



The analysis of intelligent real-time image recognition technology based on mobile edge computing and deep learning

Tao Shen¹ · Chan Gao^{2,3} · Dawei Xu⁴

Received: 8 August 2020 / Accepted: 9 October 2020 / Published online: 22 October 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

This article aims to improve the accuracy of real-time image recognition in the context of the Internet of Things (IoT), reduce the core network pressure of the IoT and the proportion of IoT broadband, and meet people's demand for internet image transmission. An intelligent image fusion system based on mobile edge computing (MEC) and deep learning is proposed, which can extract the features of images and optimize the sum of intra-class distance and inter-class distance relying on the hierarchical mode of deep learning, and realize distributed computing with the edge server and base station. Through comparison with other algorithms and strategies on the text and character data sets, the effectiveness of the constructed system is verified in the performance of the algorithm and the IoT. The results reveal that the application of the unsupervised learning hierarchical discriminant analysis (HDA) has better accuracy and recall in various databases compared with conventional image recognition algorithms. When the sum intra-class and inter-class distance K is 2, the accuracy of the algorithm can be as high as 98%. The combination of MEC and layered algorithms effectively reduces the pressures of core network and shortens the response time, greatly reduces the broadband occupancy ratio. The performance of IoT is increased by 37.03% compared with the general extraction and common cloud computing. Image recognition based on the MEC architecture can reduce the amount of network transmission and reduce the response time under the premise of ensuring the recognition rate, which can provide a theoretical basis for the research and application of user image recognition under the IoT.

Keywords Internet of things · Mobile edge computing · Deep learning · Image recognition · Feature extraction · Hierarchical discriminant analysis

1 Introduction

Economy develops rapidly these years in China, so smart city, smart network, smart shopping, and other methods have emerged, and the construction of smart network is inseparable from IoT [1]. IoT utilizes key devices such as sensors and base stations to share data through the internet, form a unified whole of all things, and achieve rapid information transmission [2]. Different types of IoT devices rely on multiple communication networks to build a complex network system with all things connected around the world, serving scenarios such as smart city, health care, