Endod. **28**(10), 679–683 (2002)
6. Brown, M.: Gingivitis identification via grey-level cooccurrence matrix and extreme learning machine. Adv. Soc. Sci. Educ. Humanit. Res. **250**, 486–492 (2018)
7. Li, W.: A gingivitis identification method based on contrast-limited adaptive histogram equalization, gray-level co-occurrence matrix, and extreme learning machine. Int. J. Imaging Syst. Technol. **29**(1), 77–82 (2019)
8. Chen, Y.: Cerebral micro-bleeding identification based on a nine-layer convolutional neural network with stochastic pooling. Concurr. Comput. Pract. Exp. **31**(1), e5130 (2020)
9. Govindaraj, V.V.: High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. J. Med. Imaging Health Inform. **9**(9), 2012–2021 (2019)
10. Jiang, X.: Chinese sign language fingerspelling recognition via six-layer convolutional neural network with leaky rectified linear units for therapy and rehabilitation. J. Med. Imaging Health Inform. **9**(9), 2031–2038 (2019)
11. Ramirez, J.: Unilateral sensorineural hearing loss identification based on double-density dual-tree complex wavelet transform and multinomial logistic regression. Integr. Comput. Aided Eng. (2019). https://doi.org/10.3233/ICA-190605
12. Li, Z.: Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling, Int. J. Imaging Syst. Technol. (2019). https://doi.org/10.1002/ima.22337

13. Hong, J., Cheng, H., Zhang, Y.-D., Liu, J.: Detecting cerebral microbleeds with transfer learning. Mach. Vis. Appl. **30**(7-8), 1123–1133 (2019). https://doi.org/10.1007/s00138-019-01029-5

14. Tang, C.: Cerebral micro-bleeding detection based on densely connected neural network. Front. Neurosci. **13**, 422 (2019)

15. Xie, S.: Alcoholism identification based on an AlexNet transfer learning model. Front. Psychiatry **10**, 205 (2019)

16. Jia, W., Muhammad, K., Wang, S.-H., Zhang, Y.-D.: Five-category classification of pathological brain images based on deep stacked sparse autoencoder. Multimedia Tools Appl. **78**(4), 4045–4064 (2017). https://doi.org/10.1007/s11042-017-5174-z

17. Zhang, Y.-D., Dong, Z., Chen, X., Jia, W., Du, S., Muhammad, K., Wang, S.-H.: Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. Multimedia Tools Appl. **78**(3), 3613–3632 (2017). https://doi.org/10.1007/s11042-017-5243-3

18. Wang, S.-H., Muhammad, K., Hong, J., Sangaiah, A.K., Zhang, Y.-D.: Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization. Neural Comput. Appl. **32**(3), 665–680 (2018). https://doi.org/10.1007/s00521-018-3924-0

19. Yang, J.: An adaptive encoding learning for artificial bee colony algorithms. J. Comput. Sci. **30**, 11–27 (2019)

20. Wang, S.-H., Sun, J., Phillips, P., Zhao, G., Zhang, Y.-D.: Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. J. Real Time Image Proc. **15**(3), 631–642 (2017). https://doi.org/10.1007/s11554-017-0717-0

21. Huang, C.: Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization dropout, and stochastic pooling. Front. Neurosci. **12**, 818 (2018)

22. Gorriz, J.M.: Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. J. Alzheimers Dis. **65**(3), 855–869 (2018)

23. Liu, G., et al.: Detection of Alzheimer's disease by three-dimensional displacement field estimation in structural magnetic resonance imaging. J. Alzheimers Dis. **50**(1), 233–248 (2016)

24. Phillips, P.: Three-dimensional eigenbrain for the detection of subjects and brain regions related with Alzheimer's Disease. J. Alzheimers Dis. **50**(4), 1163–1179 (2016)

25. Zhang, Y.-D., Sun, J.: Preliminary study on angiosperm genus classification by weight decay and combination of most abundant color index with fractional Fourier entropy. Multimedia Tools Appl. **77**(17), 22671–22688 (2017). https://doi.org/10.1007/s11042-017-5146-3

26. Zhao, G.: Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and jaya algorithm. Multimedia Tools Appl. **77**(17), 22629–22648 (2018)

27. Pan, C.: Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. J. Comput. Sci. **28**, 1–10 (2018)

28. Zhang, Y.-D., Zhang, Y., Hou, X.-X., Chen, H., Wang, S.-H.: Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed. Multimedia Tools Appl. **77**(9), 10521–10538 (2017). https://doi.org/10.1007/s11042-017-4554-8

29. Li, Y.-J.: Single slice based detection for Alzheimer's disease via wavelet entropy and multi-layer perceptron trained by biogeography-based optimization. Multimedia Tools Appl. **.**(9), 10393–10417 (2018)

30. Mousakazemi, S.M.H.: Computational effort comparison of genetic algorithm and particle swarm optimization algorithms for the proportional-integral-derivative controller tuning of a pressurized water nuclear reactor. Ann. Nucl. Energy **136**, 107019 (2020)

31. Miao, Y.F., et al.: Simulation and adaptive control of back propagation neural network proportional-integral-derivative for special launcher using new version of transfer matrix method for multibody systems. J. Vib. Control **26**, 12 (2019). Article ID Unsp. 1077546319889786
32. Cazanas, R.D., et al.: Joint programming of production-maintenance tasks: a simulated annealing-based method. Int. J. Simul. Model. **18**(4), 666–677 (2019)

A Novel Classification Method of Medical Image Segmentation Algorithm

Yu Kong[1(✉)] ⓘ, Yueqin Dun[2], Jiandong Meng[3], Liang Wang[1], Wanqiang Zhang[1], and Xinchun Li[1]

[1] Shandong Medical College, Erhuananlu, Jinan 5460, Shandong, China
kongy@sdmcjn.edu.cn
[2] School of Electrical Engineering, University of Jinan, Jinan, Shandong, China
cse_dunyq@ujn.edu.cn
[3] Shandong Medical College, 6th Jucai Road, Linyi, Shandong, China

Abstract. Medical image segmentation is a relevant and active research field of medical image processing. The proposal of various algorithms not only enriches the means to solve the problem of medical image segmentation but also makes the algorithm classification and summary urgent. At present, a variety of classification methods are mostly based on the characteristics of the algorithm itself. If the classification principle of the algorithm is determined according to the essential elements of the organ plane space, such as point, line, and surface, a new classification method will be formed, and it is more in line with people's intuitive feelings. Using this new segmentation principle to classify medical image segmentation algorithms is helpful to clarify the relationship between various algorithms.

Keywords: Image segmentation · Medical image · Algorithm classification

1 Introduction

Medical image segmentation is a fundamental and vital part of medical image processing. It is the process of splitting an image into different groups of pixels and converting them into meaningful themes. The purpose of medical image segmentation is to reduce the burden of medical staff and improve work efficiency. Because of the high variability of the image, the image processing is complicated. At present, to solve a variety of segmentation scenarios, researchers have proposed many medical image segmentation algorithms from many aspects. The classification and induction of these methods are helpful for further research work.

According to the characteristics of the algorithm, Dzung L. Pham et al. divided the segmentation methods into eight categories. See Fig. 1 [1].

According to the algorithm function and generation time, Withey and Koles divided the medical image segmentation literature into three categories and three generations, As shown in Fig. 2 [2].
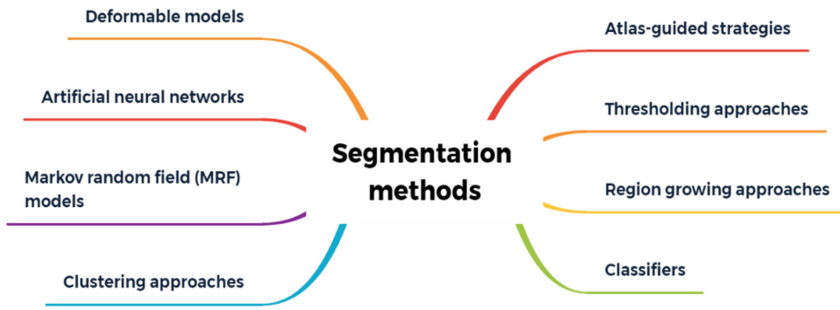
Fig. 1. Dzung L. Pham's classification method
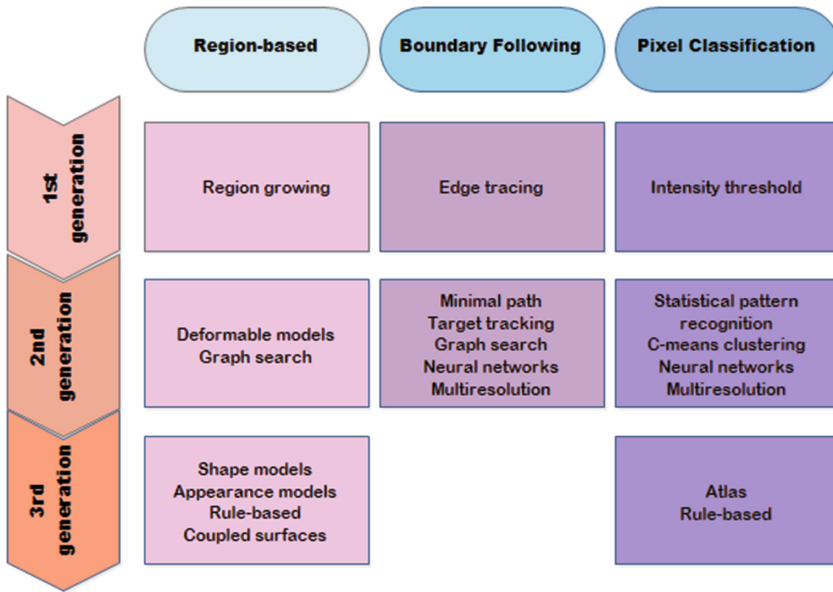
Fig. 2. Withey and Koles's classification method

Neeraj Sharma and Lalit m. Aggarwal divided the segmentation technology into grayscale segmentation and texture feature segmentation from the perspective of medical image processing [3].

Rule-based, statistical, atlas-based, and deformable models-based techniques used by Ahmed Elnakib et al. as a classification criterion [4].

The classification proposed by M. Ehrdt et al. gives the classification concept of "continuity". All the segmentation algorithms are based on the continuous process of two endpoints of the classification, namely, strong shape-based methods and purely image-based algorithms [5].

Alireza Norouzi et al. divided these algorithms into four categories: region-based method, clustering method, classifier method, and hybrid method, and discussed them in detail [6].

According to the popularity of the algorithm, Zhou LQ et al. proposed, at present, there are two kinds of artificial intelligence methods used in medical imaging, one is the traditional machine learning algorithm, the other is the deep learning algorithm [7].

For machine learning, Iman Aganj et al. divide medical image segmentation algorithms into two categories: supervised learning and unsupervised learning [8].

According to the most commonly used methods in processing CT and MRI data, Voronin, V. et al. divided the medical image segmentation algorithm into the methods of active contour, region growing, edge detection, and hybrid techniques [9].

Based on the actual clinical use at present, Tomas Sakinis et al. group current segmentation approaches in three classes: Manual, semi-automatic, and fully-automatic [10].

2   The Novel Classification Method

According to the principle of classification based on the essential elements of plane space, the medical image segmentation algorithm is divided into three methods: the segmentation method based on the plane shape principle, the method based on the contour curve, and the method based on image pixels (voxels).

2.1   Segmentation Method Based on Organ Section Shape

This kind of algorithm compares the plane image of the organ to be segmented with the prior image as a whole, to realize the whole segmentation.

The statistical shape model (SSMs) method treats the shape to be segmented as a whole and compares it with one or a group of reference shapes to segment the target. Complex objects can be segmented even in low-contrast images. Statistical shape model (SSMs) describes the shape of an object by applying PCA to a set of ground punctuation, thus establishing the shape change of an object class to create a statistical model, and using this model to prior knowledge to the algorithm. The most famous examples of SSM are active shape models and deformable models.

Atlas-guided method regards segmentation as an image registration problem and establishes the spatial correspondence between Atlas and different image coordinates. Registration is usually a computationally intensive task that requires the use of an appropriate deformation model to deform an image until it resembles another image. The resulting mapping between the two coordinate systems can be used to transfer the segmentation label from the atlas to the new image voxel. A map and a deformation model are usually not enough to capture a wide range of anatomical changes. Therefore, the method of multi-atlas cooperative work has been applied more and more widely.

2.2   Based on the Outline of the Organ Section

The idea of the segmentation method based on the model contour is that the shape can be segmented if the overview of the organ to be segmented can be depicted. First of all,

the initial integral object contour curve model based on parameters is established, and the contour curve is deformed by some method, and the whole is continuously close to the real boundary of the figure to be segmented and finally coincides with the actual contour of the target. The segmentation method based on the model contour is also called the active contour model, which mainly includes the snake model and level set model, which are energy-based methods for contour curve deformation.

Kass et al. first presented the snake model. The basic idea is to define a continuous energy function curve to represent the edge of the target. The minimum can be obtained by solving the Euler-Lagrange equation. When the energy of the curve reaches the minimum value, the position of the curve is the position of the target contour. Subsequently, some models were improved on this basis, notably the balloon model, gradient vector flow (GVF), and topology-adaptive snakes (T-snakes).

In 1988, Sethian and Osher proposed the Level Set method that elevated some calculations of lower dimensions to higher dimensions and regarded the description of N dimensions as a Level of $N + 1$ dimension. The level set model is based on the method of energy. The idea is to construct the energy with the contour as the parameter, so that the energy of the shape is minimum when it coincides with the target, and always greater than zero when it does not correspond. In segmentation, the purpose of the level set representation is to define a continuous, smooth, segmented boundary representation and to evolve it to fit the boundary position in the image in some optimal way. Level set methods include both boundary-based and region-based models. The boundary-based method is mainly based on the local information at the edge contour, which is not suitable for medical images with a lot of noise. Region-based models take variational methods as the framework for a total consideration of regional information. The classical image segmentation model is the Mumford-shah model. Chan and Vese adopted a region-based global constraint segmentation model, based on the Mumford-Shah model. The fusion of regional information enhances the robustness of the model. However, when the object to be segmented is partially occluded or partially deficient, even the deformation model of boundary and region information fusion cannot get the correct results.

2.3 Based on Pixel Features of the Organs

This kind of method mainly uses the relationship between the grayscale of the pixel itself and the surrounding pixel grayscale of the organ section image to be segmented.

Segmentation Based on Graph Algorithm

The Methods Based on Graph Cut
The graph cut (GC) segmentation method proposed by Boykov et al. in 2001, which associates the image segmentation problem with the graph minimum cut (min-cut) problem. The GC algorithm represents the pixels of an image as nodes on a graph. Firstly, an undirected graph $G = <V, E>$ is used to describe the image to be segmented. V and E are the set of vertex and edge, respectively. The minimum cost function is found by using the principle of maximum flow (minimum cut) theorem in graph algorithm, and the image is divided into background and object. The main algorithms for solving the maximum flow problem are augmented path method (Ford-Fulkerson) and indentation and

relabeling algorithm (pre-flow advance algorithm) (Goldberg-Tarjan). Ford-Fulkerson uses a depth-first search to find an increasing path in the residual graph. Goldberg-Tarjan solves this problem by gradually pushing through the residual graph.

The Intelligent Scissors (Livewire) Method
The intelligent scissors (Livewire) method was proposed by Mortensen and Barrett in 1995 and is currently used in many medical image boundary demarcation. It is based on the Dijkstra method, a modular form that finds the shortest path between two pixels as user input. In this way, the user can easily cut the image into small pieces. With minimum manual intervention and a lot of computer assistance, high-quality results can be obtained.

Edge Detection
This segmentation model is based on edge extraction to identify the boundary regions that separate two or more adjacent classes.

The Method Based on Gradient Change
Edges usually appear on the boundary between two different regions. The gradient of two adjacent areas can be calculated to determine the area where the image intensity changes from a high value to a low value (or vice versa) to determine the image edge. Edge detection methods based on gradient change can be divided into two categories: one is to calculate the first derivative of the intensity between pixels, and the other is to calculate the second derivative. They usually detect the boundary according to the principle of finding the maximum value, minimum value, and zero-crossing point of the derivative: the commonly used methods for calculating the second derivative are Laplace and Marr-Hildreth (LOG) edge detection operators. Roberts, Sobel, Prewitt, Kirsch, and Canny edge detection operators belong to the method of calculating the first derivative.

Hough Segmentation
The Hough transform was initially designed to find straight lines. It transforms a line in the cartesian plane into a point in the polar plane. We know that a line is a set of adjacent edge points in the same direction. The algorithm will collect the set of edge points in the image and effectively find all the lines where these edge points are. This method is computationally intensive, impractical for real-time applications, and provides more information than most applications require. The advantage of the Hough transform is that it allows for parameters based on a set of edge points to reduce the search time and can be adjusted according to application requirements.

Now the Hough transform is extended to the recognition of objects of arbitrary shapes, mostly circles, and ellipses.

Region-Based Methods
The region-based segmentation method is a segmentation technology based on finding regions directly. There are two necessary forms of region-based extraction methods: one is region growth, which starts from a single pixel and gradually merges to form the

desired segmentation region; the other is to cut to the desired segmentation region step by step from the overall situation.

Region Growth Method
Region growth starts with a group of seed pixels representing different growing regions, and then merges the qualified pixels in the neighborhood of the seed pixels into the growth region served by the seed pixels, and continues the merging process with the newly added pixels as new seed pixels until the qualified new pixels are not found. The key to this method is to select the appropriate initial seed pixels and reasonable growth criteria.

The watershed method is formed by introducing the concept of geodesic distance into the regional growth method.

Split and Merge (Quadtree Method)
Region growth is to start from one or some pixels, and finally get the whole region, and then achieve the target extraction. Splitting and merging can be said to be the inverse process of region growth, starting from the entire image, always split to get each sub-region, and then merge the foreground regions to get the foreground target that needs segmentation, and then achieve the target extraction. Quadtree decomposition method is a typical region splitting and merging method.

Threshold-Based Method
The method of selecting one or more appropriate grayscale thresholds to separate objects from the background is usually divided into global thresholds and local thresholds.

Global Threshold
The global threshold method refers to using global information to find the visual threshold of the whole image, which can be a single threshold or multiple thresholds. The most commonly used threshold segmentation algorithm is the largest inter-class variance method (OTSU), which selects the optimal global threshold by maximizing the variance between classes. Besides, there are threshold segmentation methods based on entropy, minimum error method, symbiosis matrix method, moment preserving method, simple statistics method, fuzzy set method and threshold method combined with other methods. In the segmentation method of determining the threshold, the accuracy of image segmentation is greatly improved by introducing the Finite Mixture Model (FMM). Karl Pearson first proposed FMM in 1894. He constructed a Gaussian Mixture Model (GMM) with two univariate Gaussian distributions and estimated the parameter set of the model using the moment method.

In 1972, C.k. chowand, T. kaneko, segmented the left ventricle and aortic root with two normal distributions in the angiocardiography image.

The Local Threshold Method
The whole original image is segmented into several non-overlapping sub-images, and then the global threshold method is used to find the optimal segmentation threshold for each sub-image.
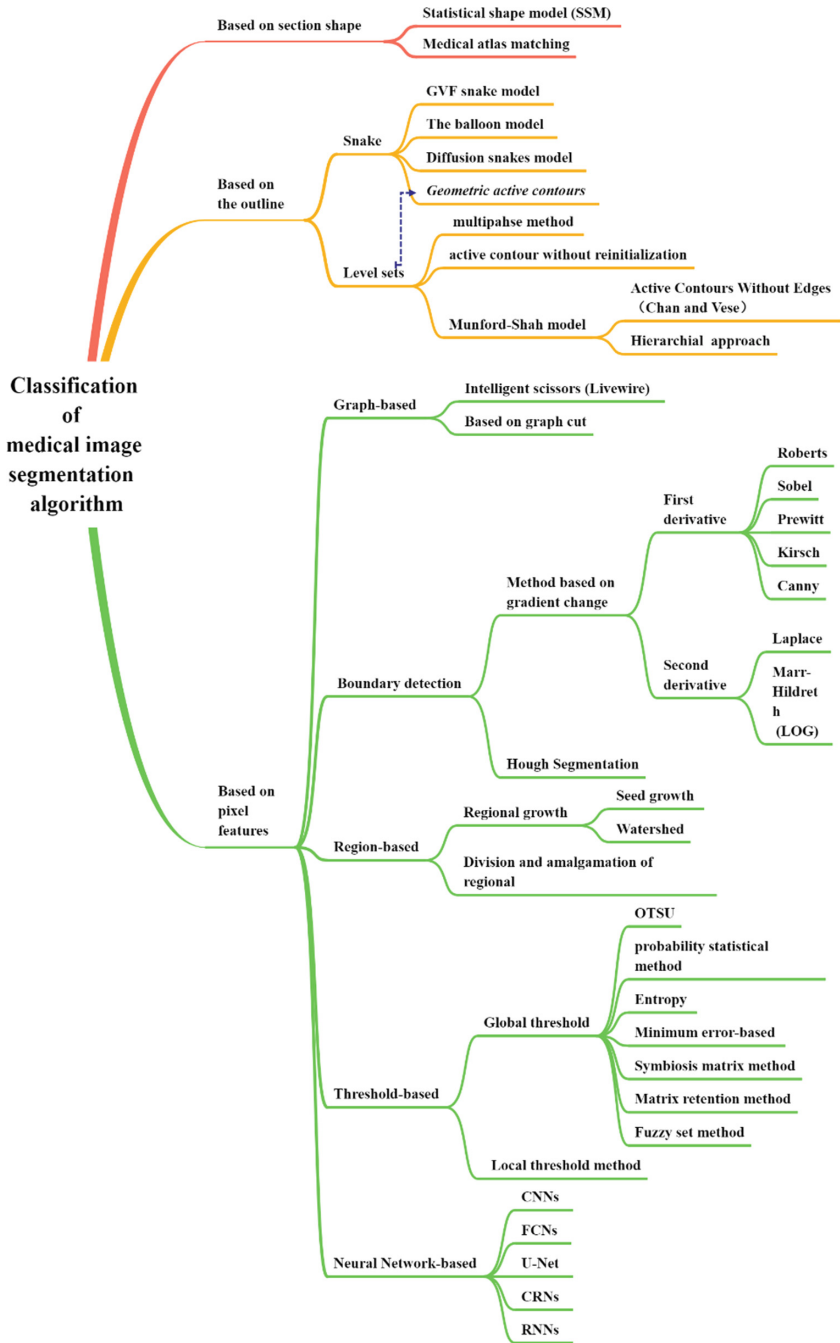
Fig. 3. Classification of medical image segmentation algorithm

Segmentation Method Based on Neural Network

To achieve image segmentation, imaging experts are required to determine the critical image feature information such as the shape and volume of organs according to their knowledge of the target domain. This complex and tedious task was completed manually in the past. Now, people use neural networks, especially deep learning methods, to achieve automatic feature extraction. According to the network structure, the current algorithm for medical image segmentation is Convolutional Neural Networks (CNN), Fully Convolutional Networks (FCN), U-Net, Convolutional Residual Networks (CRN), Recurrent Neural Networks (RNN). Mohammad Hesam Hesamian et al. had made a comprehensive overview of the algorithms of deep learning in image segmentation.

The general classification, according to the principle of classification based on the essential elements of plane space, the medical image segmentation algorithm, is shown in Fig. 3.

3   Conclusion

In order to summarize a large number of existing medical image segmentation algorithms and make it easy to clarify the relationship between various algorithms, many scholars have explored from multiple angles and described them in detail. Summing up the previous experience, this paper proposes to classify medical image segmentation algorithms from point, line, and surface levels, which helps categorize medical image segmentation algorithms from fine particles to coarse particles. The medical image segmentation algorithm can be understood as a whole from the micro point of view to the macro point of view. The hierarchical classification of point, line, and surface shows a distinct tree structure, and there is no intersection between the algorithms.

References

1. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. Annu. Rev. Biomed. Eng. **2**, 315–337 (2000)
2. Withey, D.J., Koles, Z.J.: Medical image segmentation: methods and software. In: 2007 NFSI-ICFBI, pp. 140–143 (2007)
3. Sharma, N., Aggarwal, L.M.: Automated medical image segmentation techniques. J. Med. Phys. **35**(1), 3–14 (2010)
4. Elnakib, A., et al.: Medical image segmentation: a brief survey. In: El-Baz, A., Acharya, U.R., Laine, A., Suri, J. (eds.) Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies, pp. 1–39. Springer, New York (2011)
5. Erdt, M., Steger, S., Sakas, G.: Regmentation: a new view of image segmentation and registration. J. Radiat. Oncol. Inform. **4**(1), 1–23 (2012)
6. Norouzi, A., et al.: Medical image segmentation methods, algorithms, and applications. IETE Tech. Rev. **31**(3), 199–213 (2014)
7. Zhou, L.Q., et al.: Artificial intelligence in medical imaging of the liver. World J. Gastroenterol. **25**(6), 672–682 (2019)
8. Aganj, I., et al.: Unsupervised medical image segmentation based on the local center of mass. Sci. Rep. **8**(1), 13012 (2018)

9. Voronin, V.V., et al.: Medical image segmentation by combing the local, global enhancement, and active contour model. In: Anomaly Detection and Imaging with X-Rays (ADIX) IV (2019)
10. Sakinis, T., et al.: Interactive segmentation of medical images through fully convolutional neural networks. arXiv e-prints (2019)

Pathological Changes Discover Network: Discover the Pathological Changes of Perivascular Dermatitis by Semi-supervised Learning

Xiaodong He[1], Yu Fu[2], Yingqiu Bao[2], Jianmin Chang[2], Yibo Xie[2], Weiping Li[1], and Jing Zhang[1,3(✉)]

[1] Bodhi Laboratory, Beijing BeYes Technology Co. Ltd., Beijing, China
`jane.zhang@be-yes.com`
[2] National Center of Gerontology, Beijing Hospital, Beijing, China
[3] Department of Biomedical Engineering, Tsinghua University, Beijing, China

Abstract. Deep learning method becomes increasingly popular in many fields, especially in computer vision area, and it is gradually being applied in the medical image area to solve medical image problems such as lesion segmentation, nucleus detection and disease classification. However, training deep learning segment models requires a lot of accurate pixel-level segmentation medical image labels, which is very hard to obtain. Lacking accurate pixel-level segmentation labels is an important factor restricting the development of deep learning method in medical image processing. In this paper, we propose a perivascular dermatitis classification network based on semi-supervised learning which is trained by image-level supervision. Our network can accurately and effectively segment the pathological changes contributed to the disease classification while classifying perivascular dermatitis. We use U-Net as the pathological change discover module to segment the pathological changes, propose the restricted boundary loss to improve the accuracy of the segmentation area boundary and introduce the pathological changes guided module to guide the pathological changes discover module to generate pathological changes contribute to the classification task. We evaluate our network on the dataset of skin pathological images with image-level classification labels. With the image-level labels of the perivascular dermatitis pathological images and normal skin pathological images, our network is trained to discover the pathological changes and classify the skin pathological images. Experiments show that our network discovers the skin pathological changes by using image-level classification labels. The perivascular dermatitis classification accuracy, AUC and PR of our network is 0.8793, 0.8954 and 0.8041.

1 Introduction

The rise of deep learning in computer vision and the application of deep learning methods in medical image processing provide technical support for solving medical image problems via deep learning methods. However, the training of deep learning models requires image datasets composed of a large number of accurately labeled images and labels, such as the PASCAL VOC image set [4] created by Mark Everingham, Luc van Gool, etc., with more than 17,000 images, 20 class annotated, the COCO dataset [8] which is the largest dataset for semantic segmentation, there are 80 categories, more than 330,000 images, of which 200,000 are labeled, and the number of individuals in the entire data set exceeds 1.5 million and ImageNet database [3] founded by Li Feifei, with more than 14,000,000 images, the sample images are divided into 27 categories and more than 20,000 categories.

In the application of deep learning methods in medical image processing, accurately lesion segmentation labels can effectively assist researchers in designing and training efficient and stable models [7]. However, it is very difficult to obtain a large number of medical images and accurate labels. The existing medical image datasets are much smaller than the natural image datasets. For example, LUNA dataset [12] contains 888 CT images, Kaggle-DR dataset [5] contains 35,126 training image and 53,576 testing images, IDRiD dataset [9] contains 516 image-level labeled fundus images and 81 pixel-level labeled fundus images and the ISIC 2017 dataset [2] contains 2,000 dermoscopy images with segmentation labels and 2,000 dermoscopy images with classification labels.

Obtaining a large medical image dataset with precise segmentation labels is extremely ambitious. The high cost of medical image acquisition and the lack of experienced doctors could tag the medical images are obstacles to building large and accurate medical image datasets. In this paper, we proposed a new deep neural network to obtain precise segmentation label by image-level classification labels.

There are many kinds of skin diseases and its clinical manifestations are various. Perivascular dermatitis is the main disease in clinical diagnosis. Because the clinical manifestations of skin rashes in many clinical cases lack specificity, it is necessary to use histopathological examination to assist the diagnosis of dermatitis.

Inflammatory skin diseases are divided into nine categories by Dr. A. Bernard Ackerman [1], involving about 50 different skin pathological changes. If you want to use deep learning and other methods to identify skin pathological changes, you need a senior doctor to mark a large number of skin pathological slices, which consumes a lot of doctor's effort and time. To solve this problem, we propose a pathological changes discover network that uses semi-supervised learning to assist deep learning models in finding and segmenting pathological changes.

In this paper, we introduce U-Net framework as the pathological changes discover module to segment the skin pathological changes. Drawing on the visualization result of Grad-CAM (Class Activate Mapping) [11], propose a restricted boundary loss that

refines the boundary of pathological changes. Use pathological changes guide module to guide pathological changes module to segment pathological changes, pathological changes guide module is a neural network consisting of 4 1-dimensional convolution layers and 1 full-connect layer, then use binary cross-entropy loss as a perivascular dermatitis classification loss. The results of the pathological changes module are the results of skin pathological changes. In this paper, the main contributions of our network are listed below:

(1). A pathological changes discover module based on U-Net is introduced to generate pathological changes from original skin pathological images.
(2). To improve the accuracy of pathological changes discover module generated results, a restricted boundary loss is proposed to enhance the accuracy of saliency analysis and obtain the much more accurate results of skin pathological changes.
(3). Introduce pathological changes guided module, which use classification loss to guide pathological changes discover module to learn pathological changes contribute to the perivascular dermatitis classification.

2   Proposed Methods

Our network consisted of a pathological changes discover module and a pathological changes guided module. The main framework of our network is shown as Fig. 1.
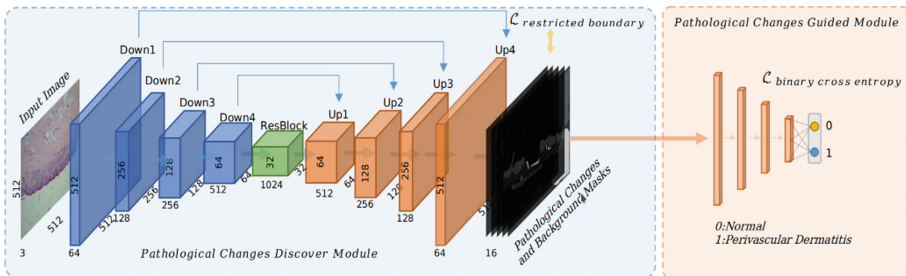
Fig. 1. The framework of pathological changes decovery network.

2.1   Pathological Changes Discover Module

The U-Net framework [10] is introduced as the pathological changes discover module to generate lesions and background masks from the original images. The pathological changes discover module could generate the pixel-level pathological change masks, we put the generated pathological change masks into pathological changes guided module, which could guide the pathological changes discover module generate pathological changes which contribute to perivascular dermatitis classification. The pathological changes discover module takes original skin pathological images as inputs, the distance map of the original images as the labels. We use the restricted boundary loss to minimize the distance between the edge of the predicted masks and the distance map of the original images.

2.2   Restricted Boundary Loss

We propose a restricted boundary loss to restrict pathological changes discover module to get more accurate segmentation result. The pipeline of restricted boundary loss is shown in Fig. 2. (The distance map generated method is clarified in the supplementary file)
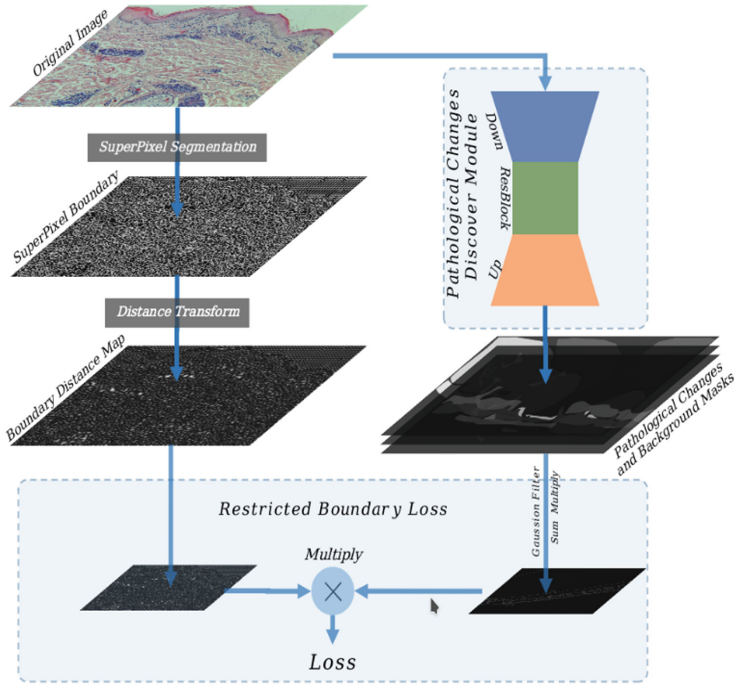
Fig. 2. The pipeline of restricted boundary loss

2.3   Pathological Changes Guided Module

In order to guide the pathological changes discover module to obtain pathological changes that could support network to distinguish normal and perivascular dermatitis images, we proposed pathological changes guided module. This module trained by the image-level labels based on semi-supervised learning methods. The main framework of our pathological changes guided module is shown in Fig. 3.

3   Experiments

3.1   Dataset and Preprocessing

Perivascular Dermatitis Dataset. We conducted a retrospective study and obtained images from the histopathology slides of 3500 patients visiting the dermatology department of Beijing hospital. We chose 160 patients histopathology slides to get the images,
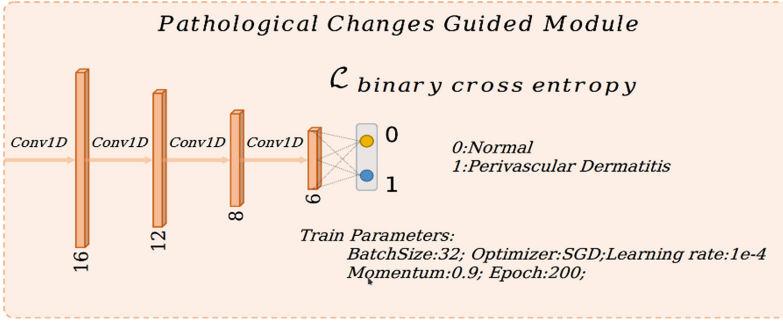
Fig. 3. The main framework of pathological changes guided module

including 86 cases of normal skin histopathology slides taking from the edge of beginning epidermal cysts slides, 74 cases of inflammatory skin diseases slides. There is a total of 1005 images including 4× images and 10× image. These images encompassed physician-patient encounters presenting from January 2016 to August 2019. The median age was 49.7 years (range from 18 to 82 years old); 77% were female.

Our perivascular dermatitis dataset contains 501 normal skin pathological images and 504 superficial perivascular dermatitis pathological images. The details of perivascular dermatitis dataset are shown in Table 1.

Table 1. The details of perivascular dermatitis dataset.

Dataset	Train	Test	Total
Normal	400	101	501
Perivascular dermatitis	400	104	504

Data Augmentation. We introduce a data augmentation strategy including random flips (horizontal flip, vertical flip), random rotations ranging from 0° to 359° random resizing by factors from 0.7 to 1.3 and random cropping the image to 512 × 512 to increasing data diversity.

3.2 Evaluation Metrics

Since the pathological changes discover module is trained by semi-supervised learning methods, it is really hard to evaluate the performance of our pathological changes discover module by some segmentation metrics, so we compare the discovered pathological changes with the result of Grad-CAM network. In our pathological changes guided module, we compute the accuracy, AUC and PR of perivascular dermatitis classification in the perivascular dermatitis dataset (detailed in Sect. 3.1) to compare the performance of our network.

Classification. We compute the accuracy, AUC and PR of perivascular dermatitis classification on the perivascular dermatitis dataset to compare the performance if our network. The higher the metrics of the testing set is, the higher the accuracy of the model is; the smaller the metrics gap between the training set and testing set is, the stronger the generalizability of the network is.

3.3 Pathological Changes Analysis

Since there is no groud-truth and pixel-level label for the pathological changes discover module, we compare the pathological change masks generated by pathological changes discover module with the output of Grad-CAM network to evaluate the performance of our methods.

Grad-CAM Visualization. We use Grad-CAM network based on ResNet34 [6] to localize class-discriminative regions for perivascular dermatitis. Grad-CAM visualization results are shown in Fig. 4.

(a) Original Image   (b) Grad − CAM 'Pathological Changes'   (c) Discover 'Lymphocytic Infiltration'   (d) Discover 'Incomplete Keratosisgical'   (e) Discover 'Acanthosis'
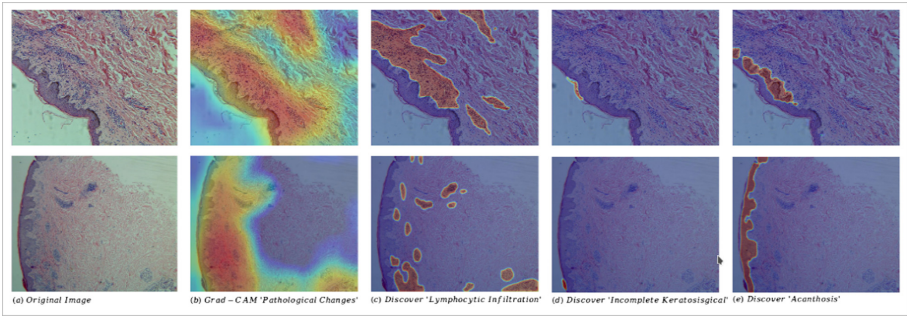
Fig. 4. Results for Grad-CAM and pathological changes discover module, (a) Original pathological image with perivascular dermatitis. (b) Grad-CAM: localizes class discriminative regions from original skin pathological images. (c–e) Discovered pathological changes from the pathological changes discover module: incomplete keratosis, acanthosis, lymphocytic infiltration.

Segmentation Results with Restricted Boundary Loss. With the comparison between Grad-CAM visualization result and discovered pathological changes in Fig. 4, we find the discovered pathological changes (incomplete keratosis, acanthosis, lymphocytic infiltration) are more accurate and clearer.

3.4 Perivascular Dermatitis Classification

The AUC, PR and accuracy of the perivascular dermatitis dataset are detailed in Table 2.

Table 2. Perivascular dermatitis classification results.

Dataset	Count	AUC	PR	Accuracy
Training set	800	0.9459	0.9784	0.9619
Testing set	205	0.8954	0.8041	0.8793

4 Conclusion

In this paper, we use semi-supervised learning methods to train our network, which could discover pathological changes by only using image-level classification tags. Precisely pathological changes such as incomplete keratosis, acanthosis and lymphocytic infiltration have been discovered by our network. The accuracy, AUC and PR of our classification result is 0.8793, 0.8954 and 0.8041. Our experiment indicates that the pathological changes discover network based on semi-supervised learning methods worked, our network could discover pathological changes contribute to perivascular dermatitis classification, the pathological changes discover module also generate precisely pathological change masks. What we should do next is to improve our network and enlarge the perivascular dermatitis dataset to get more accurate and precise pathological changes.

References

Ackerman, A.B., Termin, L.: Histologic Diagnosis of Inflammatory Skin Diseases: A Method by Pattern Analysis. Lea & Febiger, Philadelphia (1978)

Berseth, M.: Isic 2017-skin lesion analysis towards melanoma detection. arXiv preprint arXiv:1703.00523 (2017)

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. IEEE (2009)

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)

Graham, B.: Kaggle diabetic retinopathy detection competition report, University of Warwick (2015)

He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, 2015. arXiv preprint arXiv:1512.03385 (2018)

He, X., Zhou, Y., Wang, B., Cui, S., Shao, L.: Dme-net: diabetic macular edema grading by auxiliary task learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 788–796. Springer, Cham (2019)

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll´ar, P., Zitnick, C.L.: Microsoft coco: common objects in context. In: European conference on computer vision, pp. 740–755. Springer, Cham (2014)

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Meriaudeau, F.: Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research. Data **3**(3), 25 (2018)

Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer, Cham (2015)

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: visual explanations from deep networks via gradient-based localization (2016)

Setio, A.A.A., Traverso, A., De Bel, T., Berens, M.S., van den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geurts, B., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. Med. Image Anal. **42**, 1–13 (2017)

Optical Micro-scanning Reconstruction Technique for a Thermal Microscope Imaging System

Mei-Jing Gao[1(✉)], Zhu Liu[2], Liu-Zhu Wang[1], Bo-Zhi Zhang[1], and Shi-Yu Li[1]

[1] The Key Laboratory for Special Fiber and Fiber Sensor of Hebei Province, School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China
gaomeijing@126.com
[2] China Petroleum Pipeline Engineering Co. Ltd. Southeast Asia, Langfang City 065000, Heibei Province, China

Abstract. Thermal microscope imaging system can achieve the detection of subtle targets by temperature changes, which has extensive utility value in some respects that need detailed thermal analysis. Due to the mechanical preparation of the system, working vibration, external environment and other factors of interference, which will affect the precision of the micro-scanning system, Causing the micro-displacement of the four images acquired by the system deviate from the standard position. Therefore affecting the quality of subsequent image reconstruction. To minimize the impact of these factors, a micro-scanning reconstruction technique based on local gradient interpolation was proposed. The simulation and experiment demonstrate that the method can decrease the system micro-scanning error and enhance image reconstruction quality.

Keywords: Thermal microscope imaging system · Optical micro-scanning · Micro-displacement · Local gradient interpolation

1 Introduction

Since the 1990s, with the military field of strong demand for night vision and harsh environmental conditions, thermal imaging technology has developed rapidly [1–3]. However, these thermal imaging systems are mostly teleworking modes, but some practical applications often require the use of an optically amplified microscopic pattern to analyze the sample's fine temperature distribution. Compared with the visible light microscope, a thermal microscope imaging system can achieve the detection of subtle targets by temperature changes. Therefore the systems have broad research prospects in the object thermal analysis domain. In the last few years, the thermal microscope imaging system has been rapidly developed and different types of systems have been emerged [4–7]. To enhance its spatial resolution and meet actual needs, introduction of uncooled focal plane detectors into thermal microscope imaging system [8, 9], we also have developed a micro-scanning system and realized the integrated design of the system [10–12] (Fig. 1).

Fig. 1. Laboratory new thermal imaging system

Four images of the 2 × 2 micro-scanning A(i, j), B(i, j), C(i, j) and D(i, j), are obtained (i = 1, 2…, M; j = 1, 2…, N) when the tilted optical plate is rotated at 90° from the starting point [13]. According to the image acquisition sequence, Four images need to be interpolated to form a high-resolution image. The reconstruction image is increased by fourfold for the spatial resolution.

On account of the mechanical preparation of the system, working vibration, external environment factors of interference, the four images acquired by the system are not normal under-sampling images. That is, the micro-displacement between them is not standard 2 × 2 micro-scanning mode, influencing the quality of image reconstruction. Therefore, To decrease the system error, based on the micro-scanning technology and combined with the local gradient interpolation algorithm, a new reconstruction technique is proposed. The structured about paper as follows. In Sect. 2, we introduced the local gradient interpolation reconstruction model and the reconstruction technique. In Sect. 3 and Sect. 4, We conducted simulation studies and data analysis. In Sect. 5, We summarize the algorithm proposed in this paper and illustrated the superiority of this method.

2 Different Interpolation Reconstruction Models

2.1 Local Gradient Interpolation Reconstruction Model

In the image, the variance reflects the contrast of the local area. Therefore, when we deal with infrared images, the gradient G of the image can be calculated using local variance. Assume that the size of the image is M × N. The pixel gray value of the coordinate point (i, j) is $x_{i,j}$. The d × d window centered on the pixel of the coordinate is W (d >= 3 and it is odd). Sliding the window W moves in the pixel range, it covers the normal deviation $v_{i,j}$ of the center of the area. Calculated as follows (1), (2):

$$v_{i,j} = \frac{1}{d}\sqrt{\sum_{p=i-\frac{d-1}{2}}^{i+\frac{d+1}{2}}\sum_{q=j-\frac{d-1}{2}}^{j+\frac{d+1}{2}}(x_{p,q} - \overline{m}_{i,j})^2} \tag{1}$$

$$\overline{m}_{i,j} = \sum_{p=i-\frac{d-1}{2}}^{i+\frac{d+1}{2}}\sum_{q=j-\frac{d-1}{2}}^{j+\frac{d+1}{2}} x_{p,q} \tag{2}$$

Where

$X_{p,q}$ - the pixel value of the point (p,q) in window W;
$\bar{m}_{i,j}$ - The mean value of the pixels contained in the sliding window W [14–16].

From the formula (1), the gradient image G can be obtained. The gradient of position (i, j) is defined as

$$g_{i,j} = v_{i,j} \tag{3}$$

The local variance $v_{i,j}$ reflects the degree of gray change in the image. Because it is associated with multiple pixels within the sliding window, it is insensitive to single-point noise. The larger the sliding window, the more obvious the noise suppression will be. However, as the window area increases, the amount of computation will increase exponentially. Therefore, choosing a reasonably sized window is the key to the algorithm. In the real-time considerations, a window with a size of 3×3 is selected in this paper.

2.2   Interpolation Model Based on Gradient Threshold

The purpose of image interpolation is to establish a meaningful resampling method to improve image resolution. The inverse process for the linear sampling process is

$$g(x, y) = \sum_{i=0}^{M} \sum_{j=0}^{N} y_{i,j} \varphi(x - x_i, y - y_j) \tag{4}$$

Where g(x, y) refers to the reconstructed high-resolution image, $\varphi(x, y)$ refers to the interpolation function.

The gradient of the target pixel is calculated according to the difference of the heat radiation between the different targets and their parts (the gray level in the corresponding image). When the gray gradient value is smaller than the threshold which is set in this paper. It is located in the smooth area of the image. The pixel value of the pixel point can be calculated by the grayscale weighting of the neighboring pixels. If it is greater than the threshold, it is proved that it is located in the edge region. The gradient of the largest pixel value replaces the pixel grayscale value. The formula is shown in (5):

$$\begin{cases} f_{i,j} = a_{i,j-1}f_{i,j-1} + a_{i,j+1}f_{i,j+1} + a_{i-1,j}f_{i-1,j} + a_{i+1,j}f_{i+1,j}, & g_{i,j} < T \\ f_{i,j} = f_M, & g_{i,j} > T \end{cases} \tag{5}$$

$$M = \max(|f_{i,j-1} - f_{i,j}|, |f_{i,j+1} - f_{i,j}|, |f_{i-1,j} - f_{i,j}|, |f_{i+1,j} - f_{i,j}|)$$

In the above formula:

T - the gradient threshold;
M - the norm of derivative in the maximum gradient direction;
$f_{i,j}$ - the gray value of the pixel (i, j) in the coordinate position;

$a_{i,j+1}$, $a_{i,j-1}$, $a_{i+1,j}$, $a_{i-1,j}$ - the weighting factor for each direction;
$|f_{i-1,j} - f_{i,j}|$ - the gray value difference between the pixel $(i-1, j)$ and the pixel (i, j).

among them:

$$a_{i,j-1} = \frac{|f_{i,j-1} - f_{i,j}|}{\sum_{k=\pm 1} |(f_{i,j+k} - f_{i,j})| + \sum_{k=\pm 1} |(f_{i+k,j} - f_{i,j})|} \qquad (6)$$

$$a_{i-1,j} = \frac{|f_{i-1,j} - f_{i,j}|}{\sum_{k=\pm 1} |(f_{i,j+k} - f_{i,j})| + \sum_{k=\pm 1} |(f_{i+k,j} - f_{i,j})|} \qquad (7)$$

$$a_{i+1,j} = \frac{|f_{i+1,j} - f_{i,j}|}{\sum_{k=\pm 1} |(f_{i,j+k} - f_{i,j})| + \sum_{k=\pm 1} |(f_{i+k,j} - f_{i,j})|} \qquad (8)$$

$$a_{i,j+1} = \frac{|f_{i,j+1} - f_{i,j}|}{\sum_{k=\pm 1} |(f_{i,j+k} - f_{i,j})| + \sum_{k=\pm 1} |(f_{i+k,j} - f_{i,j})|} \qquad (9)$$

2.3 Micro-scanning Error Correction Technique Based on Local Gradient Interpolation Reconstruction

The low-resolution image group $\{Y_K\}_1^4$ is acquired by the system, which is $Y1 = A$, $Y2 = B$, $Y3 = C$, $Y4 = D$. Each image size is $M \times N$. The technique uses $\{Y_K\}_1^4$ and combines with the local gradient weighting polynomial to decrease the system micro-scanning error, Thereby obtaining high-resolution images G. The specific process is (Fig. 2):
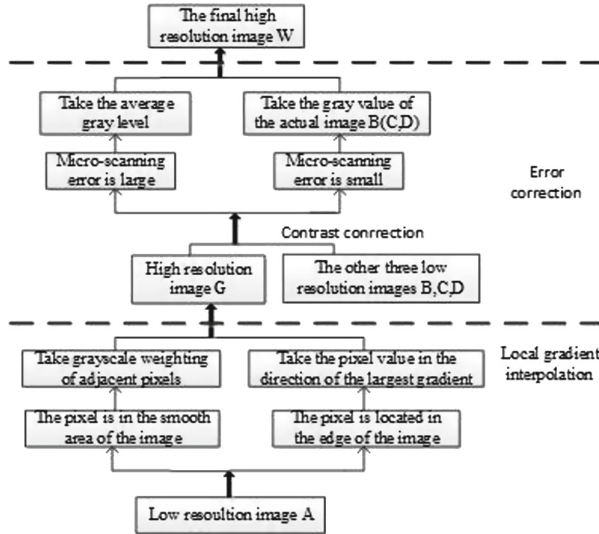
Fig. 2. Algorithm flow chart

(1) Collected four images (A, B, C, D) through the system, Selecting one of them as a base image. Assumption image A is regarded as the base image. The gray value of the coordinate pixel point (i, j) in the pixel array A is assigned to the coordinate position point $(2i, 2j)$ of the high-resolution image G, where $i = 1, 2, \cdots, M$, $j = 1, 2, \cdots N$;

(2) The A is amplified by a weighted polynomial method based on the local gradient feature. The gray value of the coordinate pixel point $(2i + 1, 2j)$, $(2i, 2j + 1)$, $(2i + 1, 2j + 1)$ in the amplified image is sequentially assigned to the pixel of the corresponding position in the high-resolution image G. Thereby a new high-resolution image G is obtained.

(3) The image G is corrected. Contrasting the pixel gray value $(2i + 1, 2j)$ in G and the pixel gray value (i, j) in the low-resolution image B. If the discrepancy is much smaller, it is obvious that the error also is small. As a result, the pixel gray value of $(2i + 1, 2j)$ in G is substituted by the pixel gray value of (i, j) in image B. If the difference between the two is relatively large, it is proved that the error is relatively large. The pixel gray value of $(2i + 1, 2j)$ in G is replaced by the mean of the two.

(4) The gray value of the coordinate pixel point $(2i, 2j + 1)$, $(2i + 1, 2j + 1)$ in the image G is the same as above. Finally, the corrected high-resolution image W can be acquired.

The image A, B, C, D acquired by the system is processed according to the above specific steps. The reconstructed image W can be obtained.

3   Simulation Research

First, the rationality of the method is verified by visible image simulation. The standard Lena image (256×256 in size) is used to obtain the downsampled images (size 128×128) with standard half-pixel displacement. The simulation and experiment are carried out.

The Lena image is processed by bilinear interpolation, direct interpolation and the technique proposed in this paper. These high-resolution images were obtained by simulation, as shown in Fig. 3. Comparing these traditional interpolation methods with the new correction algorithm, the image quality parameters were evaluated, as shown in Table 1.
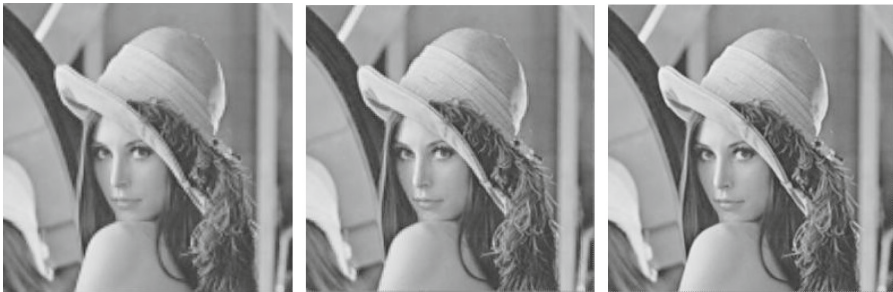
(a) Bilinear interpolation   (b) Direct interpolation   (c) The technique proposed

Fig. 3.   Simulation results

Table 1. Image quality evaluation parameters

Image quality parameters	MSE	PSNR	Q	SNT
Bilinear interpolation	108.62650	27.42400	0.99564	7.37620
Direct interpolation	104.25970	27.60220	0.99585	7.39110
The technique proposed	86.29570	28.42340	0.99744	7.39240

Observing the hair in Fig. 3, we can see that the image processed by the bilinear interpolation algorithm and the direct interpolation are not obvious compared with the technology proposed in this paper. From the Table 1, the MSE of the technology proposed in this paper is minimum, PSNR and Q are the largest. So it is closer to the original image. The image quality of this technique is higher than the image quality of bilinear interpolation and direct interpolation, so it is better than the two interpolation methods.

But in the actual micro-scanning, there are certain micro-displacement errors. The position of micro-displacement is not a strict orthographic square. In this paper, the four low-resolution images obtained by nonstandard displacement of visible light images (Lena) are simulated. In this paper, we choose the simulation results of micro-displacement (0.00, 0.00), (0.29, 0.68), (0.50, 0.52), (0.52, –0.03).

(a) Bilinear interpolation   (b) Direct interpolation   (c)The technique proposed

Fig. 4. Simulation results

Table 2. Experimental parameters of four groups of micro-displacements

Micro-displacement coordinate value	Interpolation algorithm	PSNR	MSE	SSIM	Q
(0.00, 0.00), (0.06, –0.68) (0.92, –0.62), (0.83, 0.08)	Bilinear interpolation	21.57090	418.06270	0.67514	0.96471
	Direct interpolation	23.17040	289.26450	0.74955	0.98914
	The technique proposed	25.35880	174.76550	0.78922	0.99460
(0.00, 0.00), (0.14, –0.57) (0.53, –0.55), (0.60, 0.06)	Bilinear interpolation	23.18060	288.58240	0.75545	0.98133
	Direct interpolation	24.24230	225.99550	0.80474	0.99298
	The technique proposed	26.00020	150.77050	0.82626	0.99607

(*continued*)

Table 2. (*continued*)

Micro-displacement coordinate value	Interpolation algorithm	PSNR	MSE	SSIM	Q
(0.00, 0.00), (0.04, −0.41) (0.49, −0.47), (0.46,0.02)	Bilinear interpolation	23.67000	257.82980	0.77630	0.98495
	Direct	25.17380	182.36950	0.84196	0.99535
	The technique proposed	26.52020	133.75480	0.85156	0.99711
(0.00, 0.00), (0.50, 0.52) (0.52, −0.03), (0.29, 0.68)	Bilinear interpolation	21.81390	395.31250	0.71014	0.96549
	Direct interpolation	26.53950	133.16330	0.89389	0.99411
	The technique proposed	28.06860	93.64320	0.90046	0.99677
(0.00, 0.00), (−0.01, −0.51) (0.41, −0.58), (0.46, 0.07)	Bilinear interpolation	24.11030	232.97050	0.80340	0.98667
	Direct interpolation	24.69730	203.51700	0.82193	0.99452
	The technique proposed	26.27830	141.41860	0.83882	0.99676

The simulation results are shown in Fig. 4 and Table 2. From the details of Fig. 4, the proposed technique decreases the error and improves the spatial resolution. It can be seen from the data in Table 2 that the MSE of the technique proposed in this paper is the smallest and the difference from the original image is the smallest. PSNR is the largest and the recovery effect is the best. Q and SSIM are the closest to 1, which is the most similar to the original image. It is concluded that the technique proposed in this paper achieves a better optimization effect and improves the quality of imaging. Four non-standard micro-scanning images of infrared building images were obtained by a non-standard micrograph, as shown in Table 3. In this paper, we choose the simulation results of micro-displacement (0.00, 0.00),(0.29, 0.68),(0.50, 0.52),(0.52, −0.03) for visual comparison, as shown in Fig. 5:

(a) Bilinear interpolation   (b) Direct interpolation   (c) The technique proposed

Fig. 5. Simulation results

Table 3. Experimental parameters of four groups of micro-displacements

Micro-displacement coordinate value	Interpolation algorithm	PSNR	MSE	SSIM	Q
(0.00, 0.00), (0.06, –0.68) (0.92, –0.62), (0.83, 0.08)	Bilinear interpolation	22.26090	386.35750	0.67304	0.96928
	Direct interpolation	24.30380	241.38070	0.79530	0.99141
	The technique proposed	26.15570	157.58360	0.83032	0.99561
(0.00, 0.00), (0.14, –0.5) (0.53, –0.55), (0.60, 0.06)	Bilinear interpolation	24.09690	253.15540	0.77694	0.98392
	Direct interpolation	25.44310	185.68030	0.85048	0.99462
	The technique proposed	26.75520	137.26510	0.86554	0.99669
(0.00, 0.00), (0.04, –0.41) (0.49, –0.47), (0.46, 0.02)	Bilinear interpolation	24.53540	228.84410	0.79606	0.98657
	Direct	26.30980	152.09140	0.88232	0.99643
	The technique proposed	27.24300	122.68180	0.88743	0.99748
(0.00, 0.00), (0.50, 0.52) (0.52, –0.03), (0.29, 0.68)	Bilinear interpolation	22.24990	387.33920	0.69122	0.97010
	Direct interpolation	26.84410	134.48380	0.90145	0.99469
	The technique proposed	27.93160	104.69370	0.90672	0.99674
(0.00, 0.00), (–0.01, –0.51) (0.41, –0.58), (0.46, 0.07)	Bilinear interpolation	24.52840	229.21540	0.79814	0.98663
	Direct interpolation	25.83450	169.68170	0.86987	0.99598
	The technique proposed	27.10330	126.69340	0.88145	0.99741

From the details in Fig. 5 and the data in Table 3, it can be seen that MSE of the technique proposed is the smallest, the SSIM and Q are closer to 1, compared with the bilinear interpolation and direct interpolation. The technique proposed achieves the requirement of decrease error correction and improving the spatial resolution, which makes the reconstructed image have higher resolution and better effect.

4   Experimental Research

To explain the rationality of the technique in the actual thermal imaging system, after the zero point calibration of the system, the zero point is used as the starting point. four low-resolution images were acquired by the system, and the acquired images were experimentally studied. The results as follow Fig. 6.

It can be seen from the high-resolution images processed by various algorithms and their parameters that the information entropy of the technique proposed is the largest, which contains the most information.
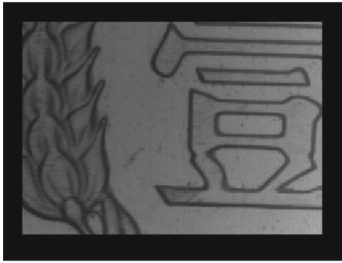
(a) Bilinear interpolation
(SNT=5.8171)
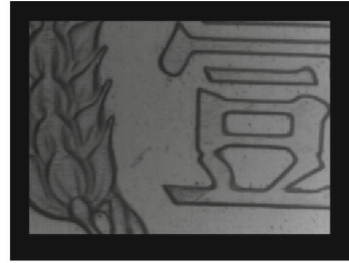
(b) Cubic spline interpolation
(SNT=5.8173)

c) Direct interpolation
(SNT=5.8311)

d) The technique proposed
(SNT=5.8409)

Fig. 6. Experimental results

5   Conclusion

For the design of the optical micro-scanning thermal microscope imaging system has errors during operation and the system spatial resolution needs to be further enhanced, the technique is proposed. The technique first uses the micro-scanning system to collect four downsampled images. Then compared with the reconstruct pixel obtained by the local gradient interpolation algorithm, the system errors generated by micro-scanning are compensated. It is proved by experiments that the technology can significantly improve the image quality, and effectively suppress the image blur caused by micro-scanning error, improving the spatial resolution of the imaging system. The technology can also be used in other photoelectric imaging systems.

References

1. Zhou, L.: Photoelectronic imaging: towards the new century. J. Beijing Inst. Technol. **22**(1), 1–12 (2002). (in Chinese)
2. McIntosh, L.M., Mansfield, J.R., Jackson, M., Neil Crowson, A., Mantsch, H.H.: Infrared imaging of skin lesions. Proc. SPIE **4577**, 226–229 (2002)
3. Jin, W., Hou, G., Liu, G.: Uncooled FPA thermal imaging technology and its application. Infrared Technology **20**(6), 6–11 (1998). (in Chinese)
4. Suriani, M.J., Ali, A., Khalina, A., Sapuana, S.M., Abdullah, S.: Detection of defects in Kenaf/Epoxy using infrared thermal imaging technique. Procedia Chem. **4**(12), 172–178 (2012)
5. Morikawa, J., Hayakawa, E., Hashimoto, T.: Application of micro-scale thermography to the thermal analysis of polymeric and organic materials. Proc. SPIE **8013**, 1–6 (2011)
6. Ankur, J.N., Goodson Kenneth, E.: Thermal microdevices for biological and biomedical applications. J. Thermal Biol. **36**, 209–218 (2011)
7. Morikawa, J., Hayakawa, E., Hashimoto, T.: Micro-scale thermal imaging of organic and polymeric materials with cooled and uncooled infrared cameras. Adv. Opt. Technol. **2012**, 1–8 (2012)
8. Gao, M., Jin, W.: Digital thermal microscope for biomedical application. In: 2007 IEEE/ ICME International Conference on Complex Medical Engineering, pp. 1878–1882 (2007)
9. Gao, M., Jin, W., Wang, X., et al.: Study on the noise equivalent temperature difference and noise equivalent eradiation difference mathematical models for micro thermal imaging systems. Trans. Beijing Inst. Technol. **27**(1), 50–54 (2007)
10. Gao, M., Jin, W., Wang, X., et al.: Design and implementation of optical micro-scanning thermal microscope imaging system with high resolution. Chin. J. Sci. Instr. **30**(5), 1037–1041 (2009)
11. Gao, M., Jin, W., Wang, X., et al.: Inter-Frame difference oversample reconstruction for optical microscanning thermal microscope imaging system. Trans. Beijing Inst. Technol. **29**(8), 705–707 (2009)
12. Gao, M., Jie, X., Tan, A., Zhenglong, Z., Yang, M., Wang, J.: Error correction based on micro-scanning preprocessing for an optical micro-scanning thermal microscope imaging system. Infrared Phys. Technol. **83**, 252–258 (2017)
13. Gao, M., Jin, W., Wang, X., Jie, Yu.: Zero calibration for the designed microscanning thermal microscopic imaging system. Acta Optica Sinica **29**(8), 2175–2179 (2009)
14. Bai, J.-q., Chen, Q.: Infrared microscanning imaging technique based on locally gradient feature. Acta Photonica Sinica **37**(11), 2253–2256 (2008)
15. Han, Y., Lenan, W.: Super resolution reconstruction of video sequence based on adaptive filter. Chin. J. Comput. **29**(4), 642–647 (2006)
16. Gao, M., Wang, J., Wei, X., Guan, C.: High-resolution over-sampling reconstruction algorithm for a microscanning thermal microscope imaging system. Infrared Phys. Technol. **76**, 661–666 (2016)

A Survey for Traditional, Cascaded Regression, and Deep Learning-Based Face Alignment

Kun Wang[✉] and Guosheng Zhao

Wuhan University, Hubei, China
`whuwangkun@163.com`

Abstract. The task of face alignment is to automatically locate key facial feature points, such as eyes, nose tips and mouth corners. There are many methods to solve the face alignment problems. These methods are classified into three categories: traditional algorithms, cascade based on regression algorithm and deep learning-based algorithm. There are still some problems existing that processing of large pose and small pose, processing of different angles and facial occlusion. The experiment results of three categories of algorithms are summarized in this paper.

Keywords: Face alignment · Cascaded regression · Deep learning · Face detection

1 Introduction

Face alignment detects eyes, nose tip and mouth corners in the true position, which is the pretreatment step of face recognition. This technology is widely used in many fields related to facial tasks. For example, this technology has been used in face recognition [1] and expression analysis [1]. Due to the influence of different postures, expression, illumination and occlusion, it is hard to locate the key feature points preciously. At present, there do exist several mainstream methods to solve these problems. In this paper, several methods are classified into three categories and are compared the results.

The three categories are that tradition algorithm, such as Supervised Descent Method (SDM) [2], cascade regression-based algorithm [3] and deep learning-based algorithm [4, 5]. The first category of the methods is traditional methods, Supervised Descent Method (SDM). This method is aimed to minimize a Non-linear Least Squares (NLS) functions. SDM learns descent directions in the training process and uses the descent directions, which have been learnt before to minimize the NLS in the testing process. SDM solve the drawbacks of 2nd order descent method in the context of computer vision. The second category of the method is the cascade regression-based algorithm. Cascade regression-based algorithm regrades face alignment as learning a regression function, which takes images as input and feature points (face shape) as output. The so-called cascade means the input of the current function depends on the output of the previous function and the learning objective of each function is to approach the real position of the feature point. The key of the cascade regression-based algorithm is that shape correlation

features are used and the optimization objective of function is shape residual, which is existing between the current shape and the real position shape. The third category of the method is the deep learning-based algorithm. This paper mainly introduces a method of face alignment based on three-level Deep Convolution Neural Network (DCNN) [4] and Mnemonic Descent Method (MDM) [6], which is a recurrent applied process in end-to-end face alignment. In three-level DCNN, the first level use three regions (face, eyes and nose tip, mouse corners) as input and train three convolution neural networks to predict the location of feature points. The latter two levels extract features near each feature point and aim to train a convolutional neural network separately for every single feature point to correct location results. The method of deep learning-based method can obtain better position results for large posture and several kinds of expressions. MDM proposes a non-linear framework, which is used cascaded method, to solve the problems about end-to-end learning of the descent direction of functions that are not linear. In MDM, a single model is usually trained in the way called end-to-end, which is from the initial image pixel points to the next predictions. MDM introduces the concept of merging the descent direction learning in different levels, which is highly discriminative and the major strength of this method.

The experiments are tested on several datasets, such as 300-W dataset [7], HELEN dataset [8], LFPW dataset [9], AFLW dataset [10]. The results of the alignment accuracy are evaluated for every sample, which is based on the mean error of the landmarks, which is normalized by the distance called inter-pupil and are reported by overall accuracy based on curves, which one we usually used is cumulative error distribution (CED). Besides, the further statistics are calculated from CED curves, for example, area-under-the-curve (AUC) or the failure rate of each method, which are effectively presented error metrics for the problem which are existing in face alignment.

2 Algorithm

2.1 Traditional Algorithms

This section describes Supervised Descent Method (SDM) [2] in the context of face alignment. SDM is based on minimizing a Non-linear Least Squares (NLS) functions. The purpose of SDM is to learn regression matrix R_0. In the supervised setting, the regression matrix can be estimated from the process of regression $\Delta x_* = x_* - x_0$ and feature vector $\emptyset_0$ which can be fixed during the optimization process. The function of SDM Eq. 1 can be learned as a generic linear function composed by regression matrix R_0, feature vector $\emptyset_0$ and bias b_0 from training process.

$$\Delta x_1 = R_0 \emptyset_0 + b_0 \tag{1}$$

Due to the problems of non-quadratic functions, SDM will learn a list of generic regression matrix $\{R_k\}$ and bias $\{b_k\}$ and the feature vector $\emptyset_{k-1}$ will be extracted from the previous landmark location x_{k-1}. The improved function Eq. 2 can be obtained by updating the new parameters and new feature vector and learning the new regression matrix and bias from the new training data set, such that the landmark x_k will be converged to x_* during the training process.

$$x_k = x_{k-1} + R_{k-1} \emptyset_{k-1} + b_{k-1} \tag{2}$$

In former works, it was unknown what the use of alignment loss function is that for discriminatory methods. This study figures out Eq. 3, which is the loss function minimized by discriminative methods, and has connection with Parameterized Appearance Models (PAMs). The loss function Eq. 3 can be obtained as followed. During the process of regression, the prediction and features $\left\{\Delta X_*^i, \emptyset_k^i\right\}$ can be updated by previous learned R_{k-1} and b_{k-1}. Thus, x_k can be updated using Eq. 2. Besides, R_k and b_k can be trained from the regression process and loss function by minimizing Eq. 3.

$$\arg\min_{R_k, b_k} \sum_{d^i} \sum_{X_k^i} \left\| \Delta X_*^{ki} - R_k \emptyset_k^i - b_k \right\|^2 \tag{3}$$

The error of this process can be decreased by increasing the number of regression times. The method of SDM only uses one-step regression and the usage of boosted regression is to approximate the function that maps from $\emptyset$ to Δx_1.

SDM learns descent directions in the training process and the usage of learned directions is to minimize the NLS in the testing process. SDM solves the drawbacks of 2nd order descent method in the context of computer vision.

2.2 Cascade Regression-Based Algorithm

The cascade regression-based algorithm proposes a face alignment structure based on coarse-to-fine shape search (CFSS) [3]. CFSS is in the beginning of a coarse search of the shape space containing different shapes, and uses a tough solution to constrain the precious shape in the subsequent search, which is a method different from traditional cascade regression method. This algorithm has its own advantage, the step-by-step and adaptive search. This advantage represents that this algorithm will prevent the final solution from local optima due to poor initialization which is the popular problem in the cascade regression. Besides, this will also increase the robustness of the large face-pose situation.

In this algorithm, the method uses the way of coarse-to-fine shape search. First step of the algorithm is to do the coarse search carried out in the shape space containing large number of candidate shapes. Second step is to identify the sub–region in the shape space and do the fine search in the fine sub-regions which are chosen. Meanwhile, the non-chosen sub-regions will be ignored for the latter finer and adaptive search. In the last step, the final shape space will converge to a small region.

Compared with the traditional methods, CFSS is attractive in the two aspects. First, CFSS is independent on the initialization, which means this method will not be trapped in the local optima. This method searches the whole shape space in the first step instead of aiming at the particular initialization point. So that this method can realize the robust face alignment. Second, the method is also suitable to large-pose variation. This method uses the formula to consider different face-pose changes at the same time. Then, the method will search the sub-regions to find the suitable space. The result of the experiments represents that the robustness of CFSS is better than the traditional algorithms.

The main part of CFSS is to learn to estimate the sub-region center. This part will conduct three steps. In the first step, this method explores larger regions in shape space under the guidance of probability distribution. The second step is to return each initial

shape to a shape closer to the real shape of the ground. The model r_k and the shape x_k after k^{th} iteration will be estimate by Eq. 5.

$$r_k = \arg\min_{r} \sum_{i=1}^{N} \sum_{j=1}^{N_l} \left\| X^{i*} - X_k^{ij} - r\left(\emptyset\left(X_k^{ij}\right)\right) \right\|_2^2 + \emptyset(r),$$

$$X_{k+1}^{ij} = X_k^{ij} + r_k\left(\emptyset\left(X_k^{ij}\right)\right) \quad k = 0, \ldots, k_l - 1 \tag{5}$$

This loss function will estimate the model r_k by the shape x_k, and the shape of $k+1^{th}$ iteration x_{k+1} will be obtained via the model r_k by the shape x_k by last iteration. Besides, k_l is less than the times of iterations, which are used in the regression process. Therefore, the predicted shape in the early stage does not need to be precise, and in the later stage, the initial candidate shape tends to be similar to the target shape. In the third step, the regression factor will be learned and the set of regressed shapes will be obtained. The specific method of determining the weight of the shape will be displayed in [3]

$$\overline{X_{(l)}^{i}} = \sum_{j=1}^{N_l} w^{ij} X_{K_l}^{ij} \tag{6}$$

CFSS learns to estimate probabilistic distribution by the model function followed, and the denominator is a normalizing factor. Thus, the factor $P(s - \bar{X})P(\emptyset(\bar{X})|s - \bar{X})$ is need to be focused and represented by Eq. 8 and Eq. 9.

$$P\left(s - \bar{X}|\emptyset(\bar{X})\right) = \frac{P\left(s - \bar{X}\right)P(\emptyset(\bar{X})|s - \bar{X})}{\sum_{y \in S} P\left(y - \bar{X}\right)P(\emptyset(\bar{X}))|y - \bar{X})} \tag{7}$$

$$P(S - \bar{X}) \propto \exp\left(-\frac{1}{2}(s - \bar{X})^T \Sigma^{-1}(s - \bar{X})\right) \tag{8}$$

$$P\left(\emptyset(\bar{X})|s - \bar{X}\right) = \prod_{j} P\left(\emptyset\left(\bar{X}^{(j)}\right)|s^{(j)} - \bar{X}^{(j)}\right) \tag{9}$$

where j is the facial component inde. The needed shape x and s will be obtained by the Probability formula. This can be identified by each face poses and the feature, which will move to the possible shape regions. After the probability estimation in the shape region, the constrained shape will be selected for the next stage by combining these factors.

A new face alignment method is proposed by searching the shape from coarse to fine. The superior error tolerance is obtained by probabilistic sub-region search. Besides, the dominant set filters the sub-region of incorrect shape. Compared with the traditional methods, CFSS is independent on the initialization and robust in the large-pose situation. The outcome of the experiment shows that this method will improve the accuracy and efficiency.

2.3 Deep Learning-Based Algorithm

Many facial applications, such as facial recognition and facial expression analysis, are inseparable from the facial detection and alignment tasks. The progress of performance

of Neural Networks makes deep learning-based algorithms in face applications possible. This algorithm proposes a network based on deep convolution neural for facial recognition. The effective response of facial region is obtained by training, and the candidate window of face is generated. At present, neural network with cascade method or multi-task method is an efficient method for facial alignment. Faced with the problems of local minimization resulted from blurring and data damage in complex image samples due to occlusion, large-pose change and extreme illumination, deep convolutional network cascade tends to be used. Deep Convolutional Network Cascade [4] proposes a way to estimate the location of key points of face with three convolution network. The first level of convolution is used to make accurate predictions. This method effectively avoids local minimization problem. The input of this convolution network is the whole face, so that the context information can be fully used, which can efficiently predict the key points in the face. The remaining two levels refine the initial prediction of the key points. The two levels of neural networks structure are shallow due to their low tasks and their inputs are in the limitation to a local area around the initial location. In each level, cascaded convolution neural networks are in the connection with improve the precise and robustness of the prediction of the key points. Because predicting key points in the large input face is a high-level task, the convolution neural network should be the deep one, and the deep structure is helpful to extract the non-linear high-level feature, which means that the non-linearity between inputs and outputs can be increased and it is possible to make connection between inputs and outputs. Aimed at leveling the accuracy and reliability of the detection, the method designed a method that used a multi-level regression networks to predict the location of every point in every level. The function of final prediction position of the key points in the face can be expressed as Eq. 10

$$ \mathrm{x} = \frac{x_1^{(1)} + \cdots + x_{l_1}^{(1)}}{l_1} + \sum_{i=2}^{n} \frac{\Delta x_1^{(i)} + \cdots + \Delta x_{l_i}^{(i)}}{l_i} \tag{10} $$

Deep multi-task learning proposes an efficient convolution neural network for facial point detection. For searching the several key features of network structure, this method proposed the high-performance convolution networks on the small scale.

For solving the problems of occlusion and pose change, another method is deep multi-task learning. This method planned to improve detection of facial landmark accuracy with the help of other related facial tasks, instead of regarding these tasks as an independent tasks. This method proposes a task-constrained model, in which tasks are stopped ahead of schedule to promote learning convergence. This method especially has advance in processing images with serious occlusion and large pose change. Compared with cascade deep-learning model, the complexity of the model is greatly reduced. This pose information of the input images can be used as an extra information to constrain the solution space estimation of landmarks, which can optimize the face landmark detection through a set of related tasks. This method proposes a loss function Eq. 11 to improve the generalization ability of landmark detection, errors of related tasks are allowed to be propagated jointly.

$$ \underset{W^r, \{w^a\}_{a \in A}}{\arg\min} \sum_{i=1}^{N} l^r \left(y_i^r, f\left(x_i; W^r \right) \right) + \sum_{i=1}^{N} \sum_{a \in A} \lambda^a l^a \left(y_i^a, f\left(x_i; W^a \right) \right) \tag{11} $$

We can suppose that we have a total of T tasks and the training data for the tth task, denoted as x_i, y_i. x_i is the feature vector and y_i is the label. w^r and w^a is the weight vector. The goal of this method is to minimize the loss function Eq. 11. The method proposed that the learning purpose of network is random gradient descent. Because different tasks have different convergence rates, it is not an easy thing to use the same way to optimize several tasks. Thus, this method proposes an efficient way to stop tasks ahead of time so that they will not exceed the training set and thus damage the main tasks. If not stopped in advance, the training error will converge slowly and show significant oscillation. This method does not learn the face landmarks detection isolate. This model is robust to the serious occlusion and large pose change face. For further use, this method will be useful in deep multi-task learning of landmarks detection and version areas.

3 Experiments

This experimental results are reported in this section which are concerned with synthetic and real data. SDM in four analytic functions and tests the performance of the SDM are compared in the first experiment due to the problem of facial feature detection in two standard databases. In the second experiment, we compare the CFSS and CFSS-Practical with methods, which is in the field of state-of-the-art and compare with cascaded regression on different initializations.

3.1 Experiment on Tradition Algorithms

The LFPW dataset [9] and the LFW-A&C dataset [11] are used in the experiment. The LFPW dataset contains images downloaded from the network, which show great differences in posture, illumination and facial expression. LFW-A&C is a subset of LFW dataset, which is consisted of 1116 images of people whose names begin with an 'A' or 'C' [11]. The model with 66 landmarks are trained in this methods on LFW and LFW-A&C datasets and evaluated tracking performance of SDM [2] on two datasets, RU-FACS [13] and YouTube Celebrities [14]. Besides, this experiment compares SDM method by icon. The performance of SDM and Newton methods in speed and accuracy is compared on four analytic functions (Fig. 1 and Table 1).

3.2 Experiment on Cascade Regression-Based Algorithm

This experiment compares CFSS [3] and CFSS-practicality with other methods, such as state-of-the-art method and only uses data from specific training sets to train our models. The average error comparison Table 2 summarizes the comparison results. It can be found that the two settings of the proposed method are better than former methods, which have been used on these datasets. The results show that the search framework has stronger robustness than the traditional cascade regression method. In Fig. 2, we plotted the CED curves of different methods.

To emphasize the usage of the proposed search method based on coarse-to-fine shape over the traditional cascade regression method, CFSS and CFSS-Practical are compared with SDM. In order to make a fair comparison, this experiment compares the cascade
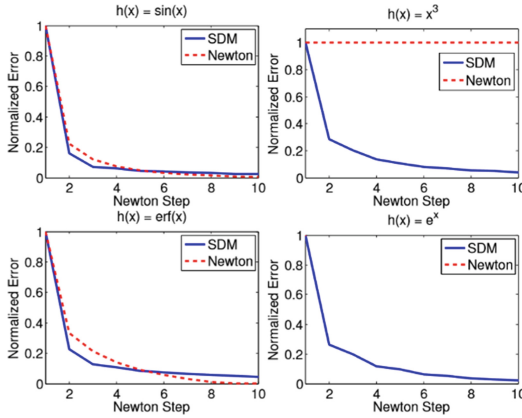
Fig. 1. Normalized error versus iterations with the Newton method and SDM.

Table 1. Experimental setup for the SDM on analytic functions [2].

Function	Training set		Test set
$h(x)$	y	x $= h^{-1}$(y)	y*
$\sin(x)$	$[-1:0.2:1]$	arcsin(y)	$[-1:0.05:1]$
x^3	$[-27:3:27]$	$y^{\frac{1}{3}}$	$[-27:0.5:27]$
$\mathrm{erf}(x)$	$[-0.99:0.11:0.99]$	erf^{-1}(y)	$[-0.99:0.03:0.99]$
e^x	$[1:3:28]$	log(y)	$[1:0.5:28]$

regression method with the application of two feature settings, that is, to select the optimum settings of the method; SIFT [12] are used in the last iteration and BRIEF is used in the other iterations. The results are shown in Table 3. Cascade regression does not perform well on mixed features, suggesting that it may not be beneficial from some different types. The competitive speed performance is achieved by cascade regression. This framework can benefit from integrating more efficient functions in the future.

3.3 Experiment on Deep Learning-Based Algorithm

This experiment evaluates the robustness of learning with related task and compares this method with the cascaded CNN. Besides, this experiment also compares this method with other methods. this experiment employs AFLW [10] for evaluation. This dataset is selected because it is more challenging than other datasets. AFLW has larger pose variations (39% of faces are non-frontal in our testing images) and severe partial occlusions. [4] As can be seen in Fig. 3, it is beneficial to optimize landmark detection by using related tasks. In particular, the performance of FLD + is much better than that of FLD, and the failure rate is reduced by more than 10%. FLD + pose performs best when there is a single related task.

Table 2. Comparison of averaged errors with state-of-the-art methods [3].

LFPW dataset			Helan dataset				300-W dataset (All 68 points)			
Method	68 -pts	49 -pts	Method	194 -pts	68 -pts	49 -pts	Method	Common subset	Challenging subset	Fullset
Zhu et al. [41]	8.29	7.78	Zhu et al. [41]	–	8.16	7.43	Zhu et al. [41]	8.22	18.33	10.20
DRMF [3]	6.57	–	DRMF [3]	–	6.70	–	DRMF [3]	6.65	19.79	9.22
			ESR [10]	5.70	–	–	ESR [10]	5.28	17.00	7.58
RCPR [8]	6.56	5.48	RCPR [8]	6.50	5.93	4.64	RCPR [8]	6.18	17.26	8.35
SDM [37]	5.67	4.47	SDM [37]	5.85	5.50	4.25	SDM [37]	5.57	15.40	7.50
							Smith et al. [32]	–	13.30	–
							Zhao et al. [40]	–	–	6.31
GN-DPM [34]	5.92	4.43	GN-DPM [34]	–	5.69	4.06	GN-DPM [34]	5.78	–	–
CFAN [38]	5.44	–	CFAN [38]	–	5.53	–	CFAN [38]	5.50	–	–
			ERT [20]	4.90	–	–	ERT [20]	–	–	6.40
			LBF [29]	5.41	–	–	LBF [29]	4.95	11.98	6.32
			LBF fast [29]	5.80	–	–	LBF fast [29]	5.38	15.50	7.37
CFSS	**4.87**	**3.78**	**CFSS**	**4.74**	**4.63**	**3.47**	**CFSS**	**4.73**	**9.98**	**5.76**
CFSS practical	**4.90**	**3.80**	**CFSS practical**	**4.84**	**4.72**	**3.50**	**CFSS practical**	**4.79**	**10.92**	**5.99**

The same 10,000 training areas are used in the experiment as the cascade CNN method. Therefore, the difference is that the experiment adopts an approach based on multi-task learning. As can be seen from Fig. 4, the method performs better on several landmarks, and its overall precise is more robust than that of cascaded CNN.

The experiment compares against: (1) Robust Cascaded Pose Regression (RCPR) [15] using the publicly available implementation and parameter settings; (2) Tree Structured Part Model (TSPM) [16], which jointly estimates the head pose and facial landmarks; (3) A commercial software, Luxand face SDK [17]; (4) Explicit Shape Regression (ESR) [18]; (5) A Cascaded Deformable Model (CDM) [19]; (6) Supervised Descent Method (SDM) [2]. For the methods which include their own face detector (TSPM [16] and CDM [19]), detection errors are avoided by cropping the image around the face. This experiment uses two datasets to evaluate: AFLW and AFW. The result is shown in [Facial Landmark Detection by Deep Multi-Task Learning]. Besides, this experiment compares this method with methods based on state-of-the-art and latest commercial software on BioID and LFPW. The results are concluded in Fig. 5.
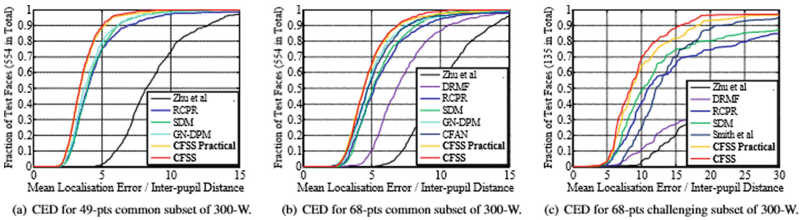

Fig. 2. CED curves of different methods. [3]

Table 3. Comparison of averaged error for cascaded regression and coarse-to-fine searching methods [3].

Feature settings	Cascaded regression				Coarse-to-fine searching		
	Random initialisation	Random voting	Mean shape initialisation	Good initialisation	After stage 1	After stage 2	After stage 3
(1) SIFT throughout	10.68	8.17	7.50	6.33	13.68	8.18	**5.76**
(2) Hybrid features	15.36	11.62	10.20	7.01	18.68	11.43	**5.99**

Fig. 3. Comparison of different model variants: the mean error over different landmarks, and the overall failure rate [5].
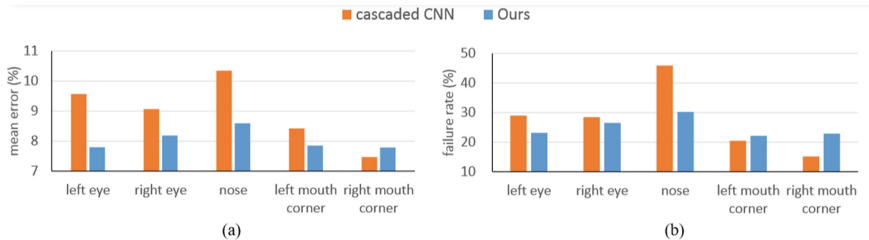
Fig. 4. Comparison with cascaded CNN. [5]

We can learned from the methods discussed above that deep learning methods are much superior to the traditional methods. The reason is that the deep learning-based methods are data-driven, which is trained based on large amount of samples. Then, the main characters of these samples can be extracted by neural networks. In contrast, traditional methods always use hand-crafted descriptors to extract features to do face alignment. The hand-crafted descriptor is based on the experience and mathematical derivation of human beings, which cannot cover amount of kinds of samples. Therefore, deep learning-based methods are better than the traditional methods. However, the training and testing complexity of deep learning-based methods are much higher than traditional methods, which will hinder their practical usage. On the other hand, multi-task based methods are usually more accurate than the single task methods. The reason is that face consists of many correlated attributes, like expression and pose. These attributes are not independent. In contrast, they can supplement more information that may be

Fig. 5. Comparison on BioID and LFPW. Since our failure rate approaches to zero on BioID, it may not be observable in the figure [4].

lacked in one single task. Therefore, exploiting the correlation among these attributes can enhance the performance of these tasks.

4 Conclusion and Future Work

In this paper, we compare different face alignment algorithms in different metrics. In this paper, face alignment algorithms are divided into three categories, namely traditional algorithms, cascade regression-based algorithms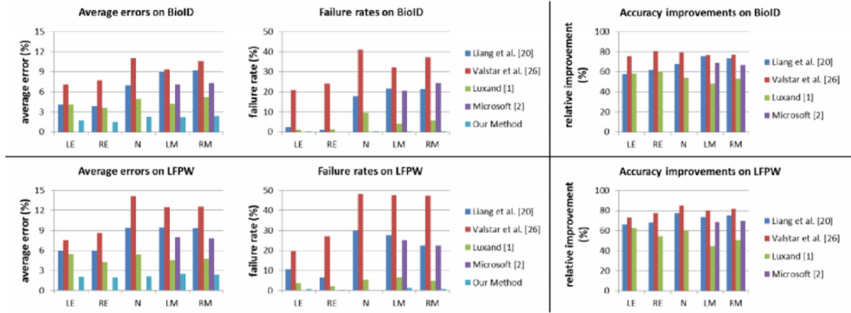 and deep learning-based algorithms. Each of them is analyzed and compared in detail. At the same time, the performance of different algorithms in adverse circumstances is concluded and discussed. In addition, the methods mentioned in this paper are tested in different datasets. The results show that the deep learning algorithms are robust for faces with large pose change or occlusion. In the future, we think cascade regression- and deep learning-based 3D face detection and alignment can further improve accuracy and efficiency.

References

1. Mo, H., Liu, L., Zhu, W., Yin, S., Wei, S.: Face alignment with expression- and pose-based adaptive initialization. IEEE Trans. Multimedia **21**(4), 943–956 (2019)
2. Xiong, X., DelaTorre, F.: Supervised descent method and its applications to face alignment
3. Zhu, S., Li, C., Loy, C.C., Tang, X.: Face alignment by coarse-to-fine shape searching
4. Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection
5. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multi-task learning
6. Trigeorgis, G., Snape, P., Nicolaou, M.A., Antonakos, E.: Mnemonic descent method: a recurrent process applied for end-to-end face alignment
7. Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: 300 faces in-the-wild challenge: the first facial landmark localization challenge. In: ICCVW, pp. 397–403 (2013)
8. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S.: Interactive facial feature localization. In: ECCV, pp. 679–692 (2012)
9. Belhumeur, P.N., Jacobs, D.W., Kriegman, D., Kumar, N.: Localizing parts of faces using a consensus of exemplars. In: CVPR, pp. 545–552 (2011)

10. Kostinger, M., Wohlhart, P., Roth, P.M., Bischof, H.: Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization. In: ICCV Workshops, pp. 2144–2151 (2011)
11. Saragih, J.: Principal regression analysis. In: CVPR (2011)
12. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2), 91–110 (2004)
13. Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Automatic recognition of facial actions in spontaneous expressions. J. Multimedia **1**(6), 22–35 (2006)
14. Kim, M., Kumar, S., Pavlovic, V., Rowley, H.: Face tracking and recognition with visual constraints in real-world videos. In: CVPR (2008)
15. Burgos-Artizzu, X.P., Perona, P., Dollar, P.: Robust face landmark estimation under occlusion. In: ICCV, pp. 1513–1520
16. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: CVPR, pp. 2879–2886 (2012)
17. Luxand Incorporated: Luxand face SDK. http://www.luxand.com/
18. Cao, X., Wei, Y., Wen, F., Sun, J.: Face alignment by explicit shape regression. In: CVPR, pp. 2887–2894 (2012)
19. Yu, X., Huang, J., Zhang, S., Yan, W., Metaxas, D.N.: Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. In: ICCV, pp. 1944–1951 (2013)

Automatic Detection and Counting of Malaria Parasite-Infected Blood Cells

Elena Doering[1,2,3], Anna Pukropski[1(✉)], Ulf Krumnack[1] , and Axel Schaffland[1]

[1] Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
apukropski@uni-osnabrueck.de

[2] Faculty of Medicine, Department of Nuclear Medicine, University of Cologne, Cologne, Germany

[3] German Center for Neurodegenerative Diseases (DZNE), Bonn/Cologne, Germany

Abstract. In this paper, we present a technique for automatic detection and counting of Plasmodium vivax-infected red blood cells by means of a convolutional neural network and a feature- based counting process. Current approaches for object detection or counting often rely on prior knowledge of certain salient features of the to-be-identified objects or require time-consuming pre-processing. For this reason, many detection problems, for example infected cell counting, remain a manual task for trained professionals, leading to potentially high amounts of time between infection and diagnosis and treatment, which, in turn, can have lethal consequences. Using the BBBC041 data set, we annotated the ground truth (GT) of infected cells with circles in each image and then trained a convolutional neural network to predict these GTs from previously unseen cell images. Subsequently, the algorithm computes the number of cells using Canny edge detection and circular Hough Transform.

Keywords: Malaria · Microscopy images · Deep learning · Convolutional neural networks · Cell counting · Computer vision

1 Introduction

1.1 Background

Automated object detection via deep learning is a branch of computer vision that is progressively becoming relevant in medical diagnoses. Many medical image analyses, for example the detection of a malaria parasite infection in red blood cells, however, remain a task to be carried out by human experts until today, as no reliable automated alternative has been established yet.

Despite enormous efforts to prevent or cure malaria, over 200 million new cases of malaria are observed worldwide each year, and around 400,000 people die from the disease per annum. Malaria is caused by parasitic infection of red blood cells, which is initiated through bites of the female Anopheles mosquito [22]. There are several species of malaria parasites, one of them are the so-called Plasmodium parasites. Plasmodium vivax (P. vivax) belongs to this species, and it is most prevalent in South-East Asia, the Middle-East and Africa [23]. Although P. vivax is commonly regarded as one of the less malicious malaria-causing species, it is considered one of the major contributors to malaria morbidity [9]. A parasitic infection with P. vivax leads to symptoms such as fever, rigors, anorexia, nausea, and psychological changes and in some cases even to a rupture of the spleen [5, 9]. Early and reliable diagnoses are therefore of major importance for the well-being of infected patients [16]. Diagnosis and treatment supervision is usually done by a specialist's microscopic investigation of the presence of Plasmodium parasites in thick blood smears, and subsequent type-identifcation and count in thin blood smears through a range of features, such as shape irregularity and size, which is a tedious task and depends heavily on the specialist's degree of expertise. Furthermore, to conserve the current state of health in the blood for diagnosis, blood smears must be investigated immediately after blood drawing, hence requiring for a specialist to be available at all times [6], which is challenging for health-care providers in disease-native regions, especially considering that the dominant areas of P. vivax infection are commonly regarded as ressource-constrained [14]. This poses strong motivation to create automated systems that quickly and reliably perform the task of detecting and counting P. vivax-infected cells.

Recent developments in computer vision and deep learning provide fast and relatively error-free means to detect objects in an image without the requirement for trained professionals. Convolutional Neural Networks (CNNs) have proven to effciently process (medical) image data [21, 24], thus offering the possibility to automatically detect and count infected cells in a blood smear. Automated cell counting models are typically either regression- [19, 24] or detection-based [11, 15]. By downstreaming the two approaches, we offer an effcient and convenient method to both detect infected cells in an image, as well as count them. Such a technique could be made available to doctors in risk areas to curtail the need for trained specialists and reduce the time between blood drawing and diagnosis, therefore lowering the increasing symptom severity over time.

1.2 The Current Study

This paper aims to show that CNNs are able to accurately detect P. vivax-infected cells among healthy blood cells, and to indicate/"predict" their location in an image. Furthermore, we show that our algorithm is able to accurately count the number of infected cells from the CNN's prediction through the application of Canny edge detection and Hough transform.

2 Related Work

Automated cell detection for subsequent counting is a technique not only desirable for malaria research. Being able to assess the number of cells of various cell types in a given

blood sample increases the reliability of diagnosing diseases in early stages and hence the quality of the treatment. State of the art cell counting approaches can be subdivided into three main areas: Manual counting, indirect counting, and automated cell counting.

Currently, selectively counting specific abnormal or uncommon cells still necessitates the knowledge and expertise of trained professionals (manual (cell) counting). Alemu [1] investigated the accuracy of the detection of malaria disease and subsequent classifcation of the malaria species in blood smears by laboratories in Ethiopia [1]. The study revealed an overall low performance for both tasks with a shift to false negativity in detecting P. vivax (37%). As a fundamental factor for this result, the authors claim missing refreshment trainings for professionals. In order to reduce this high risk of human error and shifting experts' work to other, important tasks, an automated cell counting would be of high value.

The research field of automated cell counting has seen two major branches to overcome the counting problem in the past: Detection-based and regression-based counting methods. Typical automatic detection procedures can be found in the works of the teams surrounding Khatri and Ortiz [11, 15]. The authors pre-defined particular properties of cells, so-called structuring elements, which the algorithm then automatically detected in the images. However, these methods require for prior knowledge of the structuring element, e.g. the knowledge that the cell type of interest is of a particular size, and allow for only little variability. Furthermore, adjusting the thresholds in those models to find the optimal setting for every cell type is a subjective task based on human judgement. Adding new information to an existing data set would require the whole model to be re-evaluated by a human expert. Automating the overall process has the potential to increase the eficiency and reduce the pre-processing stage to a minimum.

As opposed to manual - and automated Cell Counting, indirect counting techniques validate the number of cells based on oblique measures. An example of this approach is the spectrophotometry, using cells' light scattering properties to approximate the overall count: the more cells there are, the less light can pass through a so-called cuvette [7]. This method accurately reveals the number of cells in a blood probe, but it has no means to selectively count, for example, only infected cells.

While conventional approaches to object counting rely on prior knowledge of the object, Xie and colleagues [24] proposed a supervised fully convolutional regression network for automated cell counting. One advantage over existing counting approaches is that these networks operate much faster [2]. In their approach, [24] show that it is possible for a convolutional neural network to predict a density map of cells in an image given information on the cells' coordinates. That is, the CNN is provided an image containing multiple cells, creates a prediction as to where cells are located in the image and then receives as feedback ("labels") the precise locations of cells (see Fig. 1(c)). An error is calculated between the CNN's prediction and the label. The CNN then back-propagates the error and adapts its parameters to eventually (closely) match the prediction. In the authors' approach, this process of regression produces a density map of Gaussian representations of cells, from which a high-precision cell count can then be retrieved. Xie and colleagues [24] report that this technique outperforms previous computer vision counting strategies, thus offering promising motivation for clinical application. CNN-based methodologies have further shown to work well for selective

cell detection approaches [17]. In this paper, we developed an algorithm based on the network architectures proposed by Xie and colleagues [24, 25] to predict the presence and number of P. vivax-infected cells.
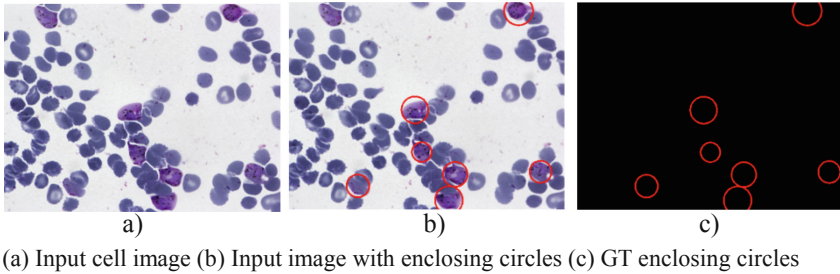

(a) Input cell image (b) Input image with enclosing circles (c) GT enclosing circles

Fig. 1. Ground truth annotation of malaria cells. (b) is only for visualization purposes, for training and validation the GT enclosing circle (c) are used.

3   Method

3.1   Data

We used image set BBBC041v1, available from the Broad Bioimage Benchmark Collection [8, 13]. In this data set, three different sets of blood smears containing P. vivax-infected cells were collected ex vivo in Manaus, Brazil, and Thailand, yielding 1328 images with a total of ~100,000 cells. These blood smears had been stained with a reagent ("Giemsa"), which attaches to the cells' DNA and provides the parasite-infected cell with a distinctive appearance. Locations of infected cells were provided for each image individually by a trained specialist. From the data set, we extracted x and y coordinates, as well as height and width of the cells and annotated the ground truths (GTs) by means of red circles on black background (see Fig. 1 (right))[1]. Four classes of P. vivax-infected cells were present in the current data set: gametocytes, rings, trophozoites, and schizonts. On average, there were ~2 infected cells in an image.

3.2   Network Architecture

CNNs are the most prominently employed type of neural networks when it comes to image analysis. Through so-called convolutional layers, they learn to detect objects in an image through kernels, which independently find particular spatial patterns and pattern combinations. Through its network structure (see Fig. 2), convolutional layers are able to detect increasingly complex pattern combinations that eventually make up the object of interest. In this process, humans do not need to pre-define properties of the object that the CNN is intended to detect, yet its detection accuracy generalizes well to previously unseen images.

---

[1] The annotation of cells in the shape of circles provides an easy way to count overlapping GTs.
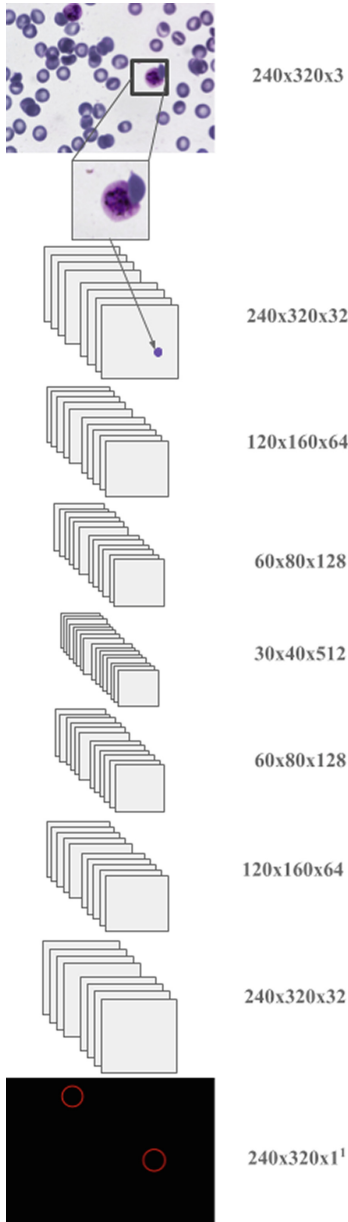
Fig. 2. Architecture of the Convolutional Neural Network of the current study. [1]The CNN predicted Gaussian normalised images with only one channel. For visualization, it is displayed in three channels in the electronic version of this paper.

The CNN of this paper was designed to selectively predict the GT of P. vivax-infected cells, given hitherto unseen images that contained both infected and uninfected blood cells. Based on the previous success in [24] to accurately predict the density surface of cells in a cell image, we implemented a variant of their FCRN-A network architecture [24] (for illustration, see Fig. 2). This network initially performs convolution and max pooling and subsequently deconvolves the images back to their original size through bilinear interpolation. As an activation function, similar to Xie [24], we found the rectified linear units (ReLU) activation function to perform well in all layers. The network had 1,560,768 trainable parameters and five modifiable hyper-parameters: Intensification value, batch size, epochs, learning rate and weight decay. The intensification value refers to the factor by which GT images were multiplied. The idea was taken from [24] and is based on the observation that non-intensified GT led to the network concluding that the smallest error was yielded by completely black images, because the black background made up most of the image space. Therefore, while the network's accuracy was high, visual inspection revealed that prediction performance was poor (see Table 1 and Figs. 3(c), 4). Batch size refers to the amount of data samples that are propagated through the network before weights are updated using the Adam optimizer proposed in [12]. Higher batch sizes are commonly associated with faster performance, while smaller batch sizes were shown to generalize better to unseen images [18]. Finding the balance between a fast learning neural network and a steady and reliable prediction is tedious work which can be controlled, inter alia, by a so-called learning rate. Neural networks are a biologically-inspired machine learning approach and can be applied for various tasks, ranging from visual detection to natural language processing applications. Even though their results are astonishing, training an artificial neural network has to be done carefully by not only observing the resulting statistical metrics as accuracy and loss. They are built to achieve an outcome with a loss as low as possible, which, in some cases, is generalizing to "wrong" or rather unwanted aspects of the data. Therefore, the concept of a learning rate was developed, which continuously controls the network's learning progress

from one batch. In the present CNN, the learning rate controls how much the new observations from one batch influence the internally built concept of an infected cell, which are weighted according to the importance of certain features. The current study uses a fixed learning rate of 0.001. The number of epochs refers to how many times the whole set of data is propagated forward, and backward again through a neural network. If the number of epochs is small, the CNN is given little opportunity to learn. If it is too large, the network's performance will not generalize well to previously unseen images, as it will overfit the training data. In a comparable model architecture aiming at identifying malaria-infected or -uninfected cells, [17] found that five iterations overfit the model for this training data. Finally, the weight decay is a factor smaller than 1, by which weights are multiplied in each iteration. This process prevents weights from growing too large, where smaller weight decays penalize weights stronger. In our case, we selected a weight decay of 0.01, based on the observation made in [24] that the network is likely to learn to achieve higher accuracies by predicting all-black images, i.e., the network will consider the GTs of cells as noise during its internal optimization process.

Importantly, we decided not to cut the images into smaller parts, as seen in [8, 24], because the average number of infected cells per image was already low in the original size (μcell count = 2). Nonetheless, to lower the computational effort and improve the network's performance, the images were scaled down to 320×240 pixels. The network consists of four downsampling layers, three upsampling layers and a final convolutional layer. Downsampling layers comprise convolutional operations, batch normalization and apply the aforementioned ReLU activation function to the data. After each of the downsampling layers, max pooling is applied to discard irrelevant information. To achieve the desired output shape of the images, the so far predicted images need to be deconvolved. This can be accomplished by upsampling. In addition to the upsampling process, two consecutive sets of convolution, batch normalization and ReLU activation were applied. The last convolution layer to generate the network's prediction uses a linear activation function with a single 1×1 kernel operation as opposed to the preceding layers using a ReLU activation.

In the overall network, a kernel size of 3×3 was sufficient. We started with 32 kernels in the first downsampling layer, increasing the number to 512 and decreasing them in the exact same manner in the upsampling layers. In order to prevent our network from overfitting, we adapted the kernel regularization of [24] in all convolutional layers with an L2-Regularization method. This penalizes parameters during the learning phase, hence forcing the network to reduce the possibility of generalization errors. To achieve the best possible prediction, we set the L2-Regularization parameter to 0.01. The loss calculation is done with a root mean squared error as opposed to the mean squared error used in [24], to avoid the likelihood of punishing the model's performance based on single prediction errors. An overall train-test split of 90–10 was incorporated, resulting in 1195 training- and 133 validation samples.

3.3   Counting

Counting of infected cells based on the CNN's output is accomplished by the usage of Canny edge detection and subsequent circular Hough transform provided by the OpenCV library [4]. Previous studies have shown the effectiveness of detecting cells with Hough

(Circle) Transform (HCT) [3, 20] and the circular nature of GTs, as opposed to, for example, rectangles, enables the precise counting of overlapping cells or GT annotations. To account for a reliable measure, the predicted images were pre-processed with a dilation operation. Morphological dilation is one of the most prominent techniques used in object shape (pre-)processing. A 3×3 elliptical structuring element was used as the basis for the dilation to process each pixel in the images. Dilation can be used to fill small holes and strengthening the contour of objects, thus accounting for a more reliable circle detection of the Hough Transform. Applying the same procedure to the validation data, detected cells were counted. A measure of network performance quality was established through the average difference of cells as provided by the prediction of infected cells and its actual GT, the latter of which was not provided to the CNN beforehand.

4   Results

In this section, we evaluate the performance of our network to accurately detect and count P. vivax-infected cells in blood smears provided in the BBBC041 P. vivax data set. We trained the network for 3 epochs and observed a mean difference of 0.78 cells per image between validation labels and the network's predictions.

Table 1. Comparison of different parameter configurations

IV	BS	Epochs	WD	Acc (%)	Diff	Visual
1	8	3	0.01	89.5	8.97	−
1	16	3	0.01	88.8	1.76	−
100	8	3	0.01	90.1	0.89	+
100	16	3	0.01	88.9	1.76	o
100	8	5	0.01	92.3	0.77	+
100	16	5	.01	91.8	2.51	−
100	**8**	**3**	**0.001**	**90.5**	**0.78**	**+**
100	16	3	0.001	90.2	0.83	−

Notes. intensification value (IV), batch size (BS), Epochs, weight decay (WD) in terms of accuracy on the entire validation set (Acc), the overall difference of cells in GT and predicted images (Diff) and subjective visual appearance of a subset of images (Visual). Visual appearance: + good, o medium, − poor. The parameters in the row in bold is used for presentation of results.

Parameter configurations were tested in the combinations presented in Table 1. The maximum accuracy of a neural network is usually expected to be reached with the best

parameter configuration. To ensure high quality standards, we also applied further measures of quality, namely the difference of cells as counted by the same Hough transform algorithm in GT and prediction images, as well as visual investigation of certain samples. This was necessary, because the CNN's accuracy was high for all configurations of parameters, leaving little room for claiming that one is more effcient than another. Prior to assessing the deviation of cell numbers between GT and prediction, the quality of our counting algorithm was ensured by comparing the number of cells indicated in the dataset to the number Hough transform predicted. Subsequently, predicted and GT cell counts were compared in order to evaluate CNN prediction performance. For example, the first and second set (row) of hyper-parameters depicted in Table 1 resulted in severe deviations of cell numbers between GT and predictions, despite showing an accuracy of almost 90%. Without intensity enhancement, a batch size of 8 resulted in false predictions of infected cells, i.e., false positives (see Fig. 3(c, a)), with a batch size of 16, the CNN failed to detect several infected cells (see Fig. 4b)). As postulated in [24], we concluded that the divergence of quality indications resulted from the fact that the area of the background made up large amounts of the image space and thus significantly affected accuracy. By increasing the intensity of the enclosing circle in GT images by a factor of 100 [24], the network's predictions became genuinely better (see Fig. 4b)). The comparison also yielded that while the CNN performed faster with a batch size of 16, its prediction quality was higher with a smaller batch size, which is in accordance with previous findings of neural networks [18]. As accuracy and difference in cell numbers were quite similar among the top configuration candidates, we eventually chose the optimal parameter settings based on visual inspection and found that the configuration as printed in bold in the table below yielded highly accurate results on all three quality scales.

5   Discussion

We have shown that it is possible to accurately detect P. vivax-infected cells among healthy blood cells from thick and thin blood smears using a convolutional regression network. Additionally, we provided a solid base for a computational method that automatically counts the number of infected cells based on Canny edge detection and Hough transform on a cell image predicted by the CNN.

The technique presented in this paper is, to our knowledge, the first automated method to detect P. vivax-infected cells without rigorous and time-intensive image-specific pre-processing algorithms such as segmentation prior to feeding data to the classification algorithms as was done for example in [10, 11, 15]. This further enables the current approach to generalize well on various appearances of cell types: Our model operated well even with blood smears collected by different researchers at various locations around the world (for an example of image quality variability, see Fig. 5). Post-hoc investigations revealed that our network also accurately predicts the presence of infected cells given rectangular GTs[2], the network thus likely operates independent of the shape of GT, provided that the shape of the infected cell's annotation makes up a significant part

---

[2] The hyper-parameters had different optimal configurations for the different GTs.
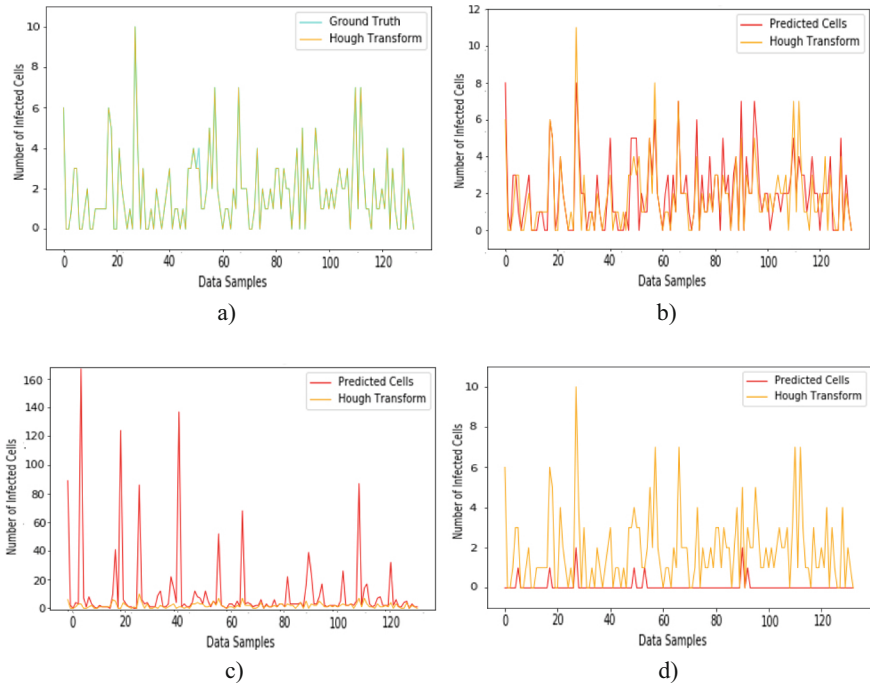
Fig. 3. Cell counting performance using HCT: (a) Nearly perfect counting performance on GT images: green line (electronic version)/dark line (print); (b) High counting performance on prediction images; (c) False positives due to low intensity; (d) False negatives due to low intensity.

of the GT image. Only labelling the centers of cells, for example, has provided poor results with the current approach. Given the large-scale independence of single features in the input and label data, we expect that the network brought forward in this paper can in the future also be used to automatically detect P. vivax infections in blood samples outside the data set used for this project. Trained on other sets of data with different Plasmodium species, it is also possible that our network could identify the presence of more malicious malaria-causing parasites. However, there is currently a lack of data sets to test this hypothesis.

[1] found that the manual counting approach currently suffers from a moderately high false negative rate. Consistently, inspection of some of the network's false positives, cells that were classified as infected by the CNN but not the human expert, commonly showed high visual resemblance to other cells that had been labeled as infected. For an example, see Fig. 6, where yellow/bright circles show the cells that were labeled as being infected by P. vivax by a human specialist, while the red/darker circles indicate which cells our model predicted to be infected. This finding stresses the importance for those who provide diagnoses to dispose of high amounts of expertise, which, especially in risk areas, is not always granted [1]. With an approach as the one brought forward in this paper, we provide a basis for the goal of reducing human error in malaria diagnoses.
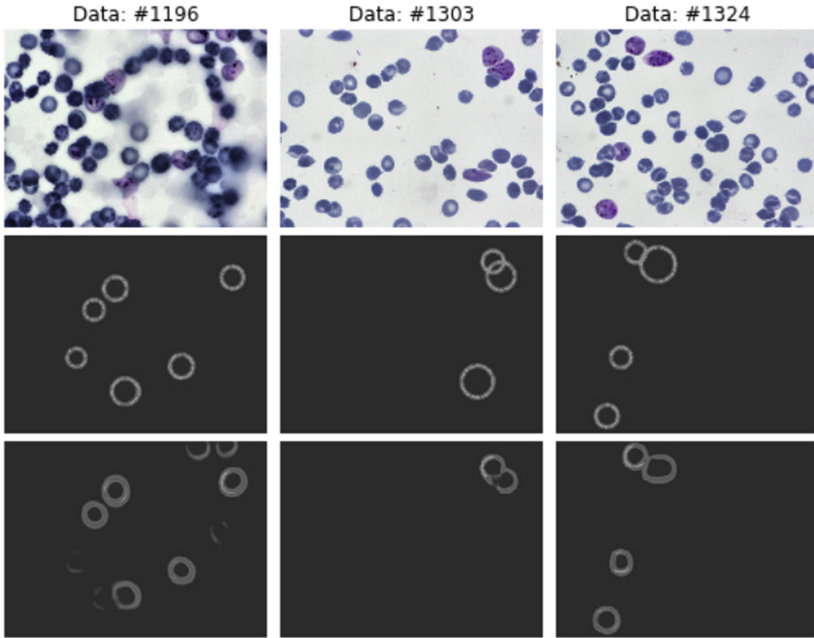
Fig. 4. Examples of three cell images (top) fed into the network for infected-cell detection. Middle: Human-labeled infected cell ocations. Bottom: Network predictions of infected cell locations.
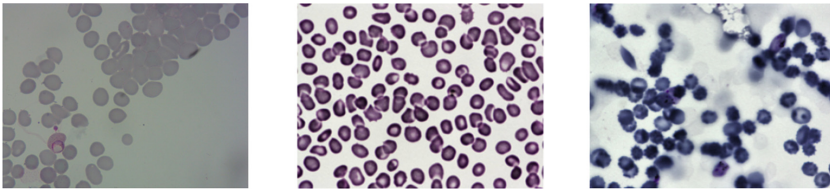
Fig. 5. Example of variability of cell image appearances.

A common misconception about deep learning algorithms that learn based on human-labeled data is that at best they could achieve human-level expertise. For future work, an investigation by different human specialists of the false positives as identified by our network would be appropriate in order to judge which of the false positives are human vs. network errors. If this information was fed into the network anew, it is likely that the network could outperform a single human specialist. For some research purposes, it is of interest to investigate the presence and number of an even smaller partition of the cells in a blood smear, for example of only trophozoites, as done for example in [10]. For such sakes, future work could add our model to build a hybrid structure similar to that in [21], where the authors trained two convolutional neural networks: one for separating cells from non-cells in histopathological images, and the second one for further classification of the cell's type. The BBBC041 data sheet offers information on the infection state of a

cell (infected vs. non-infected), as well as the type of cell (gametocyte, trophozoite, ring cell, schizont). Thus a similar approach would be possible by adding a second neural network for the classification of the specific cell type of infected cells on top of the technique brought forward in this paper. Furthermore, testing the possibility to reduce the aforementioned architecture to a single CNN could be of interest.

A drawback of our approach is that the implemented counting method is not incorporating more automatized operations. To optimize and evaluate the general effectiveness of the counting approach, we compared the found number of circles (count) on the validation images with the GT number of cells in the given image. Eventually, the fine-tuning of the parameters was based on visual inspection of the performance on the predicted images of the network. Overall, the counting approach successfully counted the cells of the validation images with only +2 cells in total. This approach is promising in the case of well shaped circles. However, the proposed CNN does not always predict the ideal cell or circular shape with good intensity values, com-

Fig. 6. Comparison of human-(yellow (electronic version)/bright (print)) and CNN-labeled (red (electronic version)/darker (print)) P. vivax infected cells.

plicating the detection of cells for the HCT. Future research could investigate the usage of a second neural network or other machine learning approach to ensure a more reliable counting method.

Although our attempt to classify P. vivax-infected cells provides a good basis for future malaria diagnosis procedures, it is not ready for immediate use for medical purposes. The mean difference of cells (0.78) between label and predicted image is still high, especially given that the average number of infected cells per blood smear image was only ~2. We acknowledge this weakness of our model, however we believe that future research could build on this study and thus enable fast and effcient malaria diagnoses all around the world.

6   Conclusion

In this paper, we presented a hybrid approach to successfully detect and count P. vivax-infected cells in microscopic images of blood smears. Combining a variant of the first fully convolutional regression network proposed in [24] with a counting technique based on circular Hough transform, we were able to provide an automatic counting approach that, if developed further, could help eliminate errors in the current diagnostic procedure of malaria.

Abbreviations

CNN   Convolutional Neural Network
GT   Ground Truth
HCT   Hough Circular Transform
P. vivax   Plasmodium vivax
ReLU   Rectified Linear Units

References

1. Alemu, M., Tadesse, D., Hailu, T., Mulu, W., Derbie, A., Hailu, T., Abera, B.: Performance of Laboratory Professionals Working on Malaria Microscopy in Tigray, North Ethiopia. J. Parasitol. Res. (2017)
2. Arteta, C., Lempitsky, V., Noble, J.A., Zisserman, A.: Interactive object counting. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2014)
3. Bewes, J., Suchowerska, N., Mckenzie, D.: Automated cell colony counting and analysis using the circular hough image transform algorithm (chita). Phys. Med. Biol. **53**, 5991–6008 (2008)
4. Bradski, G.: The OpenCV Library. Dr. Dobb's J. Softw. Tools (2000)
5. Centers for Disease Control and Prevention: Malaria Disease (2019). https://www.cdc.gov/malaria/about/disease.html
6. Centers for disease control and prevention: treatment of malaria: guidelines for clinicians (united states) (2019). https://www.cdc.gov/malaria/diagnosis_treatment/clinicians1.html
7. Coulter, W.H.: Means for counting particles suspended in a uid. US2656508A (1949)
8. Hung, J., Goodman, A., Lopes, S., Rangel, G., Ravel, D., Costa, F., Marti, M., Carpenter, A.E.: Applying faster R-CNN for Object detection on malaria images. In: CVPR (2017)
9. Imtiaz, S., Drohlia, M., Nasir, K., Hussain, M., Ahmad, A.: Morbidity and mortality associated with Plasmodium vivax and Plasmodium falciparum infection in a tertiary care kidney hospital. Saudi J. Kidney Dis. Transpl. **26**(6), 1169–1176 (2015)
10. Khan, M.I., Acharya, B., Singh, B.K., Soni, J.: Content based image retrieval approaches for detection of malarial parasite in blood images. Int. J. Biometrics Bioinform. **5**(2), 97–110 (2011)
11. Khatri, K.M., Ratnaparkhe, V.R., Ph, D., Agrawal, S.S., Bhalchandra, A.S.: Image processing approach for malarial parasite identification. Int. J. Comput. Appl. (IJCA), 5–7 (2013)
12. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization. arXiv preprint (2015)
13. Ljosa, V., Sokolnicki, K.L., Carpenter, A.E.: Annotated high-throughput microscopy image sets for validation. Nat. Methods **9**(7), 637 (2012)
14. Mitiku, K., Getahun, M., Gelaw, B.: The reliability of blood film examination for malaria at the peripheral health unit. Ethiopian J. Health Dev. **17**(3), 197–204 (2003)
15. De Solórzano, C.O., Rodriguez, E.G., Jones, A., Pinkel, D., Gray, J., Sudar, D., Lockett, S.: Segmentation of confocal microscope images of cell nuclei in thick tissue sections. J. Microsc. **193**(3), 212–226 (1999)
16. Price, R., Tjitra, E., Guerra, C., Yeung, S., White, N., Anstey, N.: Vivax malaria: neglected and not benign. Am. J. Tropical Med. Hygiene **77**(6 Supplementary), 79–97 (2007)
17. Sarkar, D.: Detecting malaria with deep learning (2019). https://opensource.com/article/19/4/detecting-malaria-deep-learning

18. Shen, K.: Effect of batch size on training dynamics. Blogpost (2018). https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e
19. Szegedy, C.: Deep neural networks for object detection. In: Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS), vol. 2, pp. 2553–2561 (2013)
20. Venkatalakshmi, B., Thilagavathi, K.: Automatic red blood cell counting using hough transform. In: 2013 IEEE Conference on Information Communication Technologies, pp. 267–271, April 2013
21. Wang, Sheng, Yao, Jiawen, Xu, Zheng, Huang, Junzhou: Subtype cell detection with an accelerated deep convolution neural network. In: Ourselin, Sebastien, Joskowicz, Leo, Sabuncu, Mert R., Unal, Gozde, Wells, William (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 640–648. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_74
22. World Health Organization: Malaria (2019). https://www.who.int/news-room/facts-in-pictures/detail/malaria
23. World Health Organization: New opportunities to prevent P. vivax malaria relapse (2019). https://www.who.int/malaria/news/2019/new-opportunities-to-prevent-vivax-malaria-relapse/en/
24. Xie, W., Noble, J.A., Zisserman, A.: Microscopy cell counting and detection with fully convolutional regression networks. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. **6**(3), 283–292 (2018)
25. Xie, Y., Xing, F., Kong, X., Su, H., Yang, L.: Beyond classification: structured regression for robust cell detection using convolutional neural network. In: Medical Image Computing and Computer-Assisted Intervention. vol. 9351, pp. 358–365, October 2015

Classification of Chest Diseases Using Wavelet Transforms and Transfer Learning

Ahmed Rasheed[1(✉)], Muhammad Shahzad Younis[1], Muhammad Bilal[2], and Maha Rasheed[1]

[1] School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad, Pakistan
`arasheed.msee17seecs@seecs.edu.pk`
[2] University of the West of England, Bristol, UK

Abstract. Chest X-ray scan is a most often used modality by radiologists to diagnose many chest related diseases in their initial stages. The proposed system aids the radiologists in making decision about the diseases found in the scans more efficiently. Our system combines the techniques of image processing for feature enhancement and deep learning for classification among diseases. We have used the ChestX-ray14 database in order to train our deep learning model on the 14 different labeled diseases found in it. The proposed research shows the significant improvement in the results by using wavelet transforms as pre-processing technique.

Keywords: Transfer learning · Wavelet transform · DNN · Chest X-Ray

1 Introduction

Chest related diseases are considered to be in the leading causes of deaths. To diagnose these diseases, X-Ray is the most regularly used modality. Automatically classifying the diseases within X-Ray image scans of chest remains a tough task. X-Ray scan being a very cost effective exam for body inspection is performed frequently in medical checkups [1]. However, it has been quite a challenging task to clinically diagnose the chest diseases from the scans. Sometimes, this is believed to be more tough than the diagnosis using chest CT scan imaging. Some highly encouraging work has surfaced in the past, and also in recent researches. But the achievement of computer-aided detection and diagnosis (CAD) that is clinically relevant in practical medical stations on chest X-ray with all data settings is still very unfeasible, if not impossible with a dataset that has a handful of images for study and training [18]. Early detection of infected areas of chest and diseases can save lives of many. Computer aided systems are now of great interest to reduce the number of casualties by detecting it in early stages and helping out the radiologists in making important decisions.

The proposed system presents an effective network which utilizes the knowledge and techniques of image processing in the form of wavelet transforms of images and deep learning resulting in classification of infected regions found in the X-Ray scans of

chests. The diseases are classified in the different classes based on the structure, shape and composition of infection present.

Training Convolutional Neural Network (CNN) model with deep connections from the scratch is not easy as a very large dataset and compute power is needed. One way to do the job is to use already trained deep models and tune them according to your needs. Transfer learning also does a similar job by removing and then adding small number of layers modified for the problem in study at the bottom of the network and then training only the newly added layers. This method is effective when you have smaller datasets. The dataset that this paper utilizes is the Chest X-Ray 14 database [1] that contains a total of 112,120 images from 30,805 patients. The dataset provides a considerably large as well as diverse repository for both training as well as testing. The approach to this paper is a comparison between the results retrieved via Transfer Learning in two cases; with wavelet transform and without it.

2 Background and Previous Work

Deep learning models can attain human like thinking and intelligence because of their complex structure which resembles the human neural system. There has been a significant amount of work done by scientists by using deep learning along with pre-processing and augmentation techniques for different problems. Some work has also been done in medical field to achieve state of art results for different problems having different datasets. While working on dataset comprising of images, the most common deep learning architecture used is CNN. Classification is widely done by using CNN architecture for image based data.

2.1 Wavelet Transform

Wavelet transform is a signal processing technique which can be used on images to extricate the important features within the image before feeding it to the classifying network. Through wavelet transform we get simultaneous localization in frequency and time domain [2]. Wavelet transforms have been used in multiple researches to help the doctors or radiologists in order to find the infected region from the X-Ray scans of patients [3–5]. It enhances the features in the imaging which then helps the neural networks to learn those features efficiently while training. It helps the network to learn the important features like the sharp edges, bright regions or other abnormalities found in the X-Ray scan [6].

The general equation representing the working of wavelet transform is provided below. Here in Eq. 1 ψ represents the transforming function, $f(x)$ original signal, k translation parameter and j is scale parameter [7] (Fig. 1).

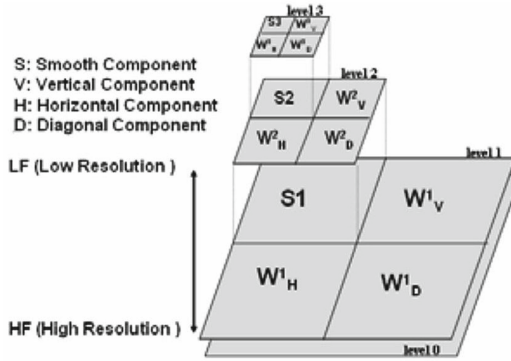$$W_k^j = \int f(x)\psi\left(\frac{x}{2^j} - k\right)dx \qquad (1)$$

Fig. 1. 2D Wavelet transform layer representation up to 3 levels [7].

2.2 Augmentation

Augmentation is the process of generating new samples of data by changing its orientation. It improves the performance of the network and prevents it to over-fit for a specific problem [8]. Augmentation is mostly used when the total quantity of images in a dataset is not enough to train the network effectively [9]. In classification problems it also serves a great role in generalizing the network to perform well for all the classes and similar datasets [10].

2.3 Deep Neural Network (DNN)

DNN are widely used for the classification of medical images [11]. A lot of work has been done in medical field by using DNNs, detection of chest diseases has also been done before through DNN in order to obtain better results compared to conventional classification techniques [12].

Convolutional Neural Network. CNNs are composed of layers with number of convolutional filters which are learned by the network in such a way that they classify the image at the end of the network. CNN are used for processing the image based data as it learns the filters to distinguish between classes [11]. The end layer of the CNN works on the basis of probability as it outputs the probability of each class and then selects the class having the highest probability for an image.

Transfer Learning. Also known as "Transfer-ability" is the use of network trained on a different dataset and problem for your own dataset and problem. In transfer learning, a pre-trained network is obtained and its last few layers are replaced in order to amend its architecture for the classification of new classes, then the filters of first few layers of network are kept frozen, remaining layers in the network are further trained on the newer dataset. Due to the limited amount of data available, associated with medical imaging, transfer learning comes in handy instead of training from scratch [13, 14]. Few known pre-trained networks are ResNet, VGG, GoogleNet and AlexNet. All these networks have been trained on the ImageNet database [15] designed to classify images with 1000

classes. To use them for the proposed problem first few starting layers are kept frozen and newly added last layers are trained again in order to work for the desired classes. Learning rate is set high for the new layers added. The network above remains unchanged and the new layers are further trained for the newer dataset.

3   Dataset

Neural networks generally require a large number sample images for the training purpose. The dataset used for this research is ChestX-ray14 database. This dataset is obtained from the online database formulated by the U.S. National Institute of Health Clinical Center, through their clinical PACS. This database has 60% scans from all frontal chest X-ray scans done within the hospitals [1].

Chest X Ray 14 database consists of total 112,120 images of frontal X-Ray scans collected from 30,805 patients having fourteen different diseases. Dataset is obtained through natural language processing of their associated radio-logical reports. There are fourteen common thoracic pathologies found in the chest X-ray scans. Table 1 shows the names and distribution of all these fourteen classes based on the total images existing in each class.

Table 1. Class distribution of database

Classes	Number of images
Infiltration	25366
Effusion	18974
Atelectasis	16057
Nodule	8409
Mass	8269
Consolidation	7177
Pneumothorax	7134
Pleural Thickening	5172
Cardiomegaly	3906
Emphysema	3586
Edema	3443
Fibrosis	2211
Pneumonia	2092
Hernia	284
Total	1121120

For training and testing purposes the dataset is further divided in to three sub parts. Training, validation and test data at 75%, 15% and 15% respectively. This splitting of dataset is totally random to ensure unbiased network training.

4   Methodology

The flow diagram in the Fig. 2 gives the overall idea of the methodology used.
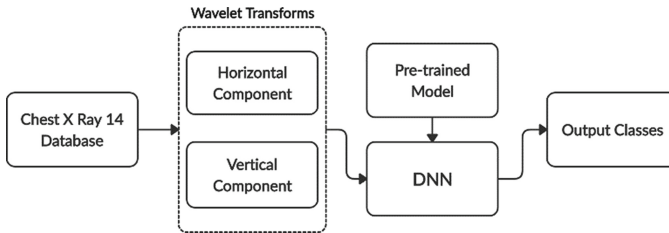
Fig. 2. Proposed system workflow diagram.

4.1   Wavelet Transforms

Wavelet transform of the X-Ray scans of complete dataset are done as image pre-processing in the proposed system. They help the network by enhancing the features in the images. Two output images are obtained from this process, one in vertical and other in horizontal directions of image. Sample images are shown in Fig. 3. Images are resized and then fed to the DNN for classification.
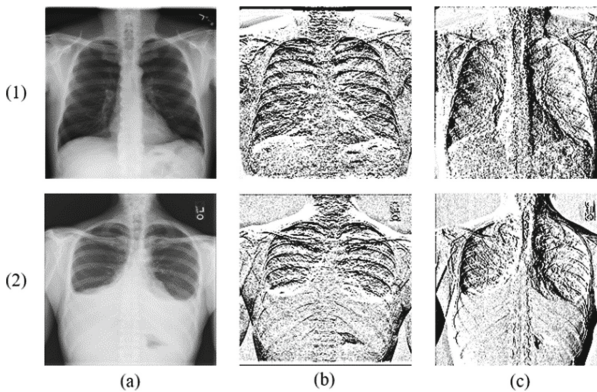
Fig. 3. Samples obtained from wavelet transform for two diseases (1) Infiltration, (2) Effusion while (a) is original and (b)(c) are vertical and horizontal component of image respectively.

4.2   Deep Neural Network (DNN)

After the pre-processing of images which is wavelet transform and resizing, the images are then fed to the DNN for classification purpose. Rather than training a whole network from scratch we have used transfer learning techniques on a pre-trained network.

Pre-trained CNN. An already trained networks is more useful instead of building a whole new CNN architecture from scratch and for this purpose we have used ResNet50 architecture which was introduced in 2015 [16]. It is one of the state of art network trained for the ImageNet database for classification task [15].

Augmentation. In order to generalize the train network and avoid the over-fitting problem, we have done augmentation by doing rotation, translation and scaling with random parameters.

Transfer Learning. In our proposed system we have transferred the knowledge of already trained network on ImageNet database to our domain which is classification of diseases in the ChestX-ray14 database by freezing the first 20 layers of pre-trained network (ResNet50) and replacing its final layers to classify our desired classes. Transfer learning comes in handy when the data on which the system is to be trained is a small database [17].

Training and Validation. The dataset is split onto three parts, training, validation and test dataset. The training dataset has been used to train the network by updating its weights for each layer added while the validation dataset serves as a feedback to the network after every few iterations and helps the system to improve and avoid over-fitting. Table 2 lists the hyper-parameters used for the training of networks.

Table 2. Hyper-parameters of Network

Parameters	Values
Total epochs	15
Batch size	20
Initial learning rate	3e−4
Activation function	ReLu
Optimizer	SGDM

5   Results and Discussion

We performed two kind of tests. First DNN was trained using transfer learning on the original dataset with 14 classes. Then the training of network was performed on the images obtained after using wavelet transform as pre-processing technique. The evaluation metric for the results is Receiver Operating Characteristic (ROC) curve. Curves for both the tests are shown in the Fig. 4 for the three major classes from the dataset.

The solid lines represent the ROC for the network trained on wavelet transform of images while dotted lines represents the network trained on original images. We can see from Fig. 4 that taking wavelet transform of X-Ray scan significantly improves the performance of the classification network. As for future work, more complex DNN architectures can give much better results when trained using wavelet transforms on high end compute resources.
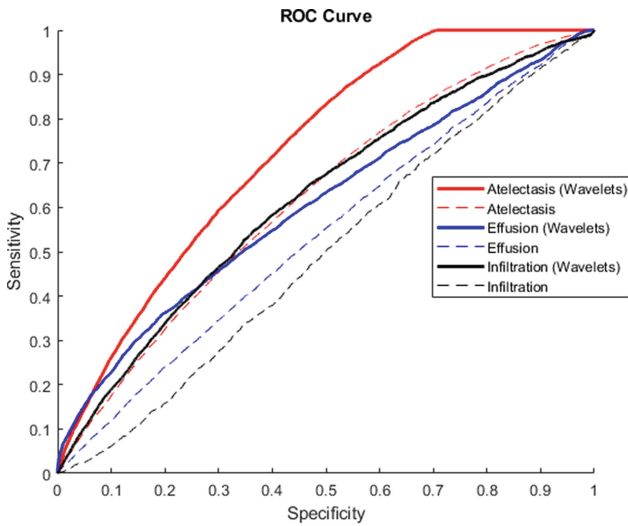
Fig. 4. ROC for three major classes Atelectasis, Effusion and Infiltration.

References

1. Wang, X., et al.: Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
2. Singh, G., Verma, D., Koranga, P.: Tech Scholar: Application of Wavelet Transform on Images: A Review (2016)
3. Li, L., et al.: Wavelet transform for directional feature extraction in medical imaging. In: Proceedings of International Conference on Image Processing, vol. 3. IEEE (1997)
4. Wang, A., Sun, H., Guan, Y.: The application of wavelet transform to multi-modality medical image fusion. In: 2006 IEEE International Conference on Networking, Sensing and Control. IEEE (2006)
5. Sifuzzaman, M., Islam, M.R., Ali, M.Z.: Application of wavelet transform and its advantages compared to Fourier transform (2009)
6. Chaplot, S., Patnaik, L.M., Jagannathan, N.R.: Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomed. Sign. Process. Control **1**(1), 86–92 (2006)
7. Mirzaalian, H., et al.: Various applying of wavelet transform in digital mammograms for detecting masses and microcalsifications. In: MVA (2007)
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Computer Vision and Pattern Recognition (CVPR) (2015)
9. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621 (2017)
10. Roth, H.R., et al.: Anatomy-specific classification of medical images using deep convolutional nets. arXiv preprint arXiv:1504.04003 (2015)
11. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)

12. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology **284**(2), 574–582 (2017)
13. Yu, Y., et al.: Deep transfer learning for modality classification of medical images. Information **8**(3), 91 (2017)
14. Kermany, D.S., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**(5), 1122–1131 (2018)
15. Deng, J., et al.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2009)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, CoRR, vol. abs/1512.03385 (2015)
17. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2009)
18. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput. Med. Imaging Graph. **31**(4-5), 198–211 (2007)

Performance Analysis of Different 2D and 3D CNN Model for Liver Semantic Segmentation: A Review

Ashfia Binte Habib[1(✉)] (iD), Mahmud Elahi Akhter[1], Rafeed Sultaan[2],
Zunayeed Bin Zahir[1], Rishad Arfin[1], Fahimul Haque[1], Syed Athar Bin Amir[1],
Md Shahriar Hussain[1], and Rajesh Palit[1]

[1] North South University, Dhaka, Bangladesh
ashfia.habib@northsouth.edu
[2] RMIT University, Melbourne, Australia

Abstract. Image segmentation is one of the most popular methods in automated computational medical image analysis. Precise and significant semantic segmentation on abdominal Magnetic Resonance Imaging (MRI) and Computer Tomography (CT) volume images, specifically liver segmentation has a lot of contribution towards clinical decision making for patient treatment. Apart from the many state-of-the-art methods, different cutting-edge deep learning architectures are being developed rapidly. Those architectures are performing better segmentation while at the same time outperforming other state-of- the-art models. Different deep learning models perform differently based on cell types, organ shapes and the type of medical imaging (i.e. CT, MRI). Starting from 2D convolutional networks (CNN), many variations of 3D convolutional neural network architectures have achieved significant results in segmentation tasks on MRI and CT images. In this paper, we review performance of different 2D and 3D CNN models for liver image segmentation. We also analyzed studies that used variants of ResNet, FCN, U-Net, and 3D U-Net along with various evaluation metrics. How these variants of 2D and 3D CNN models enhance the performance against its state-of-the-art architectures are demonstrated in the results section. Besides the architectural development, each year, new segmentation and other biomedical challenges are being offered. These challenges come with their own datasets. Apart from challenges, some datasets are provided and supported by different organizations. Use of such data set can be found in this study. Moreover, this review of reported results, along with different datasets and architectures will help future researchers in liver semantic segmentation tasks. Furthermore, our listing of results will give the insight to analyze the use of different metrics for the same organs with the change in performances.

Keywords: Liver semantic segmentation · Deep learning architecture · 2D and 3D convolutional neural network · Evaluation metrics

1 Introduction

In recent years, the field of biomedical image segmentation has seen an influx of deep learning-based segmentation architectures. Although most architectures are 2D, in recent

years the use of 3D CNN has seen increase. This is to thank U-Net [1] for this massive influx as many of these architectures are U-Net variants and or are inspired by U-Net. The increase of segmentation challenges has also helped broaden the horizon of different segmentation architectures.

Currently, many architectures have achieved significant results in different biomedical image segmentation tasks. However, in a recent paper, it was also shown that many architectures are challenge oriented and there are many ways challenges can be exploited [2]. To this end, we have reviewed performance of different deep learning architectures to see how state-of-the-art architectures and other architectures have fared on the task of liver segmentation. After introduction, this study delves into a few state-of-the-art CNN architectures in Sect. 2 which have been rather influential in the past few years. Then, in Sect. 3 we list the publications that were used in this study. For the same organ (i.e. liver) different dataset and different data type along with the data description have also been listed in tables of this section. The overview of this section will help researchers to review the performance of the different neural network architectures and its variants on liver image segmentation task.

For the purpose of having better results, many different metrics have been used to check and justify the segmentation methodologies. Many researchers claim diverse metrics to report their performances. As evaluation criteria of the different deep neural networks of liver segmentation, Dice Coefficient (Dice), Intersection over Union (IoU), Volumetric overlap error (VOE), Maximum Symmetric Surface Distance (MSSD), Relative Volume Difference (RVD), Average Symmetric Surface Difference (ASSD), Root Mean Square Symmetric Surface Distance (RMSD) have been used in many papers. Among all these metrics, the dice coefficient percentage has been used mostly. Also, many research papers use dice global, dice per case metrics.

In this paper, we reviewed the performance of different 2D and 3D CNN models on the task of liver segmentation. Our work will facilitate future researchers of liver image segmentation tasks with model selection and creation. Moreover, the variation of performance metrics can also help with choosing specific metrics for liver segmentation to achieve better performance. This review will help the future researchers search for suitable performance metrics for liver segmentation task.

2 Related Work

2.1 ResNet

ResNet [3] is a neural network that first employed skip connections from an input to a subsequent layer to minimize the loss of data which is a byproduct of convolution. The input is normalized to fit the dimension of the convolution layer of lower dimensions. Before ResNet, the main problem was that, with the addition of more layer's, error from the training would also increase. Which did not minimize the loss of the data. Afterwards, with ResNet's skip connections, the increase of layers does not result in significant loss of data as it feeds the input of the previous layer to the next layer (Fig.1).
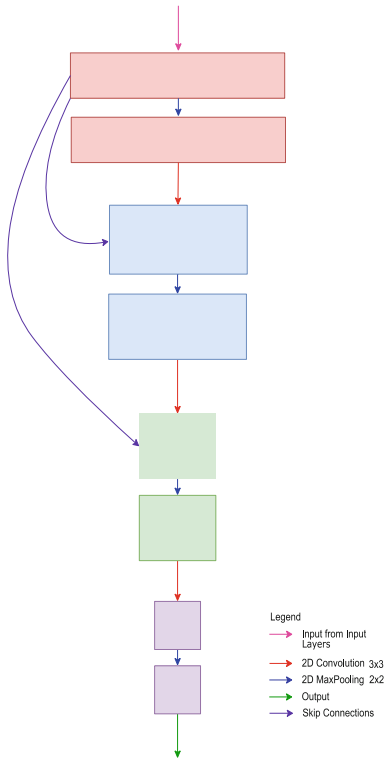
Fig. 1. ResNet
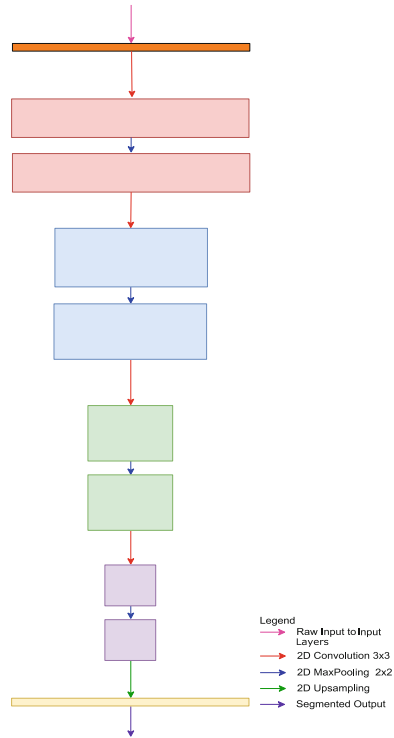
Fig. 2. FCN

2.2 FCN

Fully convolutional neural network (FCN) [4] was one of the premier neural networks for semantic segmentation that came after AlexNet [5], VGGnet [6] and GoogleLeNet [7] (Fig.2). It used pre-trained backbones from the above-mentioned architectures as its input layer and used 2D Convolution followed by 2D Max pooling layers to downsample the images using shallower layers to coarse layers until the layers became so small that the network could use pixel-wise classification to extract features. In the final layer, it used 2D upsampling to reconstruct the segmented images with the pixel-wise annotations. However, the loss of information during downsampling was considerably high, which was found by comparing the reconstructed images with the original images.

2.3 U-Net

U-Net is a neural network that is mainly used for classification and localization. It consists of an encoder and decoder structure. The encoder downsamples the images using 2D Convolution and 2D Max pooling as a way to extract pixel-wise features from the images. The decoder uses the input from the long skip connections, similar to Resnet, to reconstruct the image using 2D upsampling with segmented annotations (Fig. 3).
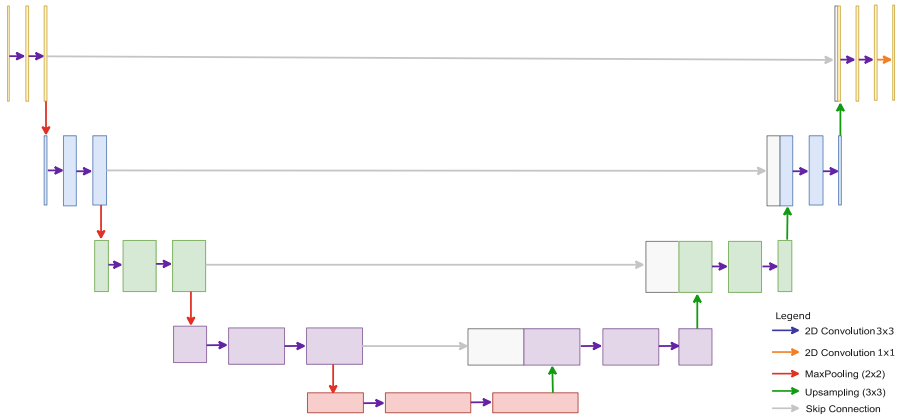
Fig. 3. U-Net

2.4 3D U-Net

3D U-Net [8] is a variant of U-Net where the inputs are voxels/3D images. It has the same encoder and decoder structure as in U-net. However, the encoder uses 3D Convolution followed by 3D Max pooling to extract the features and the decoder uses 3D upsampling to reconstruct the segmented annotated images. The main success of 3D U-Net is that the contextual information from the z-axis is preserved, which would have been lost due to generation of 2D slices from 3D images. This is prevalent in other 2D CNNs such as U-net as they tend to loses all form of contextual information in the z-axis. However, this capability of preserving the z-axis comes at a price for 3D CNNs such as 3D U-Net. These 3D architectures require a substantial amount of training time compared to their 2D counterparts.

3   Results

This section outlines our review based on the performance analysis of CNN's in different studies on liver segmentation. Table 1 shows studies and their CNN architectures that were discussed in this section. Table 2 describes the results and different metrics that those CNN architectures used. Furthermore, Table 3 describes the dataset along with the organ and the data modality. For the scope of our work, we have taken the results of architectures that best fit this studies interest. The sign * has been used after name of those papers from which we have shown partially results.

Table 1. List of our selected deep learning architectures used in different publications for liver image segmentation

Architectures	Listed as	Name of the publications
ResNet	AHCNet [9]	AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liver Tumor Segmentation in CT Volumes
	H-DenseUnet [10]	H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes
	H. S. Hoang [11]	An Evaluation of CNN-based Liver Segmentation Methods using Multi-types of CT Abdominal Images from Multiple Medical Centers
FCN	Xiao *et al.* [12]	Hematoxylin and Eosin (H&E) Stained Liver Portal Area Segmentation Using Multi-Scale Receptive Field Convolutional Neural Network
	AHCNet [9]	AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liv-er Tumor Segmentation in CT Volumes
	L. Zhang [13]	An Automatic Liver Segmentation Algorithm for CT Images U-Net with Separated Paths of Feature Extraction
	Y. Enokiya [14]	Automatic Liver Segmentation Using U-Net with Wasserstein GANs
	N. Nasiri [15]	A Controlled Generative Model for Segmentation of Liver Tumors
U-Net	M. Rezaei [16]	Instance Tumor Segmentation using Multitask Convolutional Neural Network
	H-DenseUnet [10]	H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes
	Xiao *et al.* [12]	Hematoxylin and Eosin (H&E) Stained Liver Portal Area Segmentation Using Multi-Scale Receptive Field Convolutional Neural Network
3D U-Net	Z. Tang *[17]	An Augmentation Strategy for Medical Image Processing Based on Statistical Shape Model and 3D Thin Plate Spline for Deep Learning
	Liu *et al.* *[18]	Liver CT Sequence Segmentation Based with Improved U-Net and Graph Cut

Table 2. A brief summary of different parametric estimation for review listed on Table 1

Publications	Architectures	Results (As per Evaluation Metric)					
AHCNet *[9]		Liver (Dice %)			Tumor (Dice %)		
	2D U-Net	0.923 ± 0.03			0.51 ± 25		
	ResNet	0.938 ± 0.02			0.6 ± 0.12		
L. Zhang. [13]		Iou		Recall		Accuracy	
	U-net	0.9983		0.9329		0.9983	
	Improved U-net	0.9979		0.9311		0.9979	
Y. Enokiya * [14]		Dice (%)		Precision		Recall	
	U-Net	0.89		0.95		0.84	
	Modified U-Net	0.89		0.94		0.85	
	Modified U-Net + GAN	0.91		0.95		0.88	
	Modified U-Net + WGAN	0.92		0.94		0.91	
N. Nasiri [15]		Dice (%)		VOE		MSSD	
	U-Net	0.73		39		87	
	Proposed	0.86 ± 0.06		29.43		21.76	
H. S. Hoang. * [11]		Dice (%)		Hausdorff (mm)		MSD (mm)	
		LiTS	Mayo	LiTS	Mayo	LiTS	Mayo
	FCN-CRF	92.4 ± 6.1		91.8 ± 3.4		207.7 ± 69.5	
	DRIU	93.8 ± 1.2		90 ± 2.4		428.0 ± 36.2	
	V-net	73.7 ± 15.9		91.6± 4.0		381.7 ± 40.7	
M. Rezaei. * [16]		Dice (%)	Hausdorff distance	Sensitivity		Specificity	
	U-Net	0.70	18.06	0.821		0.987	
	Our method	0.734	13.86	0.84		0.93	
Xiao et al. * [12]		Iou (mean)			Sensitivity (mean)		
	FCN	0.81			0.87		
	U-Net	0.83			0.87		
	SegNet	0.78			0.89		
	Proposed	0.87			0.92		
Z. Tang * [17]		Accuracy					
	Multi-Scale 3D CNN	90.89 %					
	3D U-Net	88.34%					

Publica-tions	Architec-tures	Results (As per Evaluation Metric)					
		Lesion		Liver			
H-Dens-eUnet. * [10]		Dice per case	Dice global	Dice per case	Dice global		
	U-Net	65.0	-	-	-		
	ResNet	67.0	-	-	-		
	2D DenseNet with pre-trained model	68.3	81.8	95.3	95.9		
	2D Dens-eUNet with pre-trained model	70.2	82.1	95.8	96.3		
	H-DenseUnet	72.2	82.4	96.1	96.5		
Liu et al. * [18]		Dice (%)	VOE (%)	RVD (%)	ASD (mm)	RMSD (mm)	MSD (mm)
	FCN-8s	89.3 ± 3.45	8.66 ± 5.04	-1.25 ± 10.93	6.19 ± 2.32	12.49 ± 6.16	52.48 ± 12.42
	U-Net	82.76 ± 4.52	25.24 ± 6.07	-2.56 ± 11.04	0.25 ± 3.12	16.72 ± 6.18	2.78 ± 42.14
	H-DenseUnet	96.5	7.4	1.8	1.450	-	27.118
	3D-FCN	-	9.36 ± 3.34	0.97 ± 3.26	1.89 ± 1.08	4.15 ± 3.16	33.14 ± 16.36
	3D U-Net	92.8	-	-	-	-	-

Table 3. A brief summary of dataset and its modality used in the reviewed publications for liver segmentation listed in Table 1, and Table 2

Dataset	Anatomy	Modality	Publication
LiTS2017: Liver Tumor Segmentation Challenge [19]	Liver	CT	L. Zhang [13]
			H. S. Hoang [11]
			H-DenseUnet [10]
			M. Rezaei [16]
	Description: LiTS 2017 is a liver CT scan dataset. It was part of ISBI2017and MICCAI2017 challenge. It consists of 130 CT scans as training and 70 CT scans as test data		
3D-IRCADb (3D Image Reconstruction for Comparison of Algorithm Database) [20]	Liver	CT	AHCNet [9]
			Y. Enokiya [14]
			H-DenseUnet [10]
	Description: 3D-IRCADb is a database that consists of two datasets. 3D-IRCADb-01 is composed of 20 CT scans 3D-IRCADb-01 is composed of 2 CT scans		
Xiao *et al.* [12]	Liver	Whole slide imaging	Xiao *et al.* [12]
	Description: The dataset contains 20 whole slide images from different cases in which the training set involves 15 cases and the testing set contains 5 cases		

4 Discussion

In the years 2017, 2018, and 2019, the number of studies on deep learning architecture for liver image segmentation has risen significantly. Most of studies use different variant of state-of-the-art architectures. The majority of the liver segmentation images are based on the CT images. As seen from result section, ResNet, FCN, U-Net, 3D U-Net have been used routinely in recent years. Some different variation of these have also been tested in search of better performance. Liver dice and tumor dice metric have been used apart from different metric like VOE, RVD, ASSD, and MSSD. Dice coefficient is the most common and widely used metric in terms of liver segmentation.

In this study, we also found differences between performance of same architecture on the task of liver segmentation. For instance, AHCNet [9] reported a dice score of 0.923 ± 0.03 for liver segmentation using 2D U-Net, whereas Enokiya [14], Nasiri [15] and Rezaei [16] respectively reported dice score of 0.89, 0.73 and 0.70 on liver segmentation task using same kind of 2D U-Net architecture. Clearly, this is a significant difference in performance when it comes to the same dataset and architecture. The different training and testing data sizes may result a change in performances apart from other factors such as hyperparameter tuning. Therefore, any fixed numbers out of certain training and

testing data may be helpful to have a benchmarking in reporting performances all over the architectures.

Z. Tang [17] used accuracy in percentage to report the results of their 3D U-Net whereas, Liu *et al.* [18] used dice coefficient score to report their results. As for IoU, L. Zhang [13] reported an IoU score of 0.89 using U-Net. On the other hand, Xiao *et al.* [12] reported a mean IoU of 0.83. This kind of variation of results from same architecture on same dataset requires more future analysis and robust benchmarking. Different metric to evaluate performance will always lead to the difference in results. Having some regulation on using specific metric measure for certain type of organ's data will be helpful to compare overall performance of different architectures.

Finally, the results of Sect. 3 will be helpful for the future studies on liver image segmentation. This study will provide some insight into different metrics and variation of results from same architecture. The variation of performance for the certain datasets on the same deep learning architecture mentioned here will in help set up better standards for evaluation metrics, performance evaluation and result submissions. Using standard set of metrics and robust benchmarking will ensure better analysis of performances of deep CNN architectures.

5   Conclusion

In this paper, we reviewed the performance of different liver segmentation studies which used a couple of major architecture like ResNet, FCN, 2D and 3D U-Net. To date, many more approaches and algorithms are on the verge of publication. For further liver semantic segmentation or other biomedical segmentation research work, we hope that this paper will provide a good starting guide for selecting different performance metrics and deep learning architecture.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. Springer, Cham (2015)
2. Reinke, A. et al.: How to exploit weaknesses in biomedical challenge design and organization. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol. 11073. Springer, Cham (2018)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 770–778 (2016). https://doi.org/10.1109/cvpr.2016.90
4. arXiv:1411.4038

5. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural net-works. In: NIPS (2012)
6. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2014)
7. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)
8. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol. 9901. Springer, Cham (2016)
9. Jiang, H., Shi, T., Bai, Z., Huang, L.: AHCNet: an application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes. IEEE Access **7**, 24898–24909 (2019). https://doi.org/10.1109/ACCESS.2019.2899608
10. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C., Heng, P.: H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from ct volumes. IEEE Trans. Med. Imaging **37**(12), 2663–2674 (2018). https://doi.org/10.1109/TMI.2018.2845918
11. Hoang, H.S., Phuong Pham, C., Franklin, D., van Walsum, T., Ha Luu, M.: An evaluation of CNN-based liver segmentation methods using multi-types of CT abdominal images from multiple medical centers. In: 2019 19th International Symposium on Communications and Information Technologies (ISCIT), Ho Chi Minh City, Vietnam, pp. 20–25 (2019)
12. Xiao, Q., et al.: Hematoxylin and Eosin (H&E) stained liver portal area segmentation using multi-scale receptive field convolutional neural network. IEEE J. Emerg. Sel. Top. Circ. Syst. **9**(4), 623–634 (2019). https://doi.org/10.1109/JETCAS.2019.2952063
13. Zhang, L., Xu, L.: An automatic liver segmentation algorithm for ct images u-net with separated paths of feature extraction. In: 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, pp. 294–298 (2018). https://doi.org/10.1109/icivc.2018.8492721
14. Enokiya, Y., Iwamoto, Y., Chen, Y.-W., Han, X.-H.: Automatic liver segmentation using U-Net with Wasserstein GANs. J. Image Graph. **6**(2), 152–159 (2018). https://doi.org/10.18178/joig.6.2.152-159
15. Nasiri, N., Foruzan, A.H., Chen, Y.: A controlled generative model for segmentation of liver tumors. In: 2019 27th Iranian Conference on Electrical Engineering (ICEE), Yazd, Iran, pp. 1742–1745 (2019). https://doi.org/10.1109/iraniancee.2019.8786681
16. Rezaei, M., Yang, H., Meinel, C.: Instance tumor segmentation using multitask convolutional neural network. In: 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, pp. 1–8 (2018). https://doi.org/10.1109/ijcnn.2018.8489105
17. Tang, Z., Chen, K., Pan, M., Wang, M., Song, Z.: An augmentation strategy for medical image processing based on statistical shape model and 3D thin plate spline for deep learning. IEEE Access **7**, 133111–133121 (2019). https://doi.org/10.1109/ACCESS.2019.2941154
18. Liu, Z., Song, Y.-Q., Sheng, V.S., Wang, L., Jiang, R., Zhang, X., Yuan, D.: Liver CT sequence segmentation based with improved U-net and graph cut. Expert Syst. Appl. (2019). https://doi.org/10.1016/j.eswa.2019.01.055
19. LiTS Challenge. https://competitions.codalab.org/competitions/17094
20. D-IRCADb database. https://www.ircad.fr/research/3dircadb/

Application of Image Segmentation and Convolutional Neural Network in Classification Algorithms for Mammary X-ray Molybdenum Target Image

Minghuan Zhang[1], Wenjian Liu[3], Xuan Zhang[1], Ying Chen[1(✉)], Yajia Gu[2], and Qin Xiao[2]

[1] Research Center of Big Data Analyses and Process,
Shanghai Sanda University, Shanghai, China
`zmh_cd@163.com`
[2] Shanghai Cancer Center, Fudan University, Shanghai, China
[3] Faculty of Humanities and Social Sciences, City University of Macau, Macau SAR, China

Abstract. Breast cancer is the second leading cause of death in women worldwide and of the methods for the detection of breast cancer, Mammography is considered promising and effective. In order to improve the detection, the study explored automatic recognition of Mammary X-ray Molybdenum target images on the basis of the clustered distribution pleomorphic calcification images of Breast Imaging Reporting & Data System category 4 (BI-RADS 4) obtained from an open access database – Digital Database for Screening Mammography (DDSM). The region of interest (ROI) of molybdenum target images was firstly segmented into sub-images by coordinate matching technology, and then the sub-images were scanned row by row and subdivided into mini-images. Those mini-images containing lesions were thus screened out and used as the objects of neural network recognition. Pattern recognition was carried out via the classical convolutional neural networks such as VGGNet16, VGGNet11 and AlexNet, and the improved AlexNet network without LRN layer. The results showed that identification and subdivision of the ROIs together with the improved AlexNet network could significantly improve the performance of recognition. By comparison with other methods, the new methods developed herein could provide additional and useful information for clinical diagnosis, and lay a technical foundation for refining classification of BI-RADS4 images into sub-categories and furthering accurate diagnosis.

Keywords: DDSM · BI-RADS4 images · Image segmentation · Pattern recognition · CNN

1 Introduction

Breast cancer, the second leading cause of death in women worldwide [1, 2], has long been presenting a serious threat. The process of detecting the breast cancer prior to symptoms is called as screening. Mammography, a kind of x-ray image screening method,

is considered the promising and effective method for the detection of breast cancer [3]. On the basis of the evidence supported by randomized controlled trails and observational studies, World Health Organization (WHO) in 2016 concluded that having women received mammography screening could be helpful to reduce the risk of breast cancer mortality by about 40% [4]. To reduce the variations in the radiologists' descriptions of mammogram findings, the American College of Radiology (ACR) developed a Breast Imaging Reporting and Data System (BI-RADS) [5]. BI-RADS provides a standard lexicon to describe mammogram findings and a classification system with 6 categories to assess the likelihood of malignancy, of which Category 4 (BI-RADS 4) indicated that the risk of malignant tumors was 3% through 94%, leaving a great uncertainty.

The Digital Database for Screening Mammography (DDSM), a digital mammography database developed by the University of Florida for breast cancer screening, contains image data of more than 2,000 cases. Each case included four images of the left and right breasts from the Cranio-Caudall (CC) and Medio-Lateral Oblique (MLO) perspectives [6]. According to the diagnostic results, these images were divided into four categories: malignant lesions, benign and without re-diagnosis and normal.

Thanks to DDSM, this study accessed to these BI-RADS 4 images with a view to efficiently identifying the nature of lesion areas contained in the images, benign or malign, by application of artificial intelligence technology. To this end, Image segmentation is the very starting step [7], which is the fundamental basis and prerequisite for image feature extraction, parameter selection and pattern recognition. The ultimate goal is to automatically recognize images with high accuracy and efficiency [8].

Because of the high spatial resolution of the molybdenum target images in DDSM and the existence of a large number of invalid regions in the image, they appear not suitable for direct analysis and recognition. In addressing this problem, an alternative is identification and utilization of the region of interest (ROI). The most commonly used methods to identify and segment the ROIs include threshold segmentation, region segmentation, edge segmentation, clustering segmentation and model segmentation [9]. A comprehensive comparison of these methods suggested a method on the basis of coordinate matching and row by row scanning to subdivide sub-images can effectively solve the problem of suitability.

With respect to analysis and recognition of medical images, computer-aided detection (CAD) has been widely used. Most of the classification algorithms in pattern recognition can be applied to the recognition of breast tumor images. Pena-Reyes and Sipper combined fuzzy system with genetic algorithm for breast tumor recognition [10]. Abonyi and Szeifert classified breast tumor data sets by constructing supervised clustering model [11]. Lin et al. used particle swarm optimization algorithm to find the best parameters of SVM for breast mass recognition, the accuracy rate of which is much higher than that of the same kind of classifiers [12]. Dheeba, Singh and Selvi designed a wavelet neural network based on PSO algorithm to identify breast tumor data [13].

In addition to tradition CADs, the development of deep learning has achieved a great success in various applications [14, 15]. Convolutional neural networks (CNNs) [16] can realize automatic feature extraction and joint training of classifiers because of its

strong ability of feature expression and information extraction. Remarkable achievements have been made on CNNs in respect of image classification [17], target detection [18] and image segmentation [19]; recently CNNs have been used in the field of automatic breast image recognition. Impressive performances were made for classifying breast mass (Huynh et al. [20], Kooi et al. [21], Kim et al. [22], Samala et al. [23]). Castro et al. developed machine learning based on natural language processing (NLP) systems, including Naïve Bayes and support vector machines (SVMs), for automatic extraction of BI-RADS assessment category values and classification of the laterality [24].

In this study, the mini-images of the ROIs automatically segmented from the molybdenum target images were then subject to the test of image classification through the use of CNNs. The test results showed that the improved neural network produced agreeable performance, which can provide additional and useful information for clinical diagnosis, and lay a technical foundation for subdividing BI-RADS 4 and furthering accurate diagnosis. All programs were implemented in Python and all models were using the TensorFlow [25] library with CUDA backend.

2 Materials and Methods

As above mentioned, all images tested in this study are available in DDSM. The X-ray images of breast cancer usually can be classified into three categories: calcification, mass and architectural distortion, of which clustered micro-calcifications constitute a major mammographic sign of breast cancer. Therefore, the detection and identification of calcified lesions in breast X-ray images has become one of research hotspots in breast cancer screening CAD system. Of calcified lesions, more common are pleomorphic calcification lesions with clustered distribution pleomorphic calcification [26]. Given the great uncertainties in BI-RADS 4 images mentioned above, this study chose this category of images as the test object. Images of 20 cases with benign and malignant calcification confirmed by pathology were selected respectively; 40 cases in total. Figure 1 are BI-RADS 4 image examples, benign (a) and malignant (b) respectively.
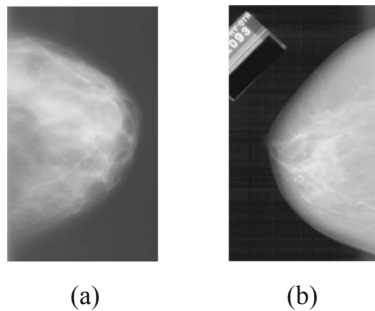

(a)   (b)

Fig. 1. Experimental image examples from DDSM. a A benign BI-RADS 4 image, b A malignant BI-RADS 4 image.

Each case in DDSM has the raw mammography (hereinafter as 'the raw image') and the subsampled image (hereinafter as the 'the thumbnail image'). The lesion areas were calibrated with red circle on the thumbnail image. But the thumbnail images were subsampled, sharpened and converted to eight bits per pixel for visual browsing purpose; because of containing artifacts introduced by this processing, they are not suitable for experimental work. On the other hand, the raw images are too large to process though DDSM suggests that any experimental work should start with the full raw images for the case. The size of a single raw image can be as large as tens of megabytes, and there are a large number of invalid areas. Therefore, the raw images are not suitable to be used for directly analysis and recognition. Because the lesion areas were not marked in the raw images and the vertical and horizontal ratio of the thumbnail images to the raw images remains unchanged, the position of the lesion areas in the raw images can be located according to the coordinates of the lesion area on the thumbnail images. This is the way that this study designed to identify the ROIs in and to segment the ROIs from the raw images.

3   Image Processing and Model Designing

3.1   ROI Location

In order to identify the ROIs in the raw images, it is necessary to acquire the coordinates of the lesion areas in the thumbnail images. To this end, it was firstly to scan the thumbnail images longitudinally and horizontally. When a red pixel was encountered, the position coordinates of that point were recorded. Thereafter, the external rectangle of the lesion area was drawn according to several relevant pixel points, and the coordinates of the upper left corner A and the lower right corner B of the external rectangle were obtained. Using the coordinates of A and B, the lesion area in the raw images can be determined. The process of locating ROI is shown in Fig. 2.
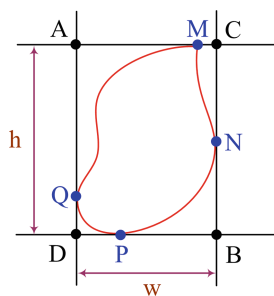

Fig. 2. The process of locating the ROIs. MNPQ is a lesion area identified in red by experienced doctors on the thumbnail images. ACBD is its external rectangle. h and w are the length of AD and DB, respectively.

As shown in Fig. 2, the coordinate of point A was set as A (XA, YA). If the height and width of the lesion area were h and w, then the coordinates of point B could be expressed as (XA+w, YA+h). The positioning process can be described as follows:

1. The position coordinates of a pixel were recorded when it was detected that the gray value of a pixel was close to a value of (255, 0, 0).
2. Horizontal and vertical coordinates recorded above were written into X_list and Y_list respectively. Therefore, the elements in X_list included at least [XM, XN, XP, XQ], and the elements in Y_list were [YM, YN, YP, YQ] accordingly.
3. In accordance with the geometric relationship, it can be concluded that the coordinates of point A and point B of rectangular ACBD outside the lesion area could be expressed as A (minimal (X_list), minimal (Y_list)) and B (maximal (X_list), maximal (Y_list)) respectively.
4. The coordinates of A and B on the thumbnail image were those of the ROI to be segmented from the raw image. Figure 3 is the example of BI-RADS 4 benign image 'B_3126_1.LEFT_MLO', showing the result of segmentation.
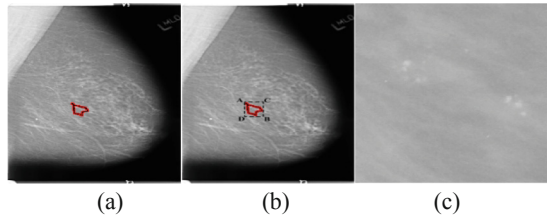

| (a) | (b) | (c) |

Fig. 3. Segmentation of the ROIs on the raw images. a The thumbnail image with marked a lesion area in red. b The external rectangle of the lesion area. c The ROI segmented from the raw image.

As Fig. 3(c) illustrating, although there are still some invalid areas in the ROI, it is easier to identify the location of the calcification with the ROI in comparison with the large size of the raw image. By virtue of this method, the ROIs of all cases were segmented in batches.

3.2   Mini-image Segmentation

According to the method designed above, the ROIs of 40 cases were acquired as the first step of this study. However, there are still some invalid regions in the acquired ROIs, which need to be further processed. Therefore the second step is to further narrow the ROIs. To this end, the ROIs were scanned row by row and then the mini-images were selected. One image of the ROI, as shown in Fig. 3(c), may be further divided into several mini-images containing calcified lesions. Under the guidance of experienced doctors, the mini-images having obvious lesions were sorted out.

To further subdivide mini-images sorted out with the above step, a frame of 64 pixels was designed moving longitudinally and horizontally from the origin of coordinates, as shown in Fig. 4. The interval of move was set at 15 pixels. Then the ROI image was scanned row by row in turn. At the end, a great number of mini-images were segmented and obtained.

Likewise, the frames of 128 pixels and 256 pixels were designed and used to scan the ROIs for further subdivision respectively. The results of subsequent experiments showed
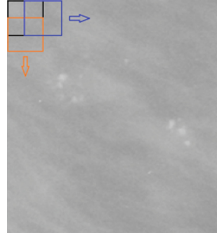
Fig. 4. The process of subdividing the ROI by row scan. The black frame is a mini-image obtained by without moving, the blue one is another mini-image obtained by moving an interval to the right by 15 pixels, and the yellow one is a mini-image obtained by moving by the same interval downwards.

that the frame of 128 pixels is the best. The frame of 64 pixels is too small, resulting in the large proportion of the lesion area in the mini-images and the characteristics not obvious. On the contrary, the frame of 256 pixels is too large, resulting in the small proportion of the lesion area in the mini-images and segmenting the ROI ineffectively. Therefore, the frame of 128 pixels is appropriate and was used for the next step.

If the characteristics of the mini-images are too obvious, there may be over-fitting in the subsequent pattern recognition. That is, the accuracy of the model on the train set is very high, but the generalization performance (the accuracy of the model on the test set) is poor. Therefore, when selecting the segmented mini-images, a number of mini-images containing incomplete lesions should be added to enhance the generalization performance of the model (Fig. 5).
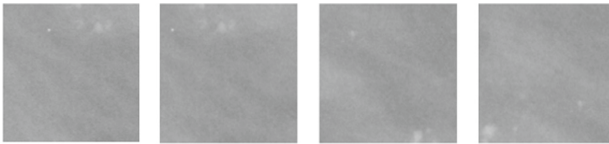

Fig. 5. Mini-images having incomplete lesions. The above four mini-images are an example of how the mini-images including incomplete lesions were sampled, to enhance the generalization performance of the model.

According to the method designed above, the images of the ROIs for all 40 cases were in batches, resulting in a total of 3886 mini-images. With the guidance of radiologists, 1236 effective learning samples were selected, of which 608 mini-images contain benign lesions and 628 mini-images malignant lesions. The mini-images are all 128 pixels in size.

3.3   Model Design

There are many classical convolution neural network models, including VGGNet [27] and AlexNet [28]. The experiments in a 16-layer VGG convolution neural network (VGG16), an 11-layer VGG convolution neural network (VGG11) and a classical

AlexNet showed that the direct application of classical models to recognize all 1236 mini-images was not satisfactory. Therefore, it is desirable to improve these classical models. Here improvement was made on the AlexNet. In general, the number of classical AlexNet network layers is 8 layers, including 5 layers of convolution layer and 3 layers of full connection layer, as shown in Fig. 6.
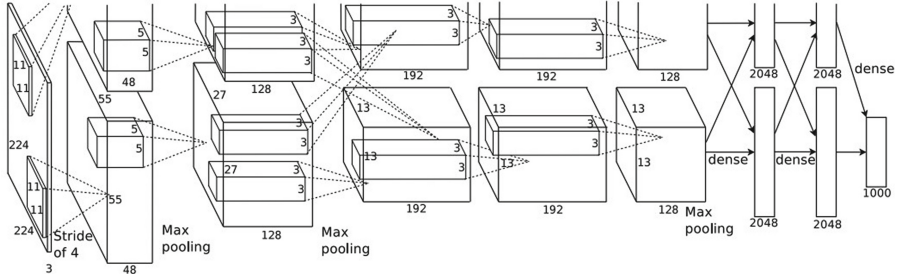

Fig. 6. The classical AlexNet network structure, including 5 layers of convolution layer and 3 layers of full connection layer.

One of the characteristics of the AlexNet is the addition of Local Response Normalization (LRN) layer, which is used to normalize the output of ReLU activation function. Formula 1 is the output of LRN.

$$b^i_{(x,y)} = \frac{a^i_{(x,y)}}{(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{(x,y)})^2)^\beta} \tag{1}$$

Where $a^i_{(x,y)}$ denotes the output of ReLU in the (x, y) position of i, n the number of adjacent pixels involved in local normalization of $a^i_{(x,y)}$, N the total number of kernel, and $b^i_{(x,y)}$ the output result of LRN. That is, LRN locally normalizes the output of ReLU and the pixels around it, creates a competitive mechanism for the activity of local neurons, suppresses the adjacent neurons of activated neurons, and makes the larger response value relatively larger, so as to improve the generalization performance of the model.

The experimental images are gray ones. Except for few calcification points having higher brightness and higher gray value, the gray values of most areas of the images are rather low. Under normal circumstance, after mini-images were segmented by using CNNs and if the pixel values of calcification points and their adjacent areas were normalized, the high brightness features of calcification points would be reduced. In this case, the characteristics of calcification points would be weakened, not helpful for pattern recognition. Therefore, it was aspired to improve the AlexNet network by removing the LRN layer. The removal of the LRN layer was carried out after activating the function ReLU in convolution layer 1 (conv1) and convolution layer 2 (conv2). Improvement was made on the convolutional layers of classical AlexNet network as shown in Table 1.

Table 1. Improvement on the convolutional layers of classical AlexNet network.

Conv layer	Description	Conv layer parameter
1	Convoluting (including activating) → pooling	Convolutional kernel size: 11 * 11 * 3; LRN layer removed
2	Convoluting (including activating) → pooling	Convolutional kernel size: 5 * 5* 64; LRN layer removed
3	Convoluting (including activating)	Convolutional kernel size: 3 * 3 * 192
4	Convoluting (including activating)	Convolutional kernel size: 3 * 3 * 384
5	Convoluting (including activating) → pooling	Convolutional kernel size: 3 * 3 * 256

With the improvement, all activating layers in the improved AlexNet network were activated by ReLU function, all pooling layers were subsampled by Max Pooling, so as to avoid the blurry effect of Average pooling and retain the most prominent features. The Gaussian distribution function with standard deviation of 0.05 was used to initialize the network weight, the loss function of the network was defined by the cross entropy function, and the Adam Optimizer function was used to minimize the loss function. Meanwhile, in order to reduce the adverse effects of over-fitting, 50% of all neurons were retained by Dropout method. Because the number of network structure randomly generated was the largest when dropout was 0.5, it can be seen after cross-verification that the classification effect is the best when dropout is 0.5.

In respect to sample data organization, random sampling method and used ten-fold cross validation method were adopted. In the training set and the test set, the distribution of benign and malignant mini-images was close to the overall distribution. Three parameters such as sensitivity, specificity and accuracy percentage were chose to assess the performance of all models [29]. The confusion matrix [30] of metrics for performance evaluation was given in Table 2.

Table 2. Confusion matrix.

	Predicted positive (malignant)	Predicted negative (benign)
Real positive (malignant)	True positive (TP)	False negative (FN)
Real negative (benign)	False positive (FP)	True negative (TN)

The sensitivity is:

$$\text{Sensitivity} = \text{TP} \times 100\%/(\text{TP} + \text{FN}) \qquad (2)$$

The specificity is:

$$Specificity = TN \times 100\%/(TN + FP) \tag{3}$$

The accuracy is:

$$Accuracy = (TP + TN) \times 100\%/(TP + FP + FN + TN) \tag{4}$$

The values of sensitivity, specificity, and accuracy were set between [0, 100%] respectively; the closer they are to 100%, the better the performance of the model is.

Receiver operating characteristic (ROC) curve can be used to measure the imbalance in classification. ROC curve and its area under the curve (AUC) are often used to evaluate the advantages and disadvantages of a binary classifier. The horizontal coordinate of ROC is false positive rate (FPR) and the vertical coordinate is true positive rate (TPR). In this experiment, FPR indicates how many of all benign samples were predicted as malignant samples incorrectly, and TPR indicates how many real malignant samples were predicted correctly. The closer the ROC curve is to the upper left corner of the graph, the better the performance of the classifier is. At the same time, the closer the value of AUC between [0.5, 1] is to 1, the better the performance of the classifier is.

4   Results

The classical networks such as VGG and AlexNet were used for train recognition models respectively, and the improved AlexNet network was designed in accordance with the section of 'Model design'. Figure 7 showed the index trend comparison of convolution neural networks.

As mentioned before, VGG16 is short for a 16-layer VGG convolution neural network and VGG11 an 11-layer VGG convolution neural network. The AlexNet_v2 denoted a classical AlexNet convolution neural network, and the Alexnet_v1 the improved AlexNet with the removal of LRN. The transverse coordinates of Fig. 7 are the test times; the longitudinal coordinates of Fig. 7(a), (b) and (c) are the sensitivity, specificity and accuracy on the test set respectively. The effective test times of each model are more than 75. In light of the trend of curve diagrams, it can be seen that the prediction effect of the AlexNet_v1 model is the best and the AlexNet_v1 is the most stable compared with other models. Table 3 shows the minimum and average sensitivity, specificity and accuracy of the 75 tests.

The results show that the ROC of AlexNet_v1 is closest to the upper left corner of the figure. Meanwhile, the AUC of AlexNet_v1 is 0.982, the highest among four tested models.
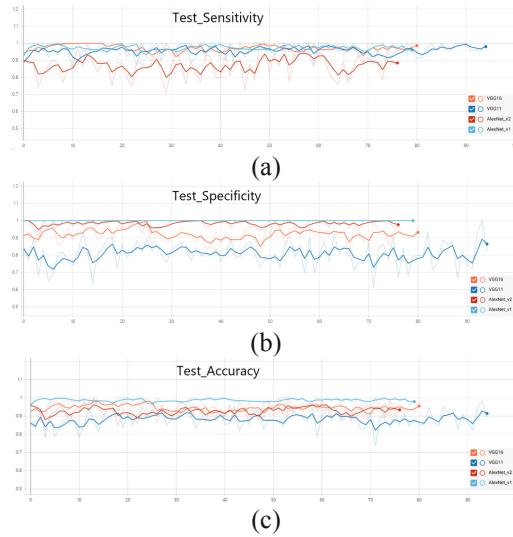
(a)

(b)

(c)

Fig. 7. The index trends of each convolution neural network. a The sensitivity on test set of each convolution neural network. b The specificity on test set of each convolution neural network, c The accuracy on test set of each convolution neural network.

Table 3. Assessment indices of all models of 75 tests.

Model	min_sens	avg_sens	min_spec	avg_spec	min_accur	avg_accur
VGG16	85.71%	97.06%	76.92%	92.22%	86.27%	94.54%
VGG11	87.50%	95.44%	61.11%	80.92%	73.53%	88.16%
AlexNet_v2	70.83%	87.06%	90.91%	98.33%	81.63%	92.84%
AlexNet_v1	86.96%	97.35%	100.00%	100.00%	93.33%	98.68%

min_sens and avg_sens represent minimum sensitivity and average sensitivity of the 75 tests, respectively. min_spec and avg_spec indicate minimum specificity and average specificity of the 75 tests, respectively. min_accur and avg_accur mean minimum accuracy and average accuracy of the 75 tests, respectively.

5   Discussion

The experimental results show that the designed methods for processing image samples and pattern recognition could achieve excellent classification performance. By contrast, should the images of the ROIs be subject to the training set directly without further segmentation, the sensitivity, specificity and accuracy of the test set would only be 54.3%, 50.3% and 52.2% respectively. Therefore, innovative methods for image process and improved models are essential for pattern recognition.

Whereas the methods herein were developed on the basis of DDSM, they could be applicable for images collected in practice from clinic. In that case, when sample images collected, the doctor should be consulted, the lesion area be marked in red as the

thumbnail images in DDSM, and the original images be maintained intact as the raw images. Because red marks of the lesion areas prevent the marked images from further process, the method developed here to acquire the coordinates of the lesion areas could be useful. Otherwise, segmentation of the marked images would be deemed to include red marks in the mini-images, giving rise to problems for pattern recognition in the later stage. Should the coordinates of the lesion area be acquired from the marked images with marks, the ROI could be identified in the original images and be segmented with the method of coordinates demonstrated in Fig. 2.

Arguably, the methods explored in this study could also be applicable for large-scale breast cancer screening. To this end, these negative breast X-ray images which usually account for the majority should be excluded by virtue of CAD technology. That will leave the positive or suspected positive breast X-ray images submitted to radiologists for further diagnosis. By this step, the efficiency of screening can be increased, the rate of misdiagnosis and missed diagnosis reduced, and the workload of doctors substantially eased [31]. For this purpose, an automatically recognition system should be developed and built on the basis these methods. Then, when these positive and suspected positive images are sorted out, the system would automatically process them, including scan and segmentation, one by one and output mini-images for pattern classification. Should a mini-image of one case contain a prediction of possible malignant category, the case might be preliminarily identified as a potential malignant tumor one. Then, the doctor in charge should be reminded for further diagnose.

It should be pointed out that because of the possible individual difference among cases, sometime substantial one, the use of unified segmentation size in some complex cases may not be conducive to the correct classification of images. Under this circumstance, close cooperation with clinicians is necessary.

According to the BI-RADS standard, all mammary X-ray molybdenum target images could be divided into 6 grades, from grade 1 through grade 6, according to the risk of malignant tumor from low to high; grade 6 is the worst.

In general, images scoring BI-RADS 4 or BI-RADS 5 have been considered positive whereas those scoring BI-RADS 1, BI-RADS 2, or BI-RADS 3 were considered negative [32]. As Table 2 indicates, however, BI-RADS 4 covers the risk of malignant tumor from 3% to 94%, which is the most complicated one among all categories. The probability range is too large to be conducive for accurate diagnosis and corresponding treatment. This is the reason to subdivide BI-RADS 4 into three sub-categories, namely, BI-RADS 4a, BI-RADS 4b and BI-RADS 4c (Table 3).Were an BI-RADS 4 image to be further identified into one of sub-categories automatically based on artificial intelligence technology, the probability range of malignant tumors would be narrowed. The technologies as such are definitely beneficial to the accurate diagnosis and formulation of corresponding treatment plans.

This study is a preliminary experiment of the above problems. The designed methods together with the improved AlexNet network are proved to be able to recognize the benign and malignant lesions of calcified points in BI-RADS 4 images. In the future, the appropriate samples will be collected and the methods and models developed herein could extend to refine the classification of BI-RADS4 images into sub-categories.

6 Conclusion

This study developed new methods for obtaining the ROIs of breast X-ray target image, and improved the AlexNet convolution neural network to pattern recognize the clustered distribution pleomorphic calcification images of BI-RADS 4. The developed methods and models can also apply for large-scale breast cancer screening, and lay a technical foundation for refining classification of BI-RADS 4 images into sub-categories.

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics. CA Cancer J. Clin. **68**, 7–30 (2018)
2. Canadian Cancer Society's Advisory Committee on Cancer Statistics: Canadian Cancer Statistics 2015. Toronto, Canadian Cancer Society (2015)
3. Khan, S., Khan, A., Maqsood, M., et al.: Optimized gabor feature extraction for mass classification using cuckoo search for big data e-healthcare. J Grid Comput **17**, 239–254 (2019)
4. Breast Cancer Screening: International Agency for Research on Cancer (IARC) Handbooks of Cancer Prevention. WHO and IARC, Lyon (2016)
5. D'Orsi, C.J.: ACR BI-RADS Atlas: Breast imaging reporting and data system. American College of Radiology, Reston (2013)
6. Lei, C., Wei, W., Liu, Z., Xiong, Q., et al.: Radiomics analysis for pathological classification prediction in BI-RADS category 4 mammographic calcifications. J. Clin. Oncol. **37**(15_suppl) (2019)
7. Pal, N.R., Pal, S.K.: A review on image segmentation techniques. Pattern Recognit. **26**, 1277–1294 (1993)
8. Laves, M.-H., Bicker, J., Kahrs, L.A., Ortmaier, T.: A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation. Int. J. Comput. Assist. Radiol. Surg. **14**(3), 483–492 (2019). https://doi.org/10.1007/s11548-018-01910-0
9. Xu, X., Xu, S., Jin, L., et al.: Characteristic analysis of Otsu threshold and its applications. Pattern Recognit. Lett. **32**, 956–961 (2011)
10. Pena-Reyes, C.A., Sipper, M.: A fuzzy-genetic approach to breast cancer diagnosis. Artif. Intell. Med. **17**, 131–155 (1999)
11. Abonyi, J., Szeifert, F.: Supervised fuzzy clustering for the identification of fuzzy classifiers. Pattern Recognit. Lett. **24**, 2195–2207 (2003)
12. Lin, S.W., Ying, K.C., Chen, S.C., et al.: Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Syst. Appl. **35**, 1817–1824 (2008)
13. Dheeba, J., Singh, N.A., Selvi, S.T.: Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. J. Biomed. Inform. **49**, 45–52 (2014)
14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)

15. Kim, S.T., Lee, J.H., Lee, H., et al.: Visually interpretable deep network for diagnosis of breast masses on mammograms. Phys. Med. Biol. **63**, 235025 (2018)

16. He, K., Zhang, X., Ren, S., et al.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: IEEE International conference on computer vision (ICCV), pp. 1026–1034 (2015)

17. Redmon, J., Divvala, S., Girshick, R., et al.: You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)

18. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**, 1137–1149 (2017)

19. Chen, L.C., Papandreou, G., Kokkinos, I., et al.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**, 834–848 (2018)

20. Huynh, B., Drukker, K., Giger, M.: MO-DE-207B-06: computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks. Med. Phys. **43**, 3705–3705 (2016)

21. Kooi, T., Litjens, G., Van Ginneken, B., et al.: Large scale deep learning for computer aided detection of mammographic lesions. Med. Image Anal. **35**, 303–312 (2017)

22. Kim, S.T., Lee, H., Kim, H.G., et al.: ICADx: interpretable computer aided diagnosis of breast masses. In: Proceedings of the SPIE, vol. 10575, p. 1057522 (2018)

23. Samala, R.K., Chan, H.-P., Hadjiiski, L.M., et al.: Evolutionary pruning of transfer learned deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis. Phys. Med. Biol. **63**, 095005 (2018)

24. Castro, S.M., Tseytlin, E., Medvedeva, O., et al.: Automated annotation and classification of BI-RADS assessment from radiology reports. J. Biomed. Inform. **69**, 177–187 (2017)

25. Wongsuphasawat, K., Smilkov, D., Wexler, J., et al.: Visualizing dataflow graphs of deep learning models in TensorFlow. IEEE Trans. Vis. Comput. Graphics **24**, 1–12 (2018)

26. Kikuchi, M., Tsunoda, H., Kaneshiro, T., et al.: A new method for differentiating benign and malignant pleomorphic clustered calcifications in mammography. J. Nippon Med. Sch. **81**(2), 70–77 (2014)

27. Singh, K.V., Romani, S., Barrena, T.J., et al.: Classification of breast cancer molecular subtypes from their micro-texture in mammograms using a VGGNet-based convolutional neural network. Recent Adv. Artif. Intell. Res. Dev. **300**, 76–85 (2017)

28. Zhu, L., Li, Z.B., Li, C., et al.: High performance vegetable classification from images based on AlexNet deep learning model. Int. J. Agric. Biol. Eng. **11**, 217–223 (2018)

29. Khan, S., Khan, A., Maqsood, M., Aadil, F., Ghazanfar, M.A.: Optimized gabor feature extraction for mass classification using cuckoo search for big data e-healthcare. Journal of Grid Computing **17**(2), 239–254 (2018). https://doi.org/10.1007/s10723-018-9459-x

30. Tang, Y., Zhang, Y.Q., Chawla, N.V., et al.: SVMs modeling for highly imbalanced classification. IEEE Trans. Syst. Man Cybern. **39**, 281–288 (2009)

31. Giger, M.L., Chan, H.P., Boone, J.: Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. Med. Phys. **35**, 5799–5820 (2008)

32. Mariscotti, G., Durando, M., Houssami, N., Fasciano, M., Tagliafico, A., Bosco, D., Casella, C., Bogetti, C., Bergamasco, L., Fonio, P., Gandini, G.: Comparison of synthetic mammography, reconstructed from digital breast tomosynthesis, and digital mammography: evaluation of lesion conspicuity and BI-RADS assessment categories. Breast Cancer Res. Treat. **166**(3), 765–773 (2017). https://doi.org/10.1007/s10549-017-4458-3

Fusion Segmentation of Head Medical Image with Partially Annotated Data

Xuzidui[(✉)], Guantian, and Heyonghong

Shenzhen Key Lab for Minimal Invasive Medical Tech, Graduate School at Shenzhen,
Tsinghua University, Shenzhen, China
xzd17@mails.tsinghua.edu.cn

Abstract. In this paper, there are 2 head computed tomography (CT) scan datasets: one with vision related organs pixel annotations and the other with auditory related organs pixel annotations. We aim to train a single neural network for vision related organs and auditory related organs segmentation at the same time with these 2 partially annotated datasets. An idea generating from co-operative training method will be applied to complete the lacking annotations of each dataset. However, it is not a proper way to treat the predicted annotations as the real annotations from professional doctors. To address this error, a semi-supervised method is chosen. Compared to the baseline method, our training pipeline is able to complete 2 segmentation tasks within only one model, and we have proved that it outperforms the baseline method. To some extent, using partially annotated medical image datasets can help to solve the problem that the scarce source of professionally annotated medical image data. What's more, the proposed method will achieve better performance.

Keywords: Medical image segmentation · Co-operative training · Semi-supervised learning

1 Introduction

It has no doubt that doctor is one of the hardest jobs in the world. One of the reasons is the examination of a lot of medical images, including computed tomography (CT) scans, magnetic resonance imaging (MRI) scans, whole slide images (WSI) and so on. Observing medical images is the most commonly used method by doctors to make diagnosis for patients. However, manually detecting specific phenomenon in these medical images can be tedious and time-consuming. Therefore, various computer-aided diagnosis (CAD) systems have been developed to assist doctors to examine these medical images.

Although deep CNN technology [10] has achieved great success in medical image processing, some difficulties still hamper its development. The most serious one is the scarce source of medical image data, especially professionally annotated medical image data. It is known to us that a large and professionally annotated dataset is necessary when training a neural network because the model's performance depends on the data it has

met. Accurate annotations are important, especially in medical image area. However, it is impossible for a professional doctor to annotate millions of medical images without any error in a short time. As a result, using partially annotated medical images is one of the proper ways to solve the problem.

The first plan for this question is also our baseline method: following supervised learning paradigm, we can train several models with these partially annotated datasets separately. Then, during the test step, a fusion process is needed to get the full segmentation results with several predicted segmentation results. Although this method is simple and reliable, large computational resource is necessary and the total number of models' parameters is too large to be applied in daily life.

Secondly, by treating this task as a multiple-objective task, multi-decoder architecture [6] can be adopted. Multi-decoder architecture is generated from the ideas of encoder-decoder architecture and multi-task learning. In this case, in order to complete several different segmentation tasks in a single model, some people proposed that multi-decoder architecture is a good choice for the reason that for similar tasks, the encoder part could be shared for the high-level semantic information of all objects is similar while the decoder part should be separate for the detail information of object boundary is different. During the same time, multi-decoder architecture has an additional advantage: it makes use of the correlation between different tasks to improve performance. However, it has 2 drawbacks: the first one is once the number of tasks is becoming larger, then the number of decoders will be larger, meaning the parameters will become too large; the second reason is, intuitively speaking, the correlation between different tasks is not only appeared in encoder part, but also in decoder part.

In this paper, we propose a pipeline to train a single model with 2 partially annotated datasets. To our knowledge, this is the first attempt to complete 2 different segmentation tasks in a single model. To achieve this objective, we have to address 2 problems: the first problem is how to complement the lacking annotations of each dataset. Because each dataset has its own annotations, 2 models will be trained to fit these 2 datasets separately, and then the prediction results by setting each dataset as the input of the model trained with the other dataset previously can be treated as the lacking annotations. [7] Obviously, the second problem comes: the quality of predicted annotations is not as good as original annotations from professional doctors. To prevent this performance degradation, a semi-supervised method should be applied for it will take the quality of annotations into consideration. What's more, our proposed method is robust to hyper parameters and models, meaning it can be applied to other areas.

2 Method

In this section, we will describe our full proposed training pipeline in detail. In our experiments, DeepLabv3+ [2] and U-Net are taken as the basic architectures of our segmentation models. As for the semi-supervised learning algorithm, the Mean Teachers algorithm [3] is adopted for its good performance. Based on these 2 ideas, we propose to train a model for 2 different but similar segmentation tasks with 2 partially annotated datasets.

2.1 DeepLabv3+

DeepLabv3+ employs the encoder-decoder structure. The encoder module is used to extract rich contextual information and the simple yet effective decoder module is adopted to recover the object boundaries. Specially, unlike common encoder-decoder architecture, the inputs of the decoder module consist of 2 different feature maps: one feature map from the end of encoder module containing high-level information, and the other feature map is from the begin of encoder module containing low-level information. High-level information tells the model semantic information about the object, while low-level information tells the model detail information about the object boundary to improve performance. DeepLabv3+ also apply the dilated convolution technology [8], which means to get various sizes of receptive field, to extract features at an arbitrary resolution, depending on the available computation resources. What's more, the Xception idea using depth-wise separable convolution technology [9] to decrease the number of parameters without hurting the performance, has also been explored.

2.2 U-Net

In order to verify our proposed method is robust to model selection, U-Net [1] is also chosen as our other base architecture. U-Net is the simplest and the most commonly used one in medical image segmentation. Encoder-decoder architecture is adopted for feature extraction and boundary recover. Skip connection idea is used to prevent the neglect of detail information about object boundary. In our experiment, because some organs are so small, for example, lens, U-Net will miss them by taking 5 down-sample operations. To address this question, only 3 down-sample operations will be adopted when extracting features.

2.3 Mean Teachers

In the region of semi-supervised learning technology, self-training method is one of the basic useful methods. For unannotated data, the predictions from pre-trained model with high confidence will be treated as real annotations to improve the generality of the model. In this view, another idea by making an ensemble operation of several pre-trained models has been proposed. Among them, Mean Teachers algorithm has been widely used and it has been proved to achieve a better performance and a faster training speed. In Mean Teachers method, 2 same models will be adopted as teacher model and student model. During the training process, the training data will be fed into these 2 models at the same time. The output of the student model is compared with the annotated label using classification cost and with the teacher model output using consistency cost. After the weights of the student model have been updated with gradient descent, the teacher model weights will be updated as an exponential moving average of the student model weights. Both the model outputs can be used for prediction, but the teacher model predictions are more likely to be correct. If the input data is an unannotated data, the training process is similar, but no classification cost would be applied.

2.4 Full Pipeline

Our full pipeline consists of 4 steps: first full-supervised training process, semi-supervised training process, prediction and fusion, second training process. All steps will be illustrated and they are essential to our performance improvement. In order to verify the performance improvement is benefit from our special training pipeline only, all hyper parameters used in our experiments will not be changed and keep same as the baseline models.

2.4.1 First Full-Supervised Training Process

As mentioned above, we have 2 partially annotated datasets: A dataset and B dataset. In A dataset, vision related organs have been annotated at pixel level; while in B dataset, auditory related organs have been annotated at pixel level. At this step, we will train 2 models with these 2 datasets using same basic architecture. DeepLabv3+ is adopted for its SOTA segmentation performance and U-Net is chosen for its wide usage. At the end of this step, there are 2 models: A model and B model. A model is trained with A dataset and it can be used to predict vision related organs segmentation. B model is trained with B dataset and it can be used to predict auditory related organs segmentation.

2.4.2 Semi-supervised Training Process

It is known to us that a model will learn to fit the training dataset, but it is impossible to achieve the perfect generalized performance because the training dataset is impossible to include all kinds of data. That means the quality of predicted annotations is not as good as original annotations from professional doctors. In our pipeline, the predicted annotations will be treated as real annotations in the following steps, so the performance will be better if the quality of the predicted annotations is better. In order to achieve better predictions, Mean Teachers method is adopted to fine-tune 2 models in semi-supervised paradigm. For example, B dataset will be treated as the unannotated data when training A model with A dataset to improve its performance, and A dataset will be the unannotated data when training B model with B dataset.

2.4.3 Prediction and Fusion

After owing 2 well-trained models for 2 partially annotated datasets, there are 2 steps for us to get full annotations. The first step is making predictions. During this step, B dataset is treated as the input of A model and A dataset is treated as the input of B model. Intuitively, the predictions are the lacking annotations of each dataset. The next step is to make fusion of predicted annotations and original annotations. In our experiment, as for the background class, only the pixel with predicted background annotations and original background annotations will be treated as the true background pixel; as for the foreground class, if a pixel belongs to a class of predicted annotations and another class of original annotations, then this pixel belongs to the class of original annotations. After making prediction and fusion, we will get 2 fully annotated datasets. By considering them as a whole fully annotated dataset, we could train a neural network for 2 segmentation tasks using full-supervised training method.

2.4.4   Second Training Process

Until now, we have a big dataset with full annotations. The simplest way is to train a model in full supervised paradigm. During this step, in order to execute control variate method, DeepLabv3+ or U-Net is still our backend model, and hyper parameters used in experiments will not change. However, it is not a proper way to treat the predicted annotations as real annotations for the poor quality of the predicted annotations. To address this problem, with the Mean Teachers algorithm, the optimization objective is the model outputs will try their best to close to the real annotations, while the model outputs of the predicted annotations will not change greatly comparing to the last epoch.

Honestly, the model trained in our pipeline should achieve better results for 2 reasons: the first reason is the larger dataset which consists 2 original datasets; the second one is not only the encoder parts are shared, but also the decoder parts, which means the correlation between these 2 similar segmentation tasks will be utilized.

2.5   Loss Function

Because our full training pipeline consist of full-supervised learning process and semi-supervised learning process, some details about the supervised loss function and unsupervised loss function should be listed. In our baseline experiment setting, cross entropy loss function and dice loss function will be used in full-supervised training process. As for our proposed pipeline, in order to make sure the performance improvement is benefit from our specially designed training pipeline, the full-supervised loss functions are still cross entropy loss function and dice loss function [5], while our unsupervised loss function used in semi-supervised training process is mean square error (MSE) loss function.

3   Experiment

In this section, we will show the experimental details about the whole training pipeline. Generally speaking, the hyper parameters used in the baseline single-task models and our proposed pipeline models must be same to make sure the performance improvement comes from our specially designed training pipeline.

3.1   Data

The dataset comes from MICCAI (the International Conference on Medical Image Computing and Computer Assisted Intervention) 2019 challenge: Automatic Structure Segmentation for Radiotherapy Planning Challenge. 22 OARs (Organs at Risk), including vision related organs and auditory related organs, of 50 nasopharynx patients will be annotated and released to public as training data. Each of the annotated CT scan is marked by one experienced oncologist and verified by another experienced one. Another 10 patients' CT scans will be used as the test data.

3.2   Preprocess

In order to verify our idea, we have divided the dataset into 2 smaller datasets: one dataset with vision related organs annotations only and the other with auditory related organs annotations only. Because different patient data may be generated from different type of devices or at different time, it is necessary to make a normalization of the input data at the first step. What's more, this paper aims at proposing a new training pipeline. It is not necessary to do data augmentation or other special tricks (Fig. 1).
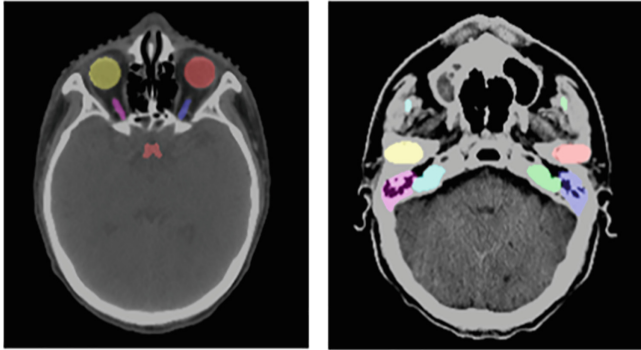

Fig. 1. The 2 datasets in our experiment. Left figure is from the dataset with vision related organs annotations while the right one is from the dataset with auditory related organs annotations.

3.3   Detail

Our proposed training process is implemented with PyTorch-1.0 and trained with an NVIDIA GeForce GTX 2080 GPU. In our experimental settings, DeepLabv3+ is chosen as basic architecture. During training process, different loss functions combinations are used in different steps. While same evaluation metric, Dice, will be used during evaluating process.

3.3.1   First Full-Supervised Training Process

For A dataset, we need to train a model for 8 vision related organs segmentation. DeepLabv3+ is chosen as the basic architecture and the number of the last output neurons is set as 9 (including 8 organs and background class). During our experiment, learning rate is 3e−4, Adam optimizer [11] is adopted, batch size is 2, the number of training dataset is 40734, loss function is cross entropy loss and dice loss.

For B dataset, we need to train a model for 10 auditory related organs segmentation. DeepLabv3+ is chosen as the basic architecture and the number of the last output neurons is set as 11 (including 10 organs and background class). During our experiment, learning rate is 3e−4, Adam optimizer is adopted, batch size is 2, the number of training dataset is 31711, loss function is cross entropy loss and dice loss.

3.3.2   Second Full-Supervised Training Process

DeepLabv3+ is still the basic architecture and the number of the last output neurons is set as 19 (including 18 organs and background class). During our experiment, learning rate is $3e-4$, Adam optimizer is adopted, batch size is 2, the number of training dataset is 72445, loss function is cross entropy loss and dice loss.

3.3.3   Semi-supervised Training Process

Initially, the student model and teacher model are all the previous trained model. The supervised loss is cross entropy loss and dice loss for annotated data (first training process) or original annotations (second training process); the unsupervised loss is MSE loss for unannotated data (first training process) or predicted annotations (second training process). And their proportion is 1:1. Afterwards, the student model is updated from gradient descent strategy while the teacher model is the exponential moving average of student models.

3.4   Result

In this part, we will show our experimental results to illustrate that our proposed pipeline is working well. At first, we will show some figures about segmentation results. Then, in order to prove that our proposed ideas are suitable in this case, we will list the results of each step.

3.4.1   Figure

The following figures are the experimental results for 2 different tasks. In these 2 figures, the left one is the original annotations from professional doctors and the right one is the output predictions from our models. From these 2 figures, we can easily arrive that our proposed model is working well. What's more, these 2 segmentation tasks are implemented in one single model (Figs. 2 and 3).
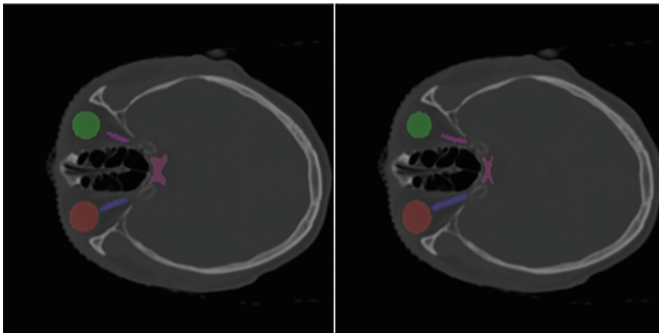

Fig. 2. The original annotations and output predictions of A dataset. The left one is the original annotations while the right one is the output predictions.
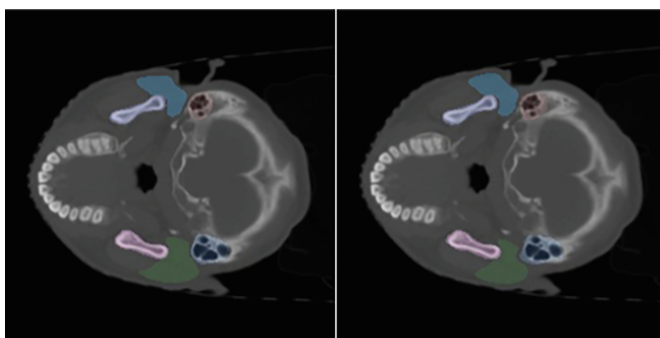
Fig. 3. The original annotations and output predictions of B dataset. The left one is the original annotations while the right one is the output predictions.

3.4.2 Table

See Table 1.

Table 1. Experimental results of DeepLabv3+

	A dataset	B dataset
First full-supervised training	71.68	84.60
Semi-supervised training	72.36	84.97
Second training	72.66	85.16

For the above table, the evaluate metric is dice for these tasks are segmentation tasks and dice is the most commonly used metric. In the table, the results of the first full-supervised training process are also the baseline results. After the semi-supervised training process, we can see that 2 models' performance has been improved: for A dataset, the dice improvement is 0.68 and the dice improvement for B dataset is 0.37. With a large dataset with all annotations, the second full-supervised training process is complemented, and we can get our proposed neural network. Obviously, the performance has been improved again: the dice improvement for A dataset is 0.3 and it is 0.19 for B dataset. Compared to the baseline methods, our proposed pipeline can achieve a great dice improvement: 0.98 for A dataset and 0.56 for B dataset. What's more, our proposed single neural network will predict these 2 kinds of organs at the same time. With the less parameters, our proposed neural network will have a shorter inference time. Additionally, with the great performance improvement, we believe our proposed pipeline should be applied in reality.

3.4.3 Robust

From the above table and figure, it is clear that our proposed method is robust to model selection. For the U-Net basic architecture, semi-supervised training process and second

full-supervised training process are useful for A dataset while second full-supervised training process is not good for B dataset. However, compared to the single-task baselines, the results show that our proposed methods are useful, and it is robust to model selection. That means our proposed method is able to be spread to other regions (Table 2).

Table 2. Experimental results of U-Net

	A dataset	B dataset
First full-supervised training	71.26	83.91
Semi-supervised training	71.60	84.39
Second training	71.79	84.39

4   Conclusion

In this paper, we have proposed a pipeline for training a neural network with various only partially annotated datasets. This idea can solve the problem of the scarce source of professionally annotated medical image data to some extent. What's more, intuitively, sharing encoder module and decoder module is benefit for taking the correlation between different segmentation tasks into consideration to improve performance. An idea generated from co-operative training is used to complement the lacking annotations of each dataset. For semantic segmentation, U-Net and DeepLabv3+ are the basic architectures in our experiments. However, the quality of predicted annotations is not as good as original annotations from professional doctors. To address this problem, Mean Teachers is the semi-supervised training method we adopted to improve the quality of predicted annotations.

Our experimental results have shown that with our proposed pipeline, it is easy for us to train a single neural network for multiple segmentation tasks at the same with only partially annotated datasets. Our contributions are three folds: First, with complementing several semantic segmentation tasks within only one neural network instead of several neural networks, the number of parameters in this CAD system is much less, so it is faster during inference process and suitable for clinical application for its less computational resource. Second, with this special training pipeline, people can train a neural network for all organs' segmentation with only partially annotated datasets. This is the good news for medical AI development because it can solve the problem that medical datasets are too hard to achieve. What's more, the proposed neural network can achieve a better performance than baseline models because it makes use of correlation between different tasks during the whole training process. As a future work, we are planning to try to complement more segmentation tasks in a neural network because our proposed training pipeline should be able to train a better neural network taking more correlation into consideration.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer, Cham (2015)
2. Chen, L.C., Zhu, Y., Papandreou, G., et al.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
3. Tarvainen, A., Valpola, H.: Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems, pp. 1195–1204 (2017)
4. Li, J., Yang, S., Huang, X., et al.: Signet ring cell detection with a semi-supervised learning framework. In: International Conference on Information Processing in Medical Imaging, pp. 842–854. Springer, Cham (2019)
5. Sudre, C.H., Li, W., Vercauteren, T., et al.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 240–248. Springer, Cham (2017)
6. Rajchl, M., Pawlowski, N., Rueckert, D., et al.: Neuronet: fast and robust reproduction of multiple brain image segmentation pipelines. arXiv preprint arXiv:1806.04224 (2018)
7. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training (1977)
8. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
9. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
10. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional networks. In: Proceedings of the Conference Neural Information Processing Systems (NIPS), pp. 1097–1105 (2012)
11. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

Application of U-Shaped Convolutional Neural Network Based on Attention Mechanism in Liver CT Image Segmentation

Chen Li[✉], Wei Chen, Xin Luo, Mingfei Wu, Xiaogang Jia, Yusong Tan, and Zhiying Wang

The School of Computer Science, National University of Defense Technology, Changsha 410073, China
lichen14@nudt.edu.cn

Abstract. Automatic segmentation of livers from medical images is a difficult task because of liver's uneven shapes. Recently, deep convolutional neural networks have become popular in this field because the deep learning method can learn hierarchical semantic information. In this paper, we propose to apply a deep learning model Attention U-Net to liver medical image segmentation. The proposed model uses classic UNet as basic architecture and combines the attention mechanism. Attention U-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task. Besides, Attention U-Net is an end-to-end image-based approach for liver segmentation and does not rely on any pre-trained models or common post-processing techniques. As a consequence, there is no need for the clipping of the region of interest (ROI) and the locating of target object in the network. The proposed model is evaluated on the public dataset from ISBI2019 Challenge for CT liver segmentation. Experiments demonstrate that Attention U-Net achieves an IoU gain of 5.15%, a Dice gain of 3.51%, a accuracy gain of 5.78% and a recall gain of 4.81% over UNet.

Keywords: Attention · UNet · Liver segmentation · Deep learning

1 Introduction

Liver cancer is the common internal malignancies in the world and the leading causes of cancer death, which pose a huge threat to human health. Identifying the position of liver from medical images is a preparation step in diagnosing liver lesions and plays an indispensable role in disease treatment. However, segmentation of liver from medical images is very challenging because of the liver's uneven shape and the low contrast between the tissues in human body.

Researches on solutions to liver segmentation have been proposed. Generally, liver segmentation methods can categorized in to 2 classes: (1) manual and semi-automatic segmentation, (2) automatic segmentation. Manual segmentation largely relies on experts with advanced technical skills to perform such tasks. As a consequence, the quality of

the segmentation relies heavily on the judgment of experts, which is time-consuming and not repeatable. These factors make manual segmentation impractical. Meanwhile, semi-automatic segmentation still requires manual intervention, which can lead to biases and errors. Because densely labeling massive medical images by manual or semi-manual methods is a tedious and high error rate task, extensive research has been conducted on automatic medical image segmentation. With the advent of convolutional neural network (CNN) and its rapid application in medical image segmentation, it has been used in many automatic medical image analysis tasks such as cardiac MRI image segmentation [1] and cancerous pulmonary nodule segmentation [2]. However, this CNN-based method usually divides the segmentation task into two steps: localization and segmentation. The extra localization step will increase the amount of model parameters and bring extra time consumption. The accuracy of model segmentation also depends heavily on the first step positioning accuracy.

In order to accelerate and facilitate diagnosis, and assist doctors in removing diseased areas such as tumors, it is necessary to develop a reliable automated solution to accurately segment organs from medical images. Inspired by attention mechanism, we use an attention-based U-shaped convolutional neural network called Attention U-Net for liver segmentation, designed to efficiently outline the liver from CT images without localization step. Attention U-Net integrates the UNet encoder-decoder architecture and Attention Gate learning mechanism, which can accelerate inference process and improve segmentation performance. The contributions of our work are summarized as follows:

- Attention U-Net is introduced for liver CT image segmentation.
- The experiments are conducted on public dataset show that Attention Gate has the ability to focus on specific parts of the whole image.
- Attention U-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task.
- Attention U-Net has superior performance on liver segmentation task compared to equivalent UNet model.

The structure of this paper is as follows: Sect. 2 will briefly review related segmentation methods. Section 3 will describe in detail our segmentation framework. Then we will introduce the settings of the experiment in Sect. 4. After that, the experimental results will be displayed and analyzed. Finally, the conclusions are given in Sect. 5.

2   Related Works

2.1   U-Shaped Convolutional Neural Networks

The state-of-art image segmentation models are variants of encoder-decoder architectures such as UNet [3] and Fully Convolutional Network (FCN) [4]. These models have a common similarity: skip connection. Skip connection combines the abstract feature information restored from the decoder and the surface feature information extracted from the encoder in the channel dimension. In complex segmentation tasks, skip connection can effectively retain the feature information of each level and generate segmentation results with low-level details.

Due to the difficulty in collecting medical image data, it is impossible to obtain sufficient mass training data in biomedical image analysis tasks. But since UNet [3] proposed in 2015, it has been widely used in the field of image segmentation due to its simple but effective structural characteristics. The network structure is shown in Fig. 1 below. UNet is composed of two parts, the concatenation path on the left and the expansive path on the right. It uses a U-shaped frame of a fully convolutional neural network to make the expansion path and the contraction path approximately symmetrical. UNet can get the feature context information from each level in the contraction path and highlight the position of the foreground target object. The restored feature information is obtained in the expansion path, and then the features extracted on multiple dimensions are merged through skip connections, combined shallow features and deep features.
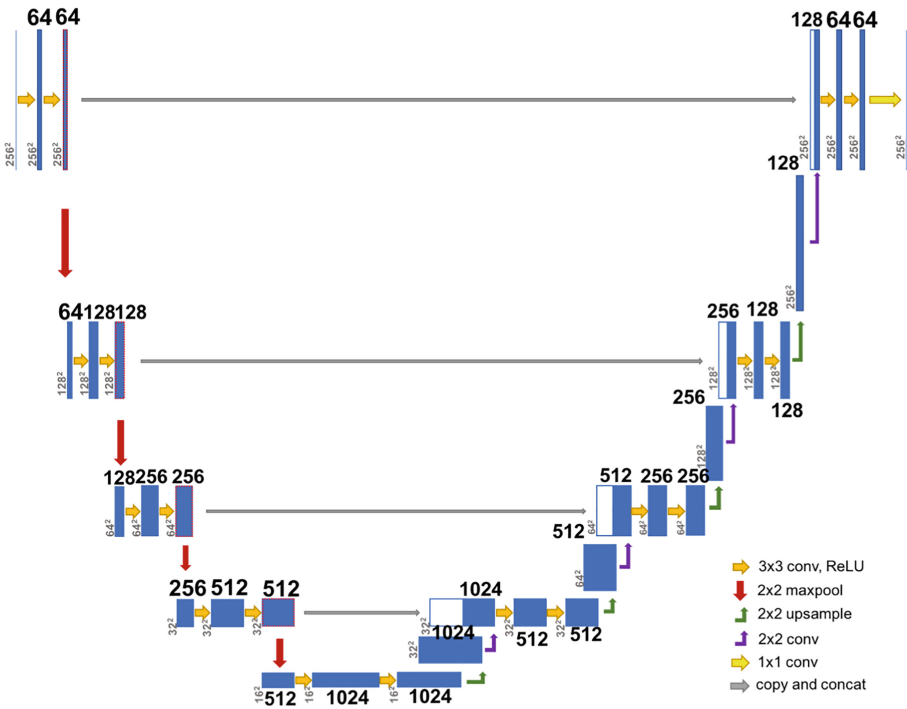

Fig. 1. UNet architecture

2.2   Attention Mechanism

Most current segmentation frameworks simplify the segmentation task into two steps: object position and segmentation. Although UNet is well-represented and widely-applied in medical image segmentation such as cardiac MRI [5] abdominal CT [6] segmentation, pulmonary CT nodule detection [2] and previous UNet-based segmentation networks still rely on multi-level cascaded CNNs to extracts the region of interest (ROI), especially when target objects show large patient-to-patient differences such as shape and

size. However, this multi-level cascaded CNN method largely relies on previous object positioning accuracy and requires more model parameters and higher computing power. In order to solve this problem, we refer to the method proposed by Oktay [7] to add a simple but effective Attention Gate mechanism to the network architecture.

Attention mechanism firstly emerged in the field of natural language processing (NLP) and quickly gained dominance. It was Nonlocal [8] proposed by He Kaiming's team that first introduced the attention mechanism to the field of computer vision. Since then, a series of research booms have focused on attention mechanisms. The attention mechanism was firstly introduced in the semantic segmentation in [9], which combined the shared network with the attention mechanism and achieved superior performance. Later in 2018, many high-impact articles were published such as DANet [10], OCNet [11], CCNet [12], and the Local Relation Net [13] in 2019. Recently, attention mechanism has gradually been applied to medical image segmentation.

Inspired by UNet [3], Oktay [9] proposed a model combining attention mechanism and UNet. The attention mechanism focuses the network learning attention on distinguishing the relevant positions of the object through the Attention Gate. The inputs of the Attention Gate are the upsampling feature in the expansion path and the corresponding feature in the contraction path. The former is used as a gating signal to enhance the learning of the task area of interest, and suppress the tissue area unrelated to the segmentation task. Therefore, the Attention Gate can increase the effectiveness of skip connections in return. Next, the S-shaped activation function sigmoid is selected to train the convergence of parameters in the Attention Gate. The output is the multiplication result of the input feature and the attention coefficient α. Another advantage of Attention Gate is that it can use standard back-propagation to update the training parameters in the gate without using the sample-based update method in [14].

3   Methodology

The proposed Attention U-Net uses UNet as the basic network framework, and an encoder and a decoder are symmetrically arranged on both sides of the network. The structure parameter configuration of Attention U-Net is listed in Table 1.

Encoder is a typical framework of a convolutional neural network and consists of downsampling modules. Each downsampling module firstly consists of two consecutive 3×3 convolutional layers. The convolutional layer processes the local information layer by layer to obtain the image feature x_l^c of the c-th channel extracted by the l-th layer. After each convolution layer, a rectified linear function (ReLU) is used as the activation function $\sigma(x_l^c) = max(0, x_l^c)$. The end of the downsampling module is a 2×2 max pooling for downsampling and doubling the number of feature channels. In summary, the feature information x_l^c can be expressed as:

$$x_l^c = \sigma(\sum_{c' \in F_l} x_{l-1}^{c'} * K^{c',c}),$$

Where c represents the channel, F_l represents the number of convolution kernels of the l-th layer, $K^{c',c}$ represents the c'-th convolution kernel, * represents the convolution operation, and σ represents the activation function ReLU.

Table 1. Attention U-Net network structure parameter configuration

Encoder	Output size	Decoder	Skip connection	Output size
Input	256^2 * 1	Up1		32^2 * 512
Conv1	256^2 * 64	Attention Gate1	[Up1, Conv4]	32^2 * 1024
Pooling	128^2 * 64	Conv6		32^2 * 512
Conv2	128^2 * 128	Up2		64^2 * 256
Pooling	64^2 * 128	Attention Gate2	[Up2, Conv3]	64^2 * 512
Conv3	64^2 * 256	Conv7		64^2 * 256
Pooling	32^2 * 256	Up3		128^2 * 128
Conv4	32^2 * 512	Attention Gate3	[Up3, Conv2]	128^2 * 256
Pooling	16^2 * 512	Conv8		128^2 * 128
Conv5	16^2 * 1024	Up4		256^2 * 64
		Attention Gate4	[Up4, Conv1]	256^2 * 128
		Conv9		256^2 * 64
		Conv10		256^2 * 1

Decoder consists of upsampling modules that upsamples the features to restore image information. Each upsampling module firstly uses 2×2 deconvolution to halve the feature size and the number of channels, then connects the compressed features after concatenation, and finally performs 3×3 convolutions twice. In the last layer, a 1×1 convolution is used to map each feature vector to the required number of classes.
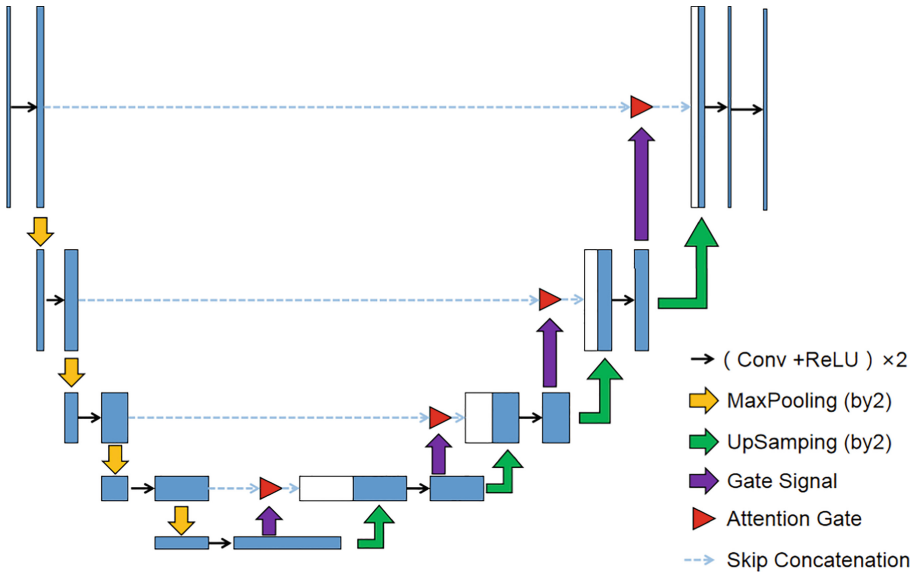

Fig. 2. Attention U-Net architecture

The structure of the Attention U-Net network is shown in Fig. 2. Attention U-Net is designed for liver segmentation and does not rely on any pre-trained models or common post-processing techniques. As a consequence, there is no need for the clipping of the region of interest (ROI) and the locating of target object at the beginning. Compared with the traditional classic UNet, Attention U-Net optimizes the previous Conv block from "Conv 3 × 3, ReLU" to "Conv 3 × 3, Batch Normalization, ReLU". We want to accelerate the model's convergence speed and alleviate the vanishing gradient problem in the deep network to a certain extent through this optimization. As a consequence, it makes training the deep network model easier and more stable.

4 Experiments and Results

4.1 Experimental Setup

In this paper, we use the public dataset CHAOS from the 2019 IEEE International Symposium on Biomedical Imaging (ISBI) to train and test the network. This database contains DICOM images of 20 different patients with a resolution of 512 × 512. CT images were obtained from the patient's upper abdomen after injection of the contrast agent. The data were collected by three instruments: Philips SecuraCT with 16 detectors, Philips Mx8000 CT with 64 detectors, and Toshiba AquilionOne with 320 detectors. All three instruments are equipped with the spiral CT option. The xy pitch is between 0.7–0.8 mm. Images in the CHAOS dataset is shown in Fig. 3.
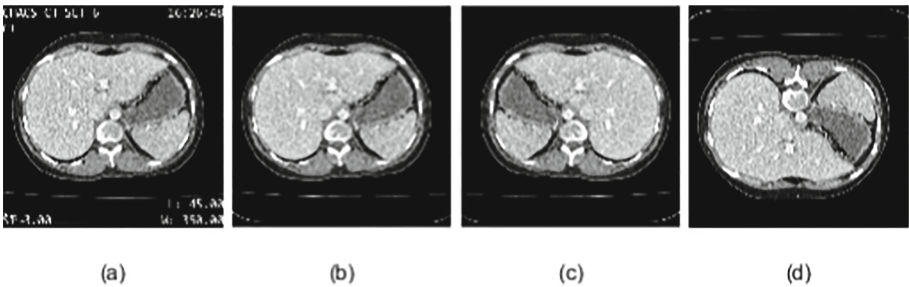

(a)   (b)   (c)   (d)

Fig. 3. Example image from the CHAOS dataset

The average number of slices in per patient sub-dataset is 90, the minimum number of slices is 79, and the maximum number of slices is 264. The total 2342 images are randomly divided into a train set and a test set. The train set includes 2303 slices and the test set includes 39 slices.

The experiment is based on 64-bit Ubuntu operating system and NVIDIA GTX 1080Ti GPU, and the deep learning framework used is Pytorch. The entire network uses Adam's method to optimize parameters, using a varying learning rate, which is initially

0.03 and decreases with epoch. Dice similarity coefficient is selected as loss function for back propagation of errors, which is defined as follows:

$$\mathrm{Dice}(\mathrm{Y}, \overline{\mathrm{Y}}) = \frac{2\mathrm{Y}\overline{\mathrm{Y}}}{\mathrm{Y} + \overline{\mathrm{Y}}},$$

Where $\overline{\mathrm{Y}}$ is the real result and Y is the segmentation result.

4.2   Experimental Results and Analysis

We use the Attention U-Net model to segment CT images of the liver to verify the effectiveness of the proposed method. Dice similarity coefficient, IoU coefficient, precision, recall are used as performance indicators to evaluate the performance of liver CT image segmentation. The larger the values of these four indicators, the larger the overlapping area between the segmentation result and the real result, the higher the similarity, and the greater the accuracy of the segmentation. The results are shown in the following Table 2.

Table 2. Segmentation results of methods on the test dataset

Network	IoU	Dice	Precision	Recall
UNet [3]	0.8350	0.8697	0.8699	0.8822
Attention U-Net	**0.8780**	**0.9002**	**0.9202**	**0.9246**

It is known from Table 2 that under the evaluation of four indicators, the U-shaped convolutional neural network after introducing the attention mechanism has greatly improved the liver CT image segmentation performance, of which the IoU ratio increased by 5.15%, the Dice coefficient increased by 3.51%, and the accuracy increased 5.78%, recall rate increased 4.81%.

By comparing the segmentation results of our model and UNet in Fig. 4, we can intuitively conclude that the Attention U-Net can better restore the characteristic structure of the liver, and more fully capture the detailed information that U-Net ignores.

By comparing the differences between the segmentation results and the actual segmentation results, we can conclude that our experimental results proved that the attention mechanism can effectively solve the problem of cascaded CNN method in Sect. 2.1. The main innovations in Attention U-Net are: Attention Gate mechanism is added and the features extracted at different levels can be merged with a focused selection. The features extracted by the encoder is propagated to the decoder through the Attention Gate before skip connection, so that more targeted hierarchical context information can be extracted; after receiving the features, the decoder restores the features in a bottom-up manner. In the expansion path, the accuracy and stability of the network are enhanced.
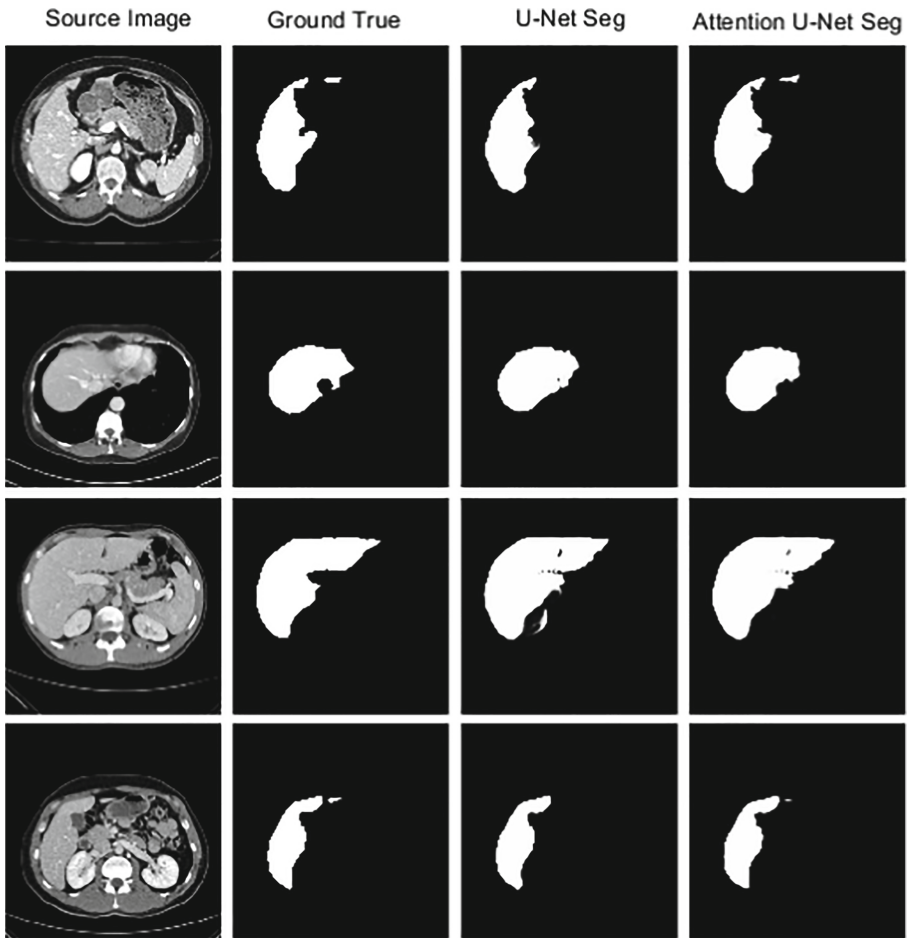
Fig. 4. Liver segmentation results of UNet model and attention U-Net model and true segmentation results

5   Conclusions

In this paper, a U-shaped convolutional neural network based on attention mechanism is proposed and applied to liver CT image segmentation. Our experimental results proved that Attention U-Net enhances the learning of task regions and increases the effectiveness of skip connections. Attention Gate merges the features selected by different levels in the encoder with a task-related selection into the decoder. Attention U-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task. As a consequence, there is no need to trim the ROI and locate the target object in our model. In addition, the method is generalized and modular in the field of medical image analysis, so it can also be applied to image classification and regression problems. We conclude that introduced Attention mechanism can enhance

the accuracy and scalability of network, and Attention U-Net used in this paper can accurately segment human tissues or organs.

References

1. Tarroni, G., Bai, W., Sinclair, M.: Human-level CMR image analysis with deep fully convolutional networks, CoRR, vol. abs/1710.09289 (2017)
2. Liao, F., Liang, M., Li, Z., Hu, X., Song, S.: Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network. IEEE Trans. Neural Netw. Learn. Syst. **30**(11), 3484–3495 (2017)
3. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation (2015)
4. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2014)
5. Khened, M., Kollerathu, V.A., Krishnamurthi, G.: Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Med. Image Anal. **51**, 21–45 (2019)
6. Roth, H.R., Lu, L., Lay, N., Harrison, A.P., Farag, A., Sohn, A., Summers, R.M.: Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation, CoRR, vol. abs/1702.00045 (2017)
7. Oktay, O., Schlemper, J., Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N., Kainz, B., Glocker, B., Rueckert, D.: Attention u-net: learning where to look for the pancreas, April 2018
8. Wang, X., Girshick, R., Gupta, A., et al.: Non-local neural networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2018)
9. Schlemper, J., Oktay, O., Chen, L., Matthew, J., Knight, C., Kainz, B., Glocker, B., Rueckert, D.: Attention-gated networks for improving ultrasound scan plane detection, CoRR, vol. abs/1804.05338 (2018)
10. Fu, J., Liu, J., Tian, H., Fang, Z., Lu, H.: Dual attention network for scene segmentation (2019)
11. Yuan, Y., Wang, J.: Ocnet: object context network for scene parsing (2018)
12. Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W.: CCNet: criss-cross attention for semantic segmentation (2019)
13. Hu, H., Zhang, Z., Xie, Z., Lin, S.: Local relation networks for image recognition (2019)
14. Mnih, V., Heess, N., Graves, A., Kavukcuoglu, K.: Recurrent models of visual attention, CoRR, vol. abs/1406.6247 (2014)

Design of Photovoltaic Power Intelligent Patrol Robot

Na Yao[✉], Xiaofang Zhao, and Huazhu Liu

Dongguan University of Technology, Dongguan 523808, Guangdong, China
yaon@dgut.edu.cn

Abstract. This paper proposes a control strategy for single-phase off-grid inverter, which integrates the three closed-loop control with the iterative-based RMS algorithm. The inverter circuit is modeled, and simulation experiment and prototype verification are performed on Matlab. The results show that the voltage drop problem caused by an abrupt variation of inverter load is improved greatly owing to the addition of voltage RMS loop. Meanwhile, the application of iterative method enhances the dynamic response performance of the system substantially; and improves the real-timeliness of three closed-loop control. The two complement each other to provide a highly effective, reliable control solution for the single-phase off-grid inverter.

Keywords: RT-Thread · Wi-Fi · Mongoose · Video transmission

1 Information

Traditional robot remote video transmission control technology is often limited to the existing communication network. Due to the particularity of the production and living environment, high quality and high density of wireless communication network without dead space is often unable to be set up and deployed, which severely limits the range of mobile robot activities. Traditional robots usually only use batteries as the sole energy source. Because of the limited capacity of batteries, mobile robots generally have lower endurance. Therefore, the application of video transmission control system of mobile robots is limited to some extent, which cannot meet the requirements of modern times. With the development of photovoltaic charging technology, web technology and Wi-Fi wireless network, mobile robot wireless video transmission control system based on photovoltaic power and cross-platform Web technology has broadened its application range, and improved the endurance of outdoor mobile robot and real-time effect and convenience of the robot remote video monitoring [1].

2 System Hardware Design

The hardware part of the system mainly consists of photovoltaic power circuit, video transmission control circuit, inclination adjustment device circuit and structure, etc. The block diagram of system hardware design is shown in Fig. 1.
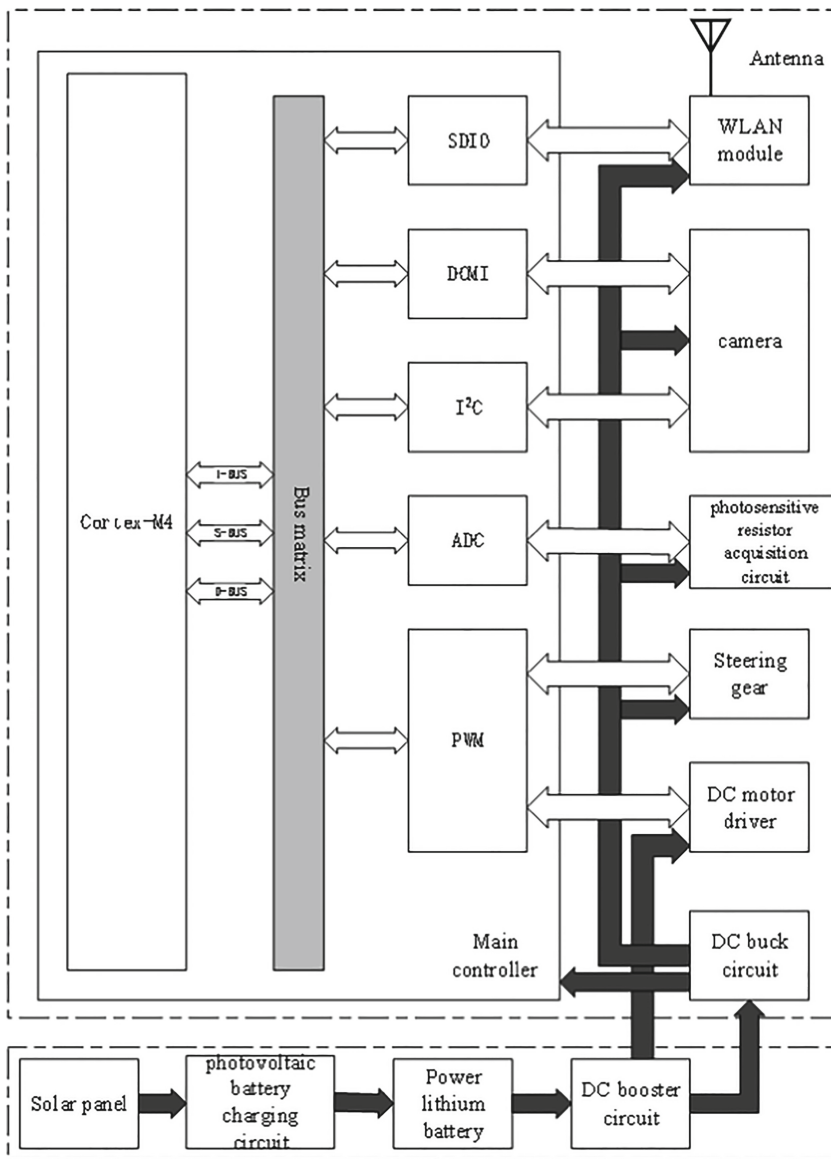
Fig. 1. Block diagram of system hardware design

2.1 Photovoltaic Power Supply Circuit Design

Photovoltaic power supply circuit is mainly composed of solar cells, photovoltaic battery charging circuit, DC booster circuit, and DC buck circuit. Electric energy output by the solar battery is the only energy source of the system. The photovoltaic battery charging circuit realizes the maximum power point tracking and completes the charging management of 2 lithium batteries. The lithium battery outputs 12 V voltage through the

DC booster circuit to provide power for the DC motor driver and DC buck circuit. The DC buck circuit provides 5 V and 3.3 V voltage stabilizers for the system. Figure 2 is the schematic diagram of photovoltaic battery charging circuit. CN3722 solar cell special charge management chip is adopted in the design to realize charging control of lithium battery group. CN3722 is a PWM switching mode battery charger controller. It can be powered by solar cells and has the maximum power point tracking function. Moreover, it is used to charge one or more lithium batteries or lithium iron phosphate batteries for constant current and constant voltage mode. It has the advantages of small package size, less peripheral components and simple use [2].
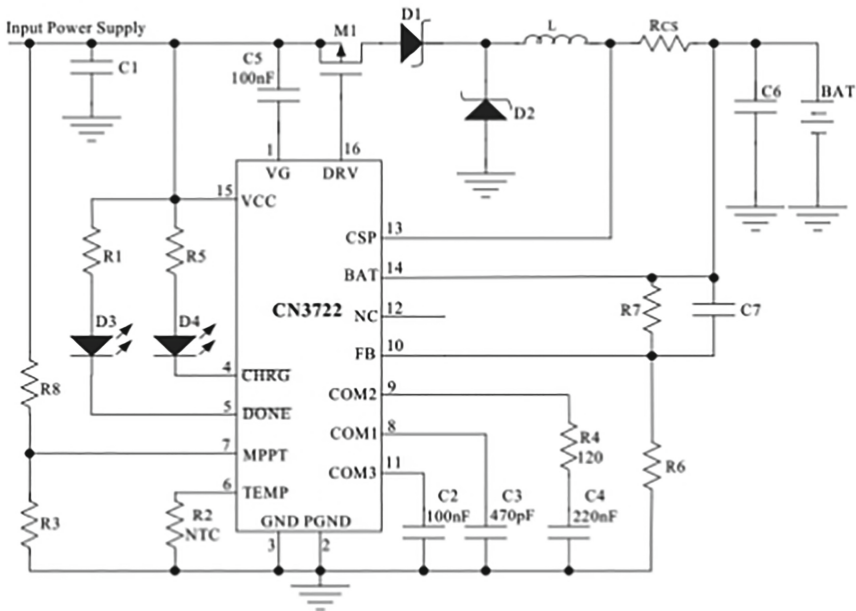
Fig. 2. Photovoltaic battery charging circuit

The DC motor used in the robot car works at the rated voltage of 12 V and the system load is heavy, so it has higher requirements on the power output. The system is powered by 2 power lithium batteries, and the DC booster circuit needs to be designed to meet the requirements of high output voltage and current. TPS61088 fully integrated synchronous boost converter is used in the design. The converter integrates 11 mΩ power switch and 13 mΩ switch rectifier. The buck circuit is mainly composed of buck switching voltage regulator LM2576 and low-voltage differential linear voltage regulator AMS1117, which provide the system with 5 V and 3.3 V DC voltage stabilizer outputs, respectively.

2.2 Video Transmission Control Circuit Design

Video transmission control circuit is mainly composed of main controller, camera, WLAN chip, DC motor and driver. WLAN chip is the physical transceiver of video

transmission control system; the main controller drives the WLAN chip through the SDIO interface to complete data communication. Besides, the main controller is responsible for buffering and packaging video data collected by the camera, and sending it to the remote host computer through the WLAN chip. The WLAN chip receives the control instructions from the remote host computer. The master controller processes the control instructions and controls the movement of the DC motor of the robot car and the rotation of the camera cradle head.

Camera Module Circuit. The main controller adopts STM32F427VIT6 of STmicroelectronics, which integrates a lot of peripheral resources, such as parallel digital camera interface, SDIO bus controller, Ethernet MAC controller, USB high-speed OTG interface, DMA controller, multichannel timer channel, etc. It is convenient to use and simplifies the hardware design. Based on the cost and the processing capacity of the main controller, homemade OV2640 camera module is selected in the design. The camera supports the direct output of video stream in JPEG format, and does not need the master controller to compress video data, so that the master controller has more sufficient resources to handle WLAN drivers and network-related tasks, thus ensuring smooth transmission of video and small delay [3]. Figure 3 is real camera and interface circuit.
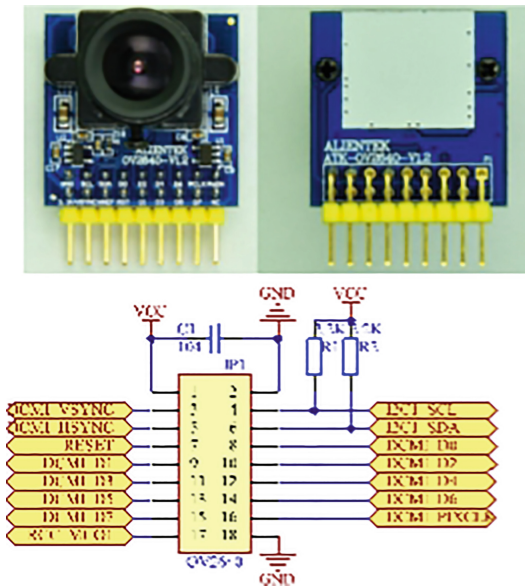

Fig. 3. Real camera and interface circuit

Wireless Local Area Network (WLAN) Module. Considering cost, stability and complexity of hardware design, BL-R8801MS1 wireless LAN module is selected for wireless video transmission system. BL-R8801MS1, as a domestic wireless module,

integrates Marvell's 88W8801 chip and its peripheral components, and it communicates with the processor through the SDIO interface. Figure 4 is the physical object and interface circuit of WLAN module.
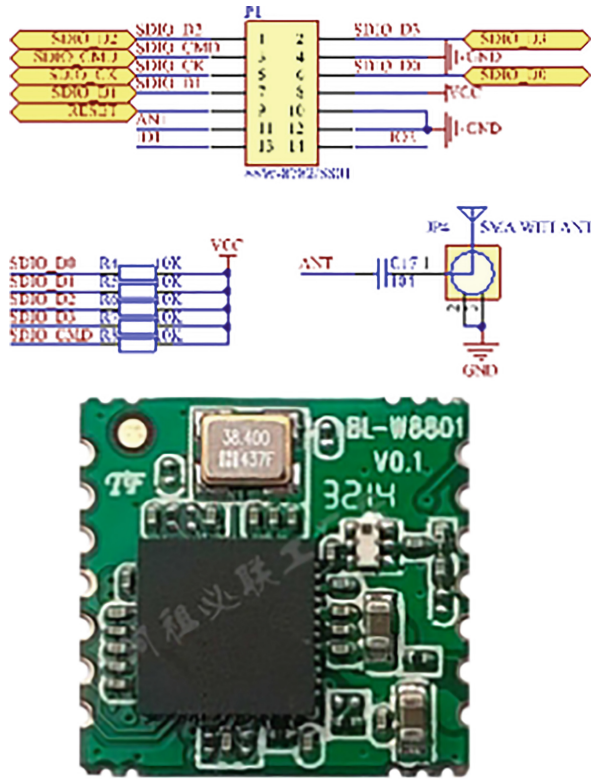

Fig. 4. Physical object and interface circuit of WLAN module

DC Motor Driver. In the design, TB6612FNG produced by Toshiba Semiconductor Company is used as the driver of DC motor of mobile robot car. It integrates two-way MOSFET-H bridge drive circuit to control two low-power DC motors, and has advantages in terms of output capacity, operation performance and heat dissipation [4].

2.3 Inclination Adjustment Circuit and Structural

The Angle adjustment device controls the Angle of solar panels installed with photosensitive resistor acquisition circuit above the robot car in the directions of X axis and Y axis. Small black in Fig. 5 represent the photosensitive resistance acquisition circuit. According to the resistance value of the photosensitive resistance, the single-chip microcomputer analyzes the distribution of solar panel light intensity, and controls the

steering gear at the bottom to drive the connecting rod device to push the solar panel to the appropriate Angle. Steering gear is selected as the actuator of the inclination adjustment device. The solar panel and the steering gear are connected by connecting rod device and bearing. The connecting rod device consists of an upper and lower arm. Its lower arm is connected with the steering engine, forming a component with linear stroke, as shown in Fig. 6. Besides, two bearings are installed at the bottom of the solar panel, which are parallel to the axis respectively. Connecting rod device and bearing are connected by self-made components, realizing independent control between the X axe and Y axe, which do not affect each other. The design drawing of bearing connection parts is shown in Fig. 7.
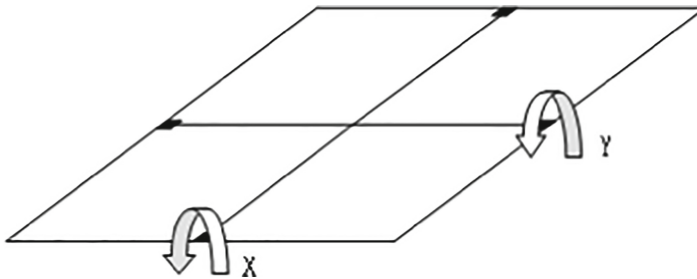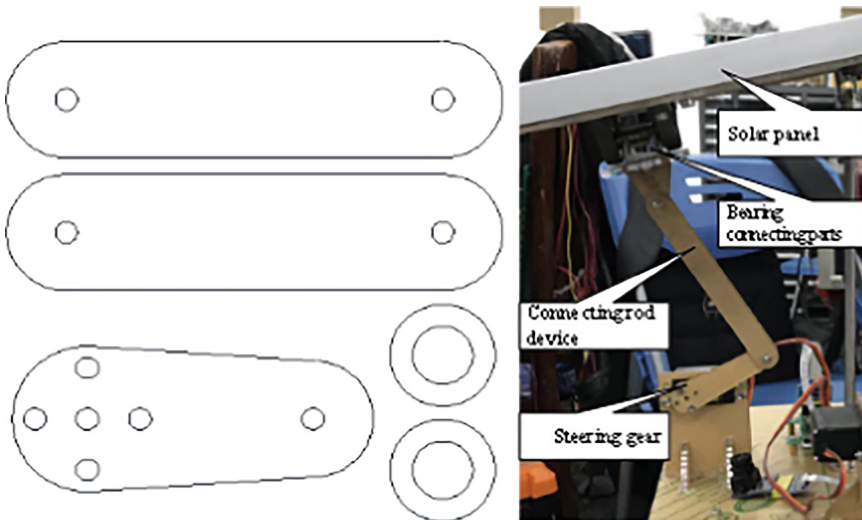

Fig. 5. Angle adjustment diagram of solar panel

Fig. 6. Physical object and design drawing of the connecting rod device
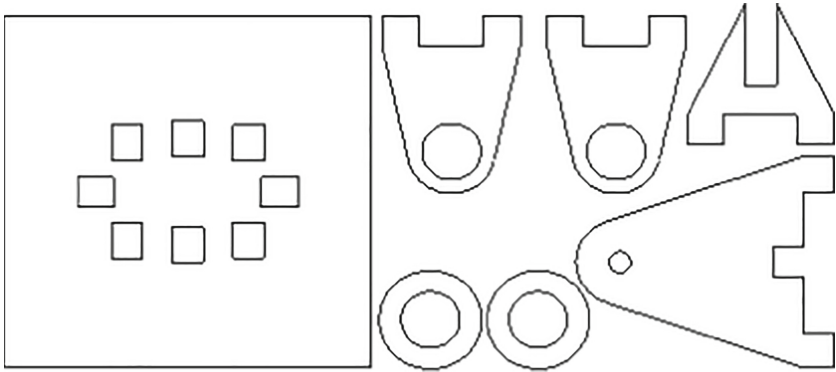
Fig. 7. Design drawing of bearing connection

3 Software Design

3.1 Wireless LAN Module Driver Transplantation

RT-Thread real-time operating system is adopted in the design. It has a good software ecology and supports all mainstream compilation tools in the market, all kinds of standard interfaces, and almost all mainstream MCU and Wi-Fi chips in the market. In the design, all network communication depends on WLAN chip to complete, and the successful transplantation of WLAN driver is the cornerstone of video transmission control system [5]. The transplantation can be roughly divided into Linux and RT-Thread communication API docking, SDIO host driver writing, network packet transceiver docking and network card registration. After the driver code is compiled on RT-Thread platform, the necessary code has to be written for WLAN operation and parameter configuration due to the lack of relevant configuration software based on nl80211 or WEXT.

3.2 Design of Video Transmission Server

The video transport server is one of the most important designs in the paper. It is an application program based on HTTP protocol and written with socket. As for its main function, the main controller drives the camera to realize the image acquisition function and sends the acquired JPEG images to the browser through TCP and HTTP protocols in real time.

All image processing functions of the designed camera, including gamma curve, white balance, contrast, chroma, etc., are programmed through the SCCB bus. Image collection and video transmission are processed in the same thread after the system starts running.

Image acquisition begins after the camera is turned on. The image acquisition process is shown in Fig. 8. HTTP is a standard for client - and server-side requests and replies. The video Server thread of the design USES the Server Push technology of HTTP protocol to send the JPEG images collected by the camera to the browser continuously.

According to the general TCP server programming process, the authors let socket binding in port 8088, enter the listening state and at the same time to process the maximum

of one client request; when the client connection read HTTP request, if the request content is "GET/?Action = stream", it indicates that the client is requesting JPEG video stream data, and the server should send an HTTP response. The server completes a push of JPEG image, and the thread will eventually return to the rt_completion_wait to wait for the completion of the next frame image acquisition, so continuously sending image data to the client realizes the function of network video transmission [6].
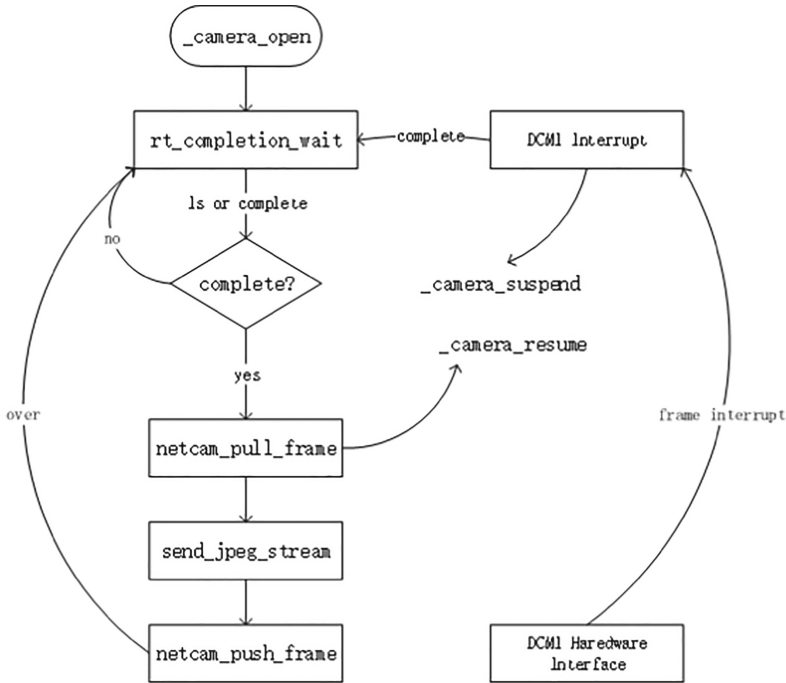

Fig. 8. Image acquisition process

3.3 Design of Control Lower Computer Program

The control lower computer is an embedded Web server written based on Mongoose network library, which USES WebSocket to communicate with the browser. It realizes the function of monitoring button events in the browser and controlling the robot car to move in all directions through JavaScript.

After the establishment of the Web server of the main controller, it is necessary to write JavaScript running on the browser side to listen for key events and run the motor control program running on the server side to control the movement of the robot car. The Mongoose event manager will generate the MG_EV_WebSOCKET_FRAME event and call back the ev_handler function registered in mg_bind when the server receives the WebSocket data frame sent by the browser [7].

In this event, the key information is read and the movement of the robot car is controlled. The core code is shown below, and the data parameter is the key information monitored.

```
static void control_handler(void *data, int size)
{
#define CAR_FORWARD  "U"          /* Key information: forward */
#define CAR_REVERSE     "D"            /* Key information: backward */
#define CAR_TURN_LF   "L"          /*Key information: turn left */
#define CAR_TURN_RT     "R"            /* Key information: turn right */

if (!strncmp(data, CAR_FORWARD, size)) {      /* Forward order */
dc_motor_reverse(motor_l, 40);     /* Control left motor reversal */
   dc_motor_forward(motor_r, 40);/* Control right motor reversal */
   rt_timer_start(timer);        /* Keep operation for a period of time */
else if (!strncmp(data, CAR_REVERSE, size)) {  /* Backward order */
  dc_motor_forward(motor_l, 40);/* Control forward rotation of left motor */
dc_motor_reverse(motor_r, 40);  /* Control backard rotation of left motor */
  rt_timer_start(timer);       /* Keep operation for a period of time */
   }
……
}
```

3.4   Design of Solar Panel Inclination Angle Adjustment Program

Inclination Angle adjustment is only an auxiliary means to improve the power generation efficiency. The design requires little accuracy of ADC voltage collection. The 12-bit ADC inside the single-chip microcomputer can be directly used for voltage collection. The program flow is shown in Fig. 9.

Voltage collection and PID control are processed in the same thread. Four channels of ADC are firstly configured for sampling in turn after the system starts to run, and then collection is started. The thread blocks and waits for ADC interrupt to send acquisition signal, and the collected data is transferred to the designated buffer through DMA.

Voltage collection stops after collecting voltage data for 10 consecutive times according to the above steps. Mean filter is conducted on the data of each channel to obtain new values. After subtracting the sampling values of 2 channels corresponding to each axis on the solar panel to obtain voltage differences, PID inclination control is carried out. A new round of voltage collection begins after system delay for a period of time.

The block diagram of solar energy inclination control system is shown in Fig. 10. Take the X axis as an example, the solar panel is equipped with two photosensitive resistance acquisition circuits on the X axis, and ADC needs to collect the two circuits of voltages. The design expects the voltage difference between them to be zero, indicating a uniform distribution of light along the X-axis. The difference between the ADC sampling values of the two circuits is regarded as the feedback system; current system error is obtained by comparing with zero value. After the PID controller algorithm, the corresponding pulse
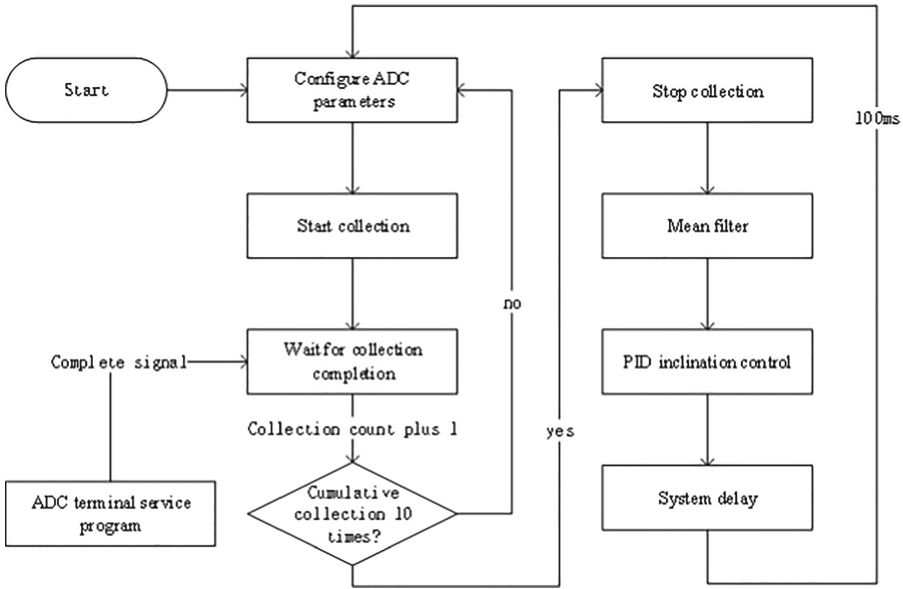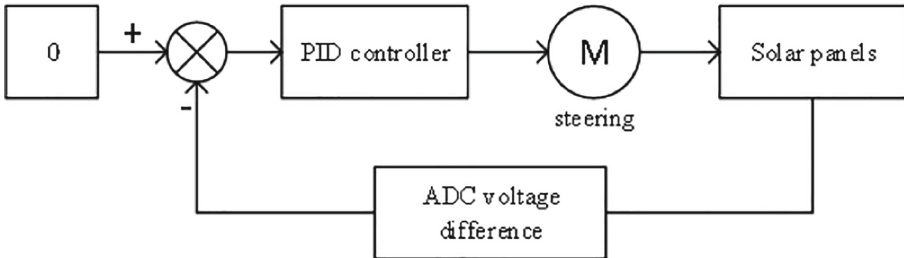
Fig. 9. ADC voltage collection process

Fig. 10. Block diagram of inclination angle control system

width is got; the actuators (steering) drive the connecting rod device is controlled and the inclination Angle of solar panels is adjusted so as to keep the solar panels towards the most sufficient illumination. Such way, to a certain extent, improves the efficiency of power generation [8].

4   System Test

4.1   Booster Circuit Conversion Efficiency

The DC digital power supply is used as the input source for the test, and the output drives the DC electronic load. The voltage and current at the input and output ends of the DC booster circuit are measured respectively. The conversion efficiency is calculated according to the test data, and the test results are shown in Table 1. As can be seen from

Table 1, when the input voltage remains unchanged (8.20 V), the conversion efficiency of light load (output current 0.094 A) can reach 91.46%, while that of heavy load (output current 1.810 A) can still reach 85.78%, and the maximum efficiency can reach 93.55%, meeting the requirements of energy conversion.

Table 1. Measurement of conversion efficiency of DC booster circuit (input voltage 8.20 V)

Input current (A)	Output current (A)	Output voltage (V)	Input power (W)	Output power (W)	Conversion efficiency (%)
0.150	0.094	11.967	1.230	1.125	91.46
0.770	0.497	11.885	6.314	5.907	93.55
1.230	0.801	11.746	10.086	9.409	93.29
1.560	1.004	11.725	12.792	11.772	92.03
1.910	1.206	11.698	15.662	14.108	90.08
2.250	1.409	11.658	18.450	16.426	89.03
2.600	1.607	11.608	21.320	18.654	87.50
2.970	1.810	11.542	24.354	20.891	85.78

4.2   Video Transmission Image Resolution and Frame Rate

PC and video robot are connected to UBNT UAP-AC-Pro wireless router to the same network, and the resolution of received image is viewed through developer tool of Web browser. Besides, the Times of image refresh within a period of time are counted and the image frame rate is calculated. Figure 11 shows the image transmitted by a video when the intelligent patrol robot is patrolling a strong electricity laboratory. The resolution of the image transmitted is 320 * 240 and the image size is 7.0 KB, which basically meets the design requirements.

The statistical data of image frame rate are shown in Table 2. The statistical frame rate is stable at around 29 frames, and the monitoring screen is basically smooth without jamming.

5   Summary

In the paper, a solution of remote video transmission control platform of mobile robot is proposed based on photovoltaic power; photovoltaic battery charging circuit is mainly designed. The software, hardware and mechanical structure of solar panel Angle adjustment device based on photosensitive resistance sensor are designed. WLAN drivers from Marvell are ported. As for STM32 platform, peripheral hardware circuits such as camera and WLAN module are designed, and corresponding underlying drivers are written. Moreover, the video transmission server and control program based on lwIP protocol stack and Mongoose network library are designed.
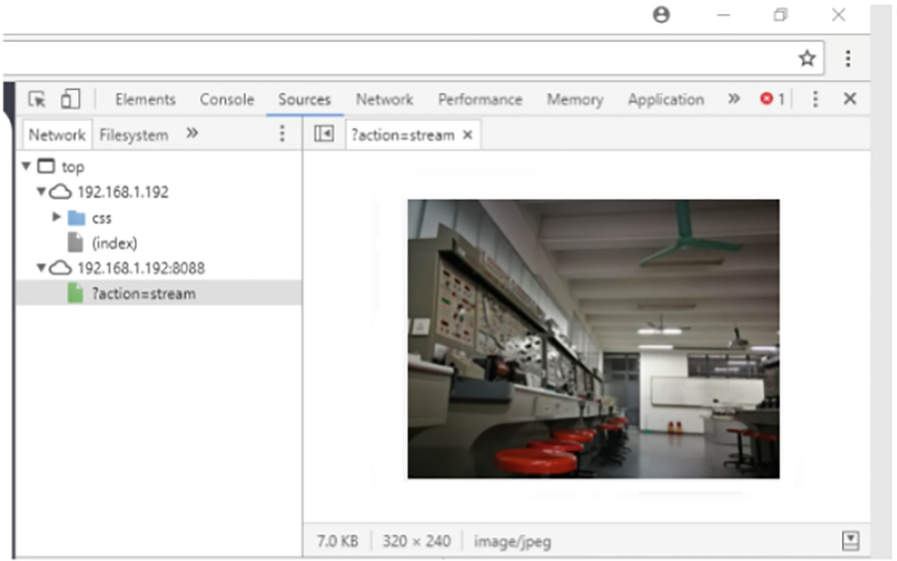
Fig. 11. Transmission image resolution of a video

Table 2. Image frame rate statistics

The starting point (ms)	The finishing point (ms)	Number of frames	Frame rate (FPS)
7.2	1016.1	30	29.74
1016.1	2031.8	30	29.54
2031.8	3059.3	30	29.20
3059.3	4064.1	30	29.86
4064.1	5119.9	29	27.47
5119.9	6129.2	31	30.71
6129.2	7147.0	30	29.48
7147.0	8156.5	29	28.73

The design initially realizes the photovoltaic charging and wireless video transmission control function of the robot car. The system can be applied to the general remote video transmission control such as environmental detection and home security.

References

1. Sun, J.: Research on mobile robot remote video transmission control. Nanchang University (2016)
2. Ding, N.: Portable solar sine wave inverter power supply design. J. Electron. (18), 10 (2014)
3. Zhu, R.: A digital image collection and processing system. J. Inf. Commun. (5), 58–59 2014
4. Wang, J.: The application of TB6612FNG in DC motor control design. J. Electron. Des. Eng. (2010)
5. Huang, Q.: Simulation research of embedded system based on RealView MDK software. Electron. Test (03), 63–66+85 (2009)
6. Acuity-. RT-Thread – domestic RTOS. Benchmarking. https://blog.csdn.net/qq_20553613/article/details/80008099
7. Xiong, P.: LwIP - lightweight TCP/IP protocol stack. https://github.com/RT-Thread/rtthread-manual-doc/blob/master/zh/1chapters/13-chapter_lwip.md
8. Chen, M.: The design of the embedded web server based on STM32. Wuhan University of Technology (2013)

Application of Intelligent Calculation Method in the Cage Simulation

Xinfeng Zhang[1], Yunwei Zhang[2], Ziqian Huang[1], Ying Yuan[1], and Xiangzhong Wei[1(✉)]

[1] College of Pharmacy, Guangxi University of Chinese Medicine, Nanning 530200, China
woxoz@163.com, xinfengzhang@yeah.net
[2] School of Foreign Languages and Literature, Nanjing Tech University, Nanjing 211800, China

Abstract. In the present study, in order to improve the calculation efficiency and reduce the computing time, we proposed the intelligent calculation method in the numerical simulation of the cage. On the basis of previous literature research, we first analyzed the force on the basic unit of mesh line in the cage, and established the mathematical model of the cage. Then the intelligent algorithm structure of different types of mass point in the cage is classified, and the specific implementation of the intelligent algorithm is illustrated. The results show that our intelligent algorithm can greatly reduce the calculation time and improve the efficiency of the numerical simulation of the cage considerably.

Keywords: Intelligent calculation · Cage simulation

1 Instruction

With the development of computer technology, numerical simulation in computer plays a more and more important role in the research of fishing nets. In the world, seawater culture cage is a popular fishing gear. The digitization and automation of the design of fishing nets has become the front field in the world. In 1998, Bris and Marichal first used the finite element method to simulate the shape of the cage in the water flow [1]. In order to avoid the generation of mesh pressure, Bris and Marichal divided each mesh foot into two. The shape of the cage in the water flow was not rough but more real. In order to study the underwater dynamic situation of the cage, Suzuki et al. established the Newtonian mechanical model by using the mass spring model in 2003, and developed a 3D visualization program with OpenGL. Suzuki et al. simulated three kinds of cage models at different flow rates, and compared with the real cage load, their fitting effect was better [2]. Li Yucheng et al. used the concentrated mass method to model the gravity cage in 2006, and the Runge-Kutta algorithm was introduced into getting the resistance value and shape of the mesh [3]. Zhao et al. used the lumped mass method and Newtonian mechanical model to simulate the deformation of cylindrical gravity cage under different currents and gravity in 2007 [4–6]. In 2008, Lee et al. simulated the dynamic behavior of cylindrical cage in current wave [7]. In 2011, Huang Xiaohua and others conducted

numerical simulation on the deformation of deep-water cage [8]. Xinfeng Zhang et al. proposed matrixing network and distributed computing method in fishing net simulation [9, 10].

Numerical simulation of fishing gear has a very huge amount of computation, which is also a difficult task for computer. In order to improve the calculation efficiency on the basis of mass spring system, lumped mass method and Newtonian mechanical modeling mentioned above, the intelligent algorithm is adopted in our present study of cage simulation. We apply this intelligent algorithm to simulate the dynamics of cage under water flow. Our intelligent algorithm can greatly improve the computing efficiency and reduce the calculation amount.

2 Modelling

The cage (Fig. 1) is composed of mesh (Fig. 2), assuming that the mesh line can be regarded as a cylinder. And the resistance coefficient is calculated with reference to the cylinder; it's assumed that the mesh line is elastic and can be stretched within a certain range without permanent deformation, and the mesh line diameter does not change. The mesh line only bears tension; it is assumed that the flow is uniform and the inertia force is zero. We take an 8 by 10 meshes cage to fix on a square frame around as an example. Given the parameters of the mesh line: diameter, original length, water resistance coefficient, elastic modulus, water density and velocity.

Each mesh line is regarded as a mass free and extensible spring (Fig. 3), which concentrates the mass of the mesh line on the nodule; The mass of each nodule is half of the sum of the four mesh lines connected with it. The mesh line is regarded as a cylinder, and its resistance coefficient is calculated according to the cylinder. The water resistance of the mesh line is also concentrated on the nodule for calculation. Figure 3 is the stress diagram of the ith mass point. Each mass point is connected with the adjacent mass point to form the mass point structure model of the cage (Fig. 1).

According to Newton's second law, the basic motion equation of the ith mass point is:

$$M_i a = T - F - W + B \tag{1}$$

a is the acceleration in the motion of mass point i. M_i is the mass of mass point i. T is the tension of mesh line acting on mass point i. F is the hydrodynamic force of mesh line (concentrated on the nodule). W gravity and B buoyancy of mesh line.

The tension of mesh line between adjacent mass points i and j is:

$$T_{ij} = E \pi d^2 (l_{ij} - l_0)/(4l_0) \tag{2}$$

l_0 is the original length of each mesh line. d is the diameter of mesh line. E is the modulus of elasticity. l_{ij} is the actual distance between two adjacent mass points i and j.

$$F = 1/2 C_D \rho d l_0 v^2 \tag{3}$$

The hydrodynamic coefficient is expressed by C_D. ρ is the density of sea water. And v is the velocity of the mass point relative to water.
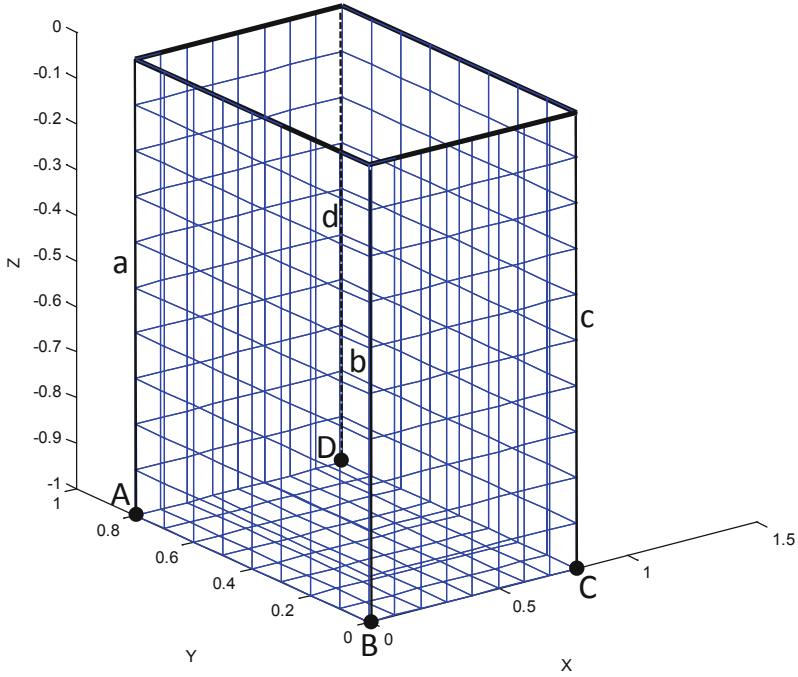
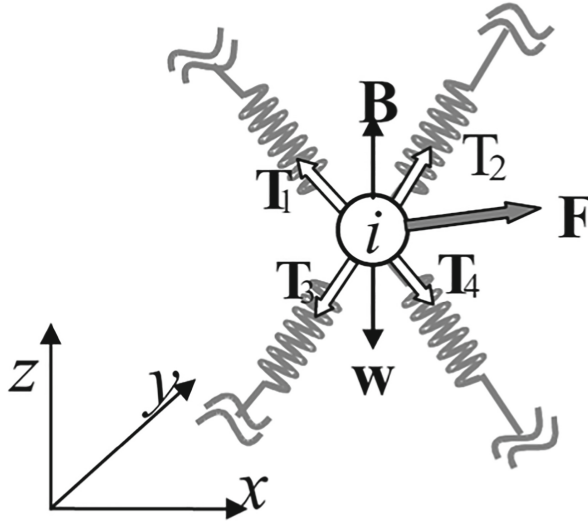**Fig. 1.** Mathematical structure of cage

Fig. 2. Part net of cage

Fig. 3. Force diagram of mass point i

To establish a three-dimensional coordinate system, the basic equation of model (1) can be written as:

$$M_i x_i'' = T\cos\alpha - 1/2 C_D \rho\, dl_0 (x_i' - v_x)^2$$
$$M_i y_i'' = T\cos\beta - 1/2 C_D \rho\, dl_0 (y_i' - y_x)^2 \tag{4}$$
$$M_i z_i'' = T\cos\gamma - 1/2 C_D \rho\, dl_0 (z_i' - z_x)^2 - W_i + B_i$$

Direction cosine is

$$\cos\alpha = (x_j - x_i)/l_{ij},\ \cos\beta = \left(y_j - y_i\right)/l_{ij},\ \cos\gamma = \left(z_j - z_i\right)/l_{ij} \tag{5}$$

3 Intelligent Calculation

The above mathematical model is too complex, and there is no analytic solution of algebraic type. So the numerical solution can only be obtained by certain numerical calculation methods. This differential equation model belongs to the rigid system differential equations. It is difficult to get stable and accurate numerical solutions by general numerical methods. Most of the general numerical methods do not converge at all. In order to calculate with high stability and accuracy, this paper uses the ultra-high level Runge-Kutta method to calculate and solve the model. Coefficients of Runge-kutta formula are:

$$a21 = 2/9;$$
$$a31 = 1/12; \quad a32 = 1/4;$$
$$a41 = 1/8; \quad a42 = 0; \quad a43 = 3/8;$$
$$a51 = 23/216; \quad a52 = 0; \quad a53 = 7/72; \quad a54 = -1/27;$$
$$a61 = -4136/729; \quad a62 = 0; \quad a63 = -4528/243;$$
$$a64 = 5264/729; \quad a65 = 1456/81;$$
$$a71 = 8087/11664; \quad a72 = 0; \quad a73 = 484/243;$$
$$a74 = -518/729; \quad a75 = -658/351; \quad a76 = 7/624;$$
$$a81 = -1217/2160; \quad a82 = 0; \quad a83 = -145/72;$$
$$a84 = 8342/6615; \quad a85 = 361/195;$$
$$a86 = 3033/50960; \quad a87 = 117/490;$$
$$a91 = 259/2768; \quad a92 = 0; \quad a93 = -84/173;$$
$$a94 = -14/173; \quad a95 = 6210/2249; \quad a96 = -99873/251888;$$
$$a97 = -29160/15743; \quad a98 = 2610/2249;$$
$$b1 = 173/3360; \quad b2 = 0; \quad b3 = 0; \quad b4 = 1846/5145; \quad b5 = 27/91;$$
$$b6 = -19683/713440; \quad b7 = -19683/713440; \quad b8 = 27/91; \quad b9 = 173/3360;$$
$$y(n+1) = y(n) + h * (b1 * k1 + b2 * k2 + b3 * k3 + b4 * k4 + b5 * k5 + b6 * k6 + b7 * k7 + b8 * k8 + b9 * k9)$$

$$(6)$$

We put the cage (Fig. 1) in a three-dimensional matrix. If a data point of the matrix is a mass point of the cage, it is represented by a solid point. Otherwise, it is represented by a hollow point. In this way, the structure of the mass points and the adjacent mass points on the cage can be roughly divided into 11 types (Fig. 4, 5 and 6). The mass points on the four edges on the bottom can be dealt with similar to those on the four edges of the cage in the vertical direction.

Then, we consider the real mass points (solid points) in the three-dimensional matrix as valid data, and other points (hollow points) in the three-dimensional matrix as invalid data (such as infinite 'Inf'). In the calculation of model solving, if two adjacent mass points are both effective mass points (solid points), the real calculation will be carried out. If one of the two adjacent mass points is invalid (hollow points), all forces acted on the mesh line between the two mass points are looked on as zero (including tension T, weight W, buoyancy B, and hydrodynamic force F, etc.)
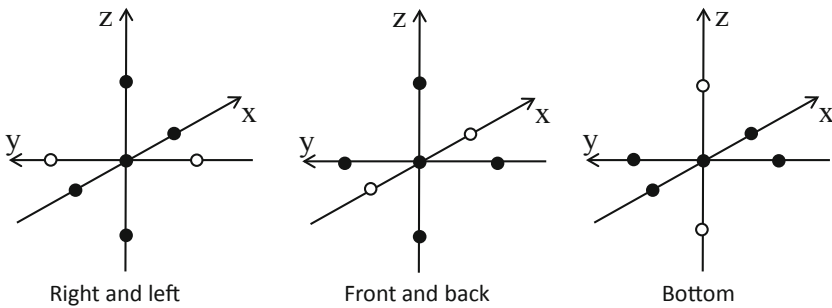

Fig. 4. The structure of mass points (solid point) on five planes of the cage
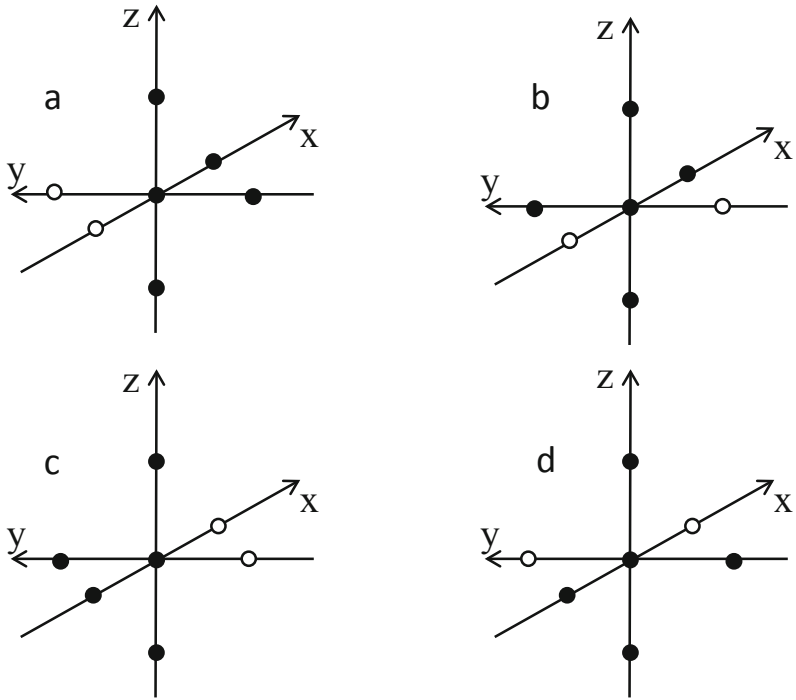
Fig. 5. The structure of mass points (solid point) on four vertical edges

The realization of intelligent calculation is in the process of programming, it is not difficult for us to intelligentize the above-mentioned calculation ideas in programming. We need to perform formula calculation between two real mass points (solid points), and carry out zero calculation for various forces between the invalid mass points (hollow points) and the real mass points (solid points); it is easy to use the judgment statements "if" and "else" to determine whether the mass point is real or not. We only give an example of the programming of the tension T in X direction (in MATLAB language) to illustrate the realization of intelligent calculation; and the programming of other force functions is completely similar, but only the calculation formula is different.

```
function [Tx]=Tx(xi,xj,yi,yj,zi,zj)
e=2.83*10^9; d=0.003; 10=0.1;
if xj==Inf
[Tx]=0;
else
  lij=sqrt((xi-xj)^2+(yi-yj)^2+(zi-zj)^2);
  [Tx]=e*pi*d^2*(lij-10)/(4*10)*[(xj-xi)/lij];
end
```

When we calculate the whole solution of the cage in the main program, in order to ensure that the current mass point i is an effective mass point (solid point), we can use the

special statement "continue" and the judgment statement "if" to intelligently avoid the invalid mass point (hollow point) at the beginning of iterative calculation. The following is a simple implementation program (in MATLAB language).
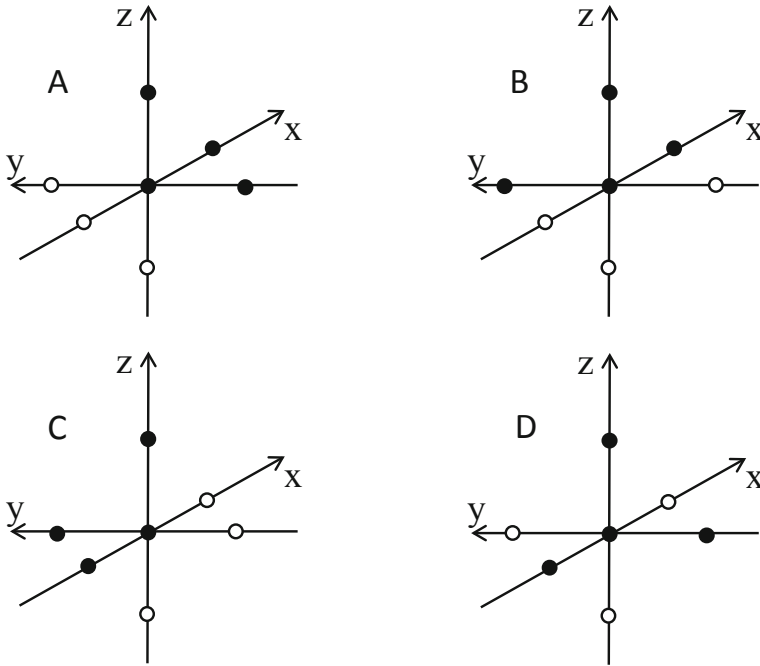
```
if xi==Inf
   continue
end
```


Fig. 6. The structure of mass points (solid point) on four corners at the bottom

4   Conclusion

The principle of running step (iterative times) and time step (s): the smaller the step, the more accurate the result is; and the larger the step, the more time saving in the calculation. Generally speaking, the time step length is not more than 0.001 s. In order to obtain more accurate results, the step length of time is 0.0002 s in the present study. After about 5 s, the cage can reach the balance state. When calculate it to 6 s, the main program needs to run 60000 steps, and the computing time in computer is about 12 min (in Matlab and under Windows 10 with CUP i5 8500) using the above intelligent calculation method. In another example, we tested with no intelligent algorithm (real mass points were programmed and

calculated one by one, for different types of the 11 mass points such as the point at the edge or corner of the cage, we programmed different function programs), the calculation time (60000 steps with time step of 0.0002 s) in the same computer was more than 4 h. Obviously, the intelligent calculation method greatly reduces the computing time and improves considerably the numerical calculation efficiency of the cage (Fig. 7).
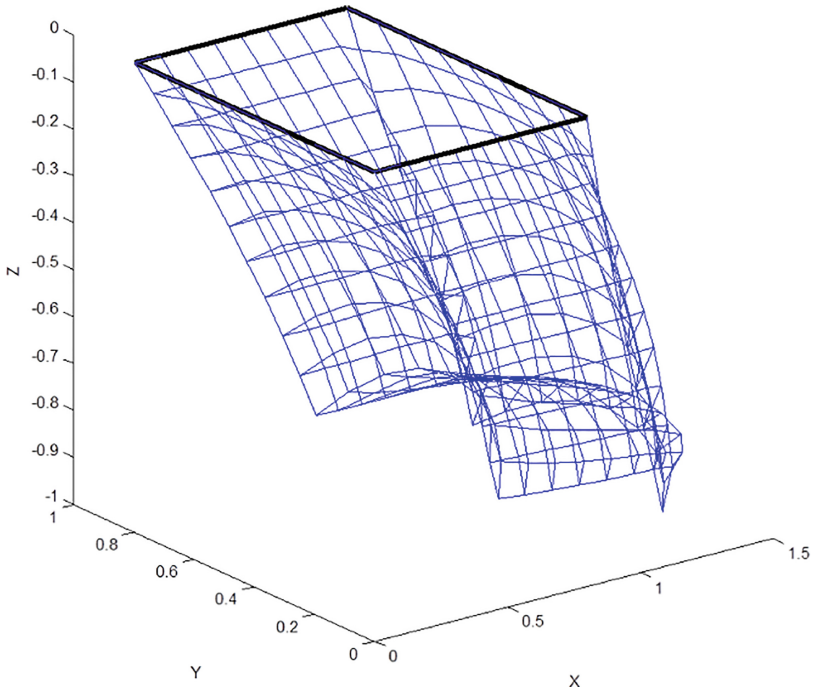

Fig. 7. The calculated shape of the cage when it reaches equilibrium

References

1. Le Bris, F., Marichal, D.: Numerical and experimental study of submerged supple nets: applications to fish farms. J. Mar. Sci. Technol. **3**, 161–170 (1998)
2. Suzuki, K., Takagi, T., Shimizu, T., Hiraishi, T., Yamamoto, K., Nashimoto, K.: Validity and visualization of a numerical model used to determine dynamic configurations of fishing nets. Fish. Sci. **69**(4), 695–705 (2003)
3. Li, Y., Zhao, Y., Gui, F.: Numerical simulation of the influencs of sinker weight on the deformation and load of net of gravity sea cage in uniform flow. Acta Oeanologica Sinica **25**(3), 125–137 (2006)

4. Zhao, Y.: Numerical simulation of the effects of weight system on the hydrodynamic behavior of 3-D net of gravity cage in current. J. Hydrodyn. **19**(4), 442–452 (2007)

5. Zhao, Y.-P., Li, Y.-C., Dong, G.-H., Gui, F.-K., Teng, B.: Numerical simulation of the effects of structure size ratio and mesh type on three-dimensional deformation of the fishing-net gravity cage in current. Aquacult. Eng. **36**, 285–301 (2007)

6. Zhao, Y.-P., Lia, Y.-C., Dong, G.-H., Gui, F.-K., Teng, B.: A numerical study on dynamic properties of the gravity cage in combined wave-current flow. Ocean Eng. **34**, 2350–2363 (2007)

7. Lee, C.-W., Kim, Y.-B., Lee, G.-H., Choe, M.-Y., Lee, M.-K., Koo, K.-Y.: Dynamic simulation of a fish cage system subjected to currents and waves. Ocean Eng. **35**, 1521–1532 (2008)

8. Huang, X., Guo, G., Hu, Y., Tao, Q., Zhang, X.: Numerical simulation of forces and motion deformation of deep-water net cages in waves and currents. J. Fish. Sci. China **18**(2), 443–450 (2011)

9. Zhang, X., Lia, Y., Song, L., Xu, L., Wang, M., Zhang, J., Zou, X., Zhang, M., Chen, X.: Matrixing network and distributed computing in the simulation of fishing nets. Procedia Eng. **37**, 79–84 (2012)

10. Zhang, X., Xu, L., Song, L., Zhang, J., Li, Y.: Effects of inertial mass coefficient on knotless netting model used in tuna purse seine. In: Applied Mechanics and Materials, vol. 256–259, pp. 1980–1984 (2013)

An Analysis of Multi-organ Segmentation Performance of CNNs on Abdominal Organs with an Emphasis on Kidney

Mahmud Elahi Akhter, Ashfia Binte Habib[✉] [iD], Rishad Arfin, Fahimul Haque,
Syed Athar Bin Amir, Zunayeed Bin Zahir, Md Shahriar Hussain, and Rajesh Palit

North South University, Dhaka, Bangladesh
ashfia.habib@northsouth.edu

Abstract. Manual hand tailored biomedical image segmentation of different
organs is a time consuming and laborious task. However, for the last decade or so,
many deep learning convolutional neural network (CNN) models have emerged
claiming to have close to human level results on biomedical image segmentation
of different type organs while automating the task. Multi-organ segmentation is
the process of segmenting multiple organs of the same patient. This offers a con-
venient solution to automation by providing segmentation of multiple organs at
a time. Since 2015, we seen massive improvements of deep CNNs. This has led
to better multi-organ segmentation architectures and has influenced the study of
multi-organ segmentation. In this paper, we analyze the performance of different
multi-organ segmentation studies. Our main focus was kidney, spleen and pan-
creas. We emphasized on kidney with the believe that we will see an increase in
kidney segmentation tasks and challenges. We found that multi-organ segmenta-
tion architectures have been improving over time and are performing quite well.
However, we also found that there is substantial performance variance across
the different studies even after using the same architecture and datasets on those
studies.

Keywords: Multi-organ segmentation · Deep learning · Convolutional neural
network

1 Introduction

In recent years, in the biomedical semantic segmentation and image analysis field, deep
CNNs has seen an upsurge of usage and also has been quite successful. We are seeing
much more sophisticated architectures that are performing better than our current existing
state-of-the-art architectures. Although Most of the architectures tend to address single
organ segmentation tasks, one of the most challenging task of biomedical semantic seg-
mentation is multi-organ semantic segmentation. Currently biomedical image segmen-
tation tasks are carried out by professionals. This is time consuming and labor intensive.
An automated solution can help with this issue and ensure efficient distribution of human
resources. Single organ segmentations are certainly helpful but automated multi-organ

segmentation can help much more in this regard because multi-organ segmentation can segment multiple organs from a single patients MR image or CT image.

One major challenge for multi-organ segmentation is to address the issue of different sized organ which is also true for single organ segmentation tasks. CNN architectures especially U-Net [1] like architectures suffer from blurred image features. The lack of high resolution edge information of the input and difficulty of extracting high level global features is also another constraint for these architectures [2]. It is also known that small organ segmentation is still suboptimal in terms of results [3].

With that in mind, in this paper, we analyze the performance of abdominal multi-organ segmentation from different studies to show the variation between the performance of different architectures. We believe that this study can help future multi-organ segmentation researches to setup more robust benchmarking and use different metrics to better evaluate their architectures. This is also true for single organ segmentation tasks. The scope of our study is abdominal multi-organ and in order to keep it concise we are only showing the results of segmentation on kidney, spleen, aorta, pancreas. However, we emphasized on kidney more as we believe with the release of KiTS19 [4] we will see more kidney segmentation. In the following subsection, we will present different approaches that have been taken for multi-organ segmentation. Afterwards, we move on to results of different studies and finally discuss our findings.

2   Related Work

There are many different approaches towards multi-organ segmentation which range from shape learning models, Multi-Atlas Segmentation (MAS), machine learning, statistical models, Convolutional neural network (CNN) and many more. However, most popular approach in the recent years seems to have been CNN architectures, specifically encoder decoder CNN architectures. In this section, we explore some different approaches that have been employed over the years for multi-organ segmentation.

Shape learning and machine learning models has success when used for the purpose of multi-organ segmentation. Cerrolaza J.J. proposed a shape model using their own statistical framework, which used canonical correlation analysis (CCA) along with generalized PCA to segment spleen, pancreas, liver, kidney, gallbladder and stomach [5]. Heinrich M.P trained a random forest based on vantage points tree which used binary context feature and achieved accuracy of ≥90% Dice on multi-organ abdominal segmentation on CT scans [6]. Cai J proposed a hybrid model which was used to segment abdominal organs and it consisted of a shape learning architecture and a CNN based architecture. They used an adversarial function to optimize the end to end shape learning architecture which as stated earlier was trained together with a CNN to perform both shape learning and segmentation task [7]. Different FCN and variants of FCN models have also been used for multi-organ segmentation. Wang C. used Multi-task FCN for different organs of the chest, where objective function for each organ was different. Their method converged faster and produced better segmentation than that of a single FCN with combined objective function for multiple classes [8]. Chen *et al.* used three 3D FCN networks for voxel wise dense prediction to segment thoracic (Lungs) and abdominal organs (kidney, liver and spleen). They used linear combination network, multi-channel

network and multi-image network, and achieved a Dice coefficient score of 98% on right lung [9]. Use of densely connected networks have also been explored. Qin Y. *et al.* used autofocus convolutional layers to segment prostate, bladder rectum, femur (both) and brain. They used auto focus to generate more powerful context as autofocus layers adaptively vary different effective receptive field [10]. Gibson E. *et al.* used a dense dilated convolutional network which used dense skip connections along with a spatial prior map to segment abdominal organs (liver, pancreas, stomach and esophagus) [11]. L. Chen *et al.* proposed an architecture called DRINet which consisted of a densely connected convolutional layer, a deconvolutional block with residual inception and an unpooling block. In terms of location, intensity, shape and size of an organ, standard convolutional layers cannot distinguish between their slight differences, which DRINet tried to address. In terms of multi-organ segmentation of abdominal CT images, it surpassed standard U-Net [12]. Chen *et al.* used cascaded 3D U-Net to segment dual energy computed tomography (DECT). The architecture had two stages. First stage generated the region of interest (ROI), which was used as an input (input mask) for the second stage. The second stage used two cascaded 3D U-Net with an analysis and synthesis path [13]. Navarro F. et al. showed that using complementary task learning could improve the results of multi-organ segmentation. He used a distance map regression and contour map detection to encode geometric properties of organs. These were used along with a U-Net like architecture which resulted in a segmentation map, a distance map and a contour map which in terms improved their multi-organ segmentation results [14]. Lei W. created DeepIGeo-V2 architecture which consisted of 2 stages: L1 Net and L2 Net. As input, the L1 Net takes in raw images and afterwards produces segmented images as output. In these outputs, the miss-segmentations are refined by L2 Net to produced better and refined segmentation results compared to automatic CNNs [15]. Multi planar segmentation methods have also been used for multi-organ segmentation. In their paper, Xia Y. *et al.* showed that there are many pros and cons when it comes to using 2D or 3D networks. They utilized 2D networks to train on each orthogonal plane of a 3D voxel. Afterwards, used a 3D Volumetric Fusion Network (VFN), which fused the 2D segmentation results for very refined localization of multi-organ [16]. Zhou, Y. showed that Deep Multi-Planar Co- Training (DMPCT) performed 4% better than that of fully supervised methods when only small number of annotations are used [17]. Her results have been used in this study in the results section. Lu X. proposed an algorithm that breakdowns a volumetric 3D image into slices and pixels in the 2D space. Afterwards, they applied dual architectures, mainly FCN and CNN, on each orthogonal plane and combined the feature probability maps using cross-sectional fusion which showed very promising results for organ localization [18]. Apart from these, vanilla networks have also been used to achieve good results. Wanyi Fu *et al.* used 3D U-Net and segmented 23 different organs and they got a DSC of 0.95 for liver and 0.94 for lungs [19]. On the other hand, Chen *et al.* used 2D U-Net and DenseUnet on MRI to segment abdominal organs which achieved dice scores between 0.87 to 0.96 except for duodenum [20].

3   Results

In this section, we present our analysis on performance of different architectures. In Table 1, we have listed all the studies that were covered in this paper. In Table 2, we have presented results of kidney segmentation. In Table 3 and 4 spleens, and pancreas segmentation results have been shown respectively. To keep our scope concise, we have only reported partial results of these architectures. We have used the # sign to signify partially reported results. Almost all studies used CT scans as dataset. However, a few studies used MRI as dataset and ~ sign has been used to signify those.

Table 1. List of publications on multi-organ semantic segmentation

Architectures	Listed as	Name of the publications
ResNet	Sinha [21]	Multi-scale guided attention for medical image segmentation
	Gibson et al. [22]	Automatic multi-organ segmentation on abdominal CT with dense V-networks
	Fang [23]	Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction
FCN	Zhou et al. 1 [17]	Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training
	Zhou et al. 2 [24]	Prior-aware neural network for partially-supervised multi-organ segmentation
	Valindria [25]	Multi-modal learning from unpaired images: application to multi-organ segmentation in CT and MRI Vanya
	Wang et al. [26]	Abdominal multi-organ segmentation with organ-attention networks and statistical fusion
U-Net	Heinrich [27]	OBELISK-Net: fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions
	Groza et al. [28]	Comparison of deep learning-based techniques for organ segmentation in abdominal CT images V
	Kakeya et al. [29]	3D U-JAPA-Net: mixture of convolutional networks for abdominal multi-organ CT segmentation
CNN	Berger [30]	An adaptive sampling scheme to efficiently train fully convolutional networks for semantic segmentation
	Hu et al. [31]	Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets

Table 2. Results of kidney segmentation as per publications

Publications	Architectures	Results (as per evaluation metric)			
		BTCV dataset (Dice %)		**KiTS dataset (Dice %)**	
Fang # [23]	U-Net	89.7		95.8 ± 0.91	
	ResU-Net	91.3		94.8 ± 1.06	
	DenseU-Net	–		94.2 ± 2.08	
	Proposed	92.7		96.2 ± 1.02	
		Left kidney (DSC %)		**Right kidney (DSC %)**	
Zhou et al. 1 # [17]	FCN	95.12 ± 5.01		95.69 ± 2.36	
	SPSL	95.88 ± 3.68		96.14 ± 2.94	
	Proposed	96.09 ± 3.42		96.26 ± 2.29	
		Left kidney		**Right kidney**	
		Dice (%)	**ASD (mm)**	**Dice (%)**	**ASD (mm)**
Wang et al. # [26]	3D U-Net	83.9 ± 22.4	0.77 ± 1.04	88.0 ± 14.4	1.39 ± 2.01
	HFCN	95.02 ± 2.6	0.41 ± 0.42	95.6 ± 4.5	1.03 ± 1.68
	FCN MV	96.01 ± 2.0	0.36 ± 0.47	95.8 ± 4.9	1.05 ± 1.74
	Proposed	96.8 ± 1.9	0.30 ± 0.30	98.4 ± 2.1	0.45 ± 0.89
		Left kidney (DSC%)		**Right kidney (DSC%)**	
Groza et al. [28]	F-Net	0.890		0.883	
	BRIEF	0.747		0.749	
	U-Net	0.876		0.872	
		DSC (%)	**JI (%)**	**ASD (mm)**	
Hu et al. # [31]	3D CNN	95.4 ± 3.2	91.3 ± 5.3	2.4 ± 3.4	

Table 2. (*continued*)

Publications	Architectures	Results (as per evaluation metric)					

Valindria # [25]

	Architectures	Left kidney (Dice %)					
		MR	**CT**		**MR**	**CT**	
Valindria # [25]	DLTK FCN	0.833	0.851		0.833	0.851	

Gibson et al. # [22]

	Architectures	Left kidney					
		Dice (%)	**Mean boundary distance (mm)**	**95% Hausdorff distance (mm)**			
Gibson et al. # [22]	VoxResNet	0.91	1.6	4.2			
	VNet	0.92	1.6	3.7			
	DenseVNet	0.95	0.9	3.1			

Sinha # ~ [21]

	Architectures	Left kidney			Right kidney		
		DSC	**VS**	**MSD**	**DSC**	**VS**	**MSD**
Sinha # ~ [21]	U-Net	82.51 ± 7.48	89.55 ± 4.68	0.61 ± 0.19	79.14 ± 15.23	87.68 ± 5.77	0.69 ± 0.38
	DA Net	84.49 ± 8.60	91.52 ± 6.73	0.67 ± 0.30	83.85 ± 9.40	92.88 ± 5.12	0.65 ± 0.31
	PAN (ResNet.34)	83.62 ± 6.21	91.83 ± 7.75	0.69 ± 0.21	81.51 ± 9.03	90.89 ± 5.64	0.75 ± 0.31
	PAN (ReNet101)	85.36 ± 4.87	92.69 ± 6.88	0.64 ± 0.15	85.02 ± 5.16	93.77 ± 4.63	0.61 ± 0.19
	Proposed	88.01 ± 6.16	94.95 ± 4.48	0.48 ± 0.14	87.96 ± 6.46	96.14 ± 3.15	0.48 ± 0.18

Kakeya et al. # [29]

	Architectures	Left kidney			Right kidney		
		Dice	**Recall**	**Precision**	**Dice**	**Recall**	**Precision**
Kakeya et al. # [29]	2D U-Net	**0.969**	0.965	0.973	0.972	**0.970**	0.973

Table 2. (*continued*)

Publications	Architectures	Results (as per evaluation metric)				
	3D U-Net	0.968	0.948	**0.992**	**0.978**	0.969	0.987
	3D M-U-Nets	0.981	0.977	0.985	0.943	0.929	0.988
	3D U-JAPA-Net	0.984	**0.981**	0.987	0.975	0.969	**0.990**
Berger # [30]		**Dice**				
	Dual CNN(val)	0.934				
	Dual CNN + isample (val)	0.938				
Heinrich # [27]		**Dice score**				
	Hybrid-OBELISK+CNN	942				
	All-conv. U-Net	946				
	Global FCN U-Net	943				
Zhou et al. 2 # [24]		**Left kidney (DSC)**			**Right kidney (DSC)**	
	AutoContext3DFCN DLTK U-Net	0.897			0.866	
	Proposed	0.915			0.895	
	AutoContext3DFCN	0.920			0.920	

Table 3. Results of spleen segmentation as per publications

Publications	Architectures	Results (as per evaluation metric)		
Fang # [23]		**(Dice %)**		
	U-Net	91.0		
	ResU-Net	90.9		
	PIPO-FAN(DPS+AF)	92.3		
Groza et al. [28]		**Dice (%)**		
	F-Net	0.885		
	BRIEF	0.751		
	U-Net	0.887		
Wang et al. # [26]		**Dice (%)**		**ASD (mm)**
	3D U-Net	89.6 ± 9.5		0.98 ± 0.56
	HFCN	93.1 ± 2.1		0.59 ± 0.37
	FCN MV	96.3 ± 1.9		0.60 ± 0.36
	Proposed	97.1 ± 1.5		0.42 ± 0.25
Kakeya et al. # [29]		**Dice**	**Recall**	**Precision**
	2D U-Net	0.946	0.946	0.952
	3D U-Net	0.932	0.904	**0.982**
	3D M-U-Nets	0.968	0.963	0.973
	3D U-JAPA-Net	**0.969**	**0.966**	0.973
Hu et al. # [31]		**Dice (%)**	**JI (%)**	**ASD (mm)**
	3D CNN	94.2 ± 5.4	89.5 ± 6.9	1.2 ± 1.8
Gibson et al. # [22]		**Dice**	**Mean boundary distance (mm)**	**95% Hausdoff distance (mm)**
	VoxResNet	0.91	2.1	3.8
	VNet	0.94	1.8	3.6
	DenseVNet	0.96	0.8	2.4
Sinha # ~ [21]		**DSC**	**VS**	**MSD**
	U-Net	71.95 ± 21.61	83.28 ± 14.78	1.76 ± 2.57
	DA Net	75.54 ± 16.08	84.37 ± 16.15	1.17 ± 0.94
	PAN (ResNet.34)	73.70 ± 19.97	81.98 ± 20.67	1.37 ± 1.43
	PAN (ResNet101)	74.84 ± 21.23	84.24 ± 17.37	1.30 ± 1.47
	Proposed	78.61 ± 18.69	87.87 ± 15.23	1.13 ± 1.24

(*continued*)

Table 3. (*continued*)

Publications	Architectures	Results (as per evaluation metric)		
Zhou et al. 1 # [17]		**DSC (%)**		
	FCN	94.81 ± 2.64		
	SPSL	95.58 ± 1.90		
	Proposed	95.69 ± 1.59		
Zhou et al. 2 # [24]		**DSC (%)**		
	AutoContext3DFCN	0.926		
	DLTK U-Net	0.939		
	Proposed	0.961		
Berger # [30]		**Dice (%)**		
	Dual CNN (val)	0.927		
	Dual CNN + isample (val)	0.941		

Table 4. Results of pancreas segmentation as per publications

Publications	Architectures	Results (as per evaluation metric)		
Zhou et al. 1 # [17]		**DSC (%)**		
	FCN	76.49 ± 11.6		
	SPSL	80.93 ± 6.84		
	Proposed	82.03 ± 6.16		
Heinrich # [27]		**Dice score**		
	Hybrid-OBELISK + CNN	70.2		
	All-conv. U-Net	66.9		
	Global FCN Unet	68.8		
Kakeya et al. # [29]		**Dice**	**Recall**	**Precision**
	2D U-Net	0.758	0.689	0.880
	3D U-Net	0.688	0.564	**0.968**
	3D M-U-Nets	0.806	0.776	0.867
	3D U-JAPA-Net	**0.861**	**0.816**	0.922
Gibson et al. # [22]		**Dice**	**Mean boundary distance (mm)**	**95% Hausdorff distance (mm)**
	VoxResNet	0.74	2.2	6.3
	VNet	0.71	2.9	9.5
	DenseVNet	0.78	1.9	5.9

4   Discussion

Although multi-organ segmentation is quite a challenging task, in our study we found that the interest in multi-organ segmentation study is growing as most of the publications covered in this paper was from 2018 and 2019. However, it is evident from this study that there is massive variance of performance even if the architecture is same (Fig. 1).
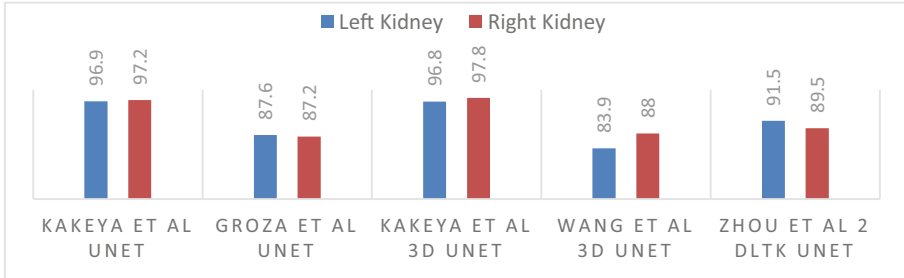
Fig. 1. Variation between 2D and 3D U-Net results for kidney segmentation

We can also see that in some publications 2D U-Net performed better compared to 3D U-Net, whereas 3D CNNs are known to work better than 2D CNNs. We see this trend in spleen segmentation results also. Fang et al. [23], Groza et al. [28], Kakeya et al. [29] and Zhou et al. 2 [24] all had DSC of 91.0, 88.5, 94.6 and 93.9 respectively, which shows the variance between different 2D U-Nets. We can also see variance between left and right kidney whereas both can be considered as the same organ. We can also see that even though other organs have high dice coefficient rate, pancreas segmentation tends to vary within 70% to 86% dice coefficient. Therefore, we believe having certain standard on benchmarking will surely help to see and compare performance analysis. Maintaining a database for result crosschecking will help future researcher to optimize more in different architectures. We would also like to mention DLTK toolkit [32] and NiftyNet [33] for future researchers who want to break into deep learning based biomedical image analysis. DLTK and NiftyNet both are open source. DLTK is a library whereas NiftyNet is a CNN platform. Both are tensorflow based. As stated, these can be a useful tool for future researchers who would like to start out in biomedical image analysis and or practitioners who would like to benchmark different architectures more rigorously.

5   Conclusion

In this paper, we analyzed the performance of different CNN architectures that are being developed and used for the purpose of multi-organ segmentation. Scope of this work was limited to multi-organ segmentation only, within certain limitation. We believe that in the light of the huge difference in performance that has been shown and discussed in the results section, there needs to be much more rigorous analysis to look into the variance of results for same architectures. We believe that having proper benchmarking

standards (No of test/training case for use, metrics etc.) to reduce the huge variations of results within same architecture class can be a solution.

Manual hand tailored segmentation can be time consuming and an arduous task. Automated multi-organ segmentation can help manage this tedious workload more efficiently. Automatic multi organ segmentation can help with both tumor and lesion detection. In this paper, we emphasized mostly on kidney as we believe we will see an increase in kidney segmentation tasks. Apart from that, we believe that, reducing the variance in performance will help achieve better performance and somewhat standardize the state-of-the art architectures.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. Springer, Cham (2015)
2. Seo, H., Huang, C., Bassenne, M., Xiao, R., Xing, L.: Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images. IEEE Trans. Med. Imaging (2019). https://doi.org/10.1109/TMI.2019.2948320
3. Valindria, V.V., et al.: Small organ segmentation in whole-body MRI using a two-stage FCN and weighting schemes. In: Shi, Y., Suk, H.I., Liu, M. (eds.) Machine Learning in Medical Imaging. MLMI 2018. Lecture Notes in Computer Science, vol. 11046. Springer, Cham (2018)
4. KiTS19 Challenge. http://kits19.grand-challenge.org/data/
5. Cerrolaza, J.J., Summers, R.M., Linguraru, M.G.: Soft multi-organ shape models via generalized PCA: a general framework. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol. 9902. Springer, Cham (2016)
6. Heinrich, M.P., Blendowski, M.: Multi-organ segmentation using vantage point forests and binary context features. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol. 9901. Springer, Cham (2016)
7. Cai, J., Xia, Y., Yang, D., Xu, D., Yang, L., Roth, H.: End-to-end adversarial shape learning for abdomen organ deep segmentation. In: Suk, H.I., Liu, M., Yan, P., Lian, C. (eds.) Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science, vol. 11861. Springer, Cham (2019)
8. Wang, C.: Segmentation of multiple structures in chest radiographs using multi-task fully convolutional networks. In: Sharma, P., Bianchi, F. (eds.) Image Analysis. SCIA 2017. Lecture Notes in Computer Science, vol. 10270, pp. 282–289. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59129-2_24
9. Chen, S., Zhong, X., Hu, S., Dorn, S., Kachelrieß, M., Lell, M., Maier, A.: Automatic multi-organ segmentation in dual-energy CT (DECT) with dedicated 3D fully convolutional DECT networks. Med. Phys. **47**, 552–562 (2020). https://doi.org/10.1002/mp.13950

10. Qin, Y., et al.: Autofocus layer for semantic segmentation. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol. 11072. Springer, Cham (2018)

11. Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B.R., Pereira, S.P., Clarkson, M.J., Barratt, D.C.: Towards image-guided pancreas and biliary endoscopy: automatic multi-organ segmentation on abdominal CT with dense dilated networks, pp. 728–736 (2017). https://doi.org/10.1007/978-3-319-66182-7_83

12. Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., Rueckert, D.: DRINet for medical image segmentation. IEEE Trans. Med. Imaging **37**(11), 2453–2462 (2018). https://doi.org/10.1109/TMI.2018.2835303

13. arXiv:1710.05379

14. Navarro, F., et al.: Shape-aware complementary-task learning for multi-organ segmentation. In: Suk, H.I., Liu, M., Yan, P., Lian, C. (eds.) Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science, vol. 11861. Springer, Cham (2019)

15. Lei, W., Wang, H., Gu, R., Zhang, S., Zhang, S., Wang, G.: DeepIGeoS-V2: deep interactive segmentation of multiple organs from head and neck images with lightweight CNNs. In: Zhou, L., et al. (eds.) Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention. LABELS 2019, HAL-MICCAI 2019, CuRIOUS 2019. Lecture Notes in Computer Science, vol. 11851. Springer, Cham (2019)

16. Xia, Y., Xie, L., Liu, F., Zhu, Z., Fishman, E.K., Yuille, A.L.: Bridging the gap between 2D and 3D organ segmentation with volumetric fusion net. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Lecture Notes in Computer Science, vol. 11073. Springer, Cham (2018)

17. Zhou, Y., Wang, Y., Tang, P., Bai, S., Shen, W., Fishman, E.K., Yuille, A.: Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training. In: Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019 (pp. 121-140) (2019). [8658899] (Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/WACV.2019.00020

18. Lu, X., Xu, D., Liu, D.: Robust 3D organ localization with dual learning architectures and fusion. In: Carneiro, G., et al. (eds.) Deep Learning and Data Labeling for Medical Applications. DLMIA 2016, LABELS 2016. Lecture Notes in Computer Science, vol. 10008. Springer, Cham (2016)

19. Fu, W., et al.: Multi-organ segmentation in clinical-computed tomography for patient-specific image quality and dose metrology. In: Medical Imaging 2019: Physics of Medical Imaging, vol. 10948, International Society for Optics and Photonics, p. 1094829 (2019). www.spiedigitallibrary.org. https://doi.org/10.1117/12.2512883

20. arXiv:1912.11000

21. arXiv:1906.02849

22. Gibson, E., et al.: Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE Trans. Med. Imaging **37**(8), 1822–1834 (2018). https://doi.org/10.1109/TMI.2018.2806309

23. arXiv:2001.00208

24. arXiv:1904.06346

25. Valindria, V.V., et al.: Multi-modal learning from unpaired images: application to multi-organ segmentation in CT and MRI. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, pp. 547–556 (2018)

26. Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E.K., Yuille, A.L.: Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Med. Image Anal. **55**, 88–102 (2019)

27. Heinrich, M.P., Oktay, O., Bouteldja, N.: OBELISK-Net: fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions. Med. Image Anal. **54**, 1–9 (2019). https://doi.org/10.1016/j.media.2019.02.006

28. Groza, V., Brosch, T., Eschweiler, D., Schulz, H., Renisch, S., Nickisch, H.: Comparison of deep learning-based techniques for organ segmentation in abdominal CT images (2018)

29. Kakeya, H., Okada, T., Oshiro, Y.: 3D U-JAPA-Net: mixture of convolutional networks for abdominal multi-organ CT segmentation. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol. 11073. Springer, Cham (2018)

30. Berger, L., Eoin, H., Cardoso, M.J., Ourselin, S.: An adaptive sampling scheme to efficiently train fully convolutional networks for semantic segmentation. In: Nixon, M., Mahmoodi, S., Zwiggelaar, R. (eds.) Medical Image Understanding and Analysis. MIUA 2018. Communications in Computer and Information Science, vol. 894. Springer, Cham (2018)

31. Hu, P., Wu, F., Peng, J., et al.: Int. J. CARS **12**, 399 (2017). https://doi.org/10.1007/s11548-016-1501-5

32. arXiv:1711.06853

33. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D.C., Ourselin, S., Cardoso, M.J., Vercauteren, T.: NiftyNet: a deep-learning platform for medical imaging. Comput. Methods Progr. Biomed. **158**, 113–122 (2018). https://doi.org/10.1016/j.cmpb.2018.01.025. ISSN 0169-2607

Correction to: Automatic Detection and Counting of Malaria Parasite-Infected Blood Cells

Elena Doering, Anna Pukropski, Ulf Krumnack ⓘ,
and Axel Schaffland ⓘ

Correction to:
Chapter "Automatic Detection and Counting of Malaria
Parasite-Infected Blood Cells" in: R. Su and H. Liu (Eds.):
***Medical Imaging and Computer-Aided Diagnosis*,**
LNEE 633, https://doi.org/10.1007/978-981-15-5199-4_15

In the original version of the book, the name of the fourth author of Chapter 15 was spelt incorrectly, which has now been corrected and updated.

Author Index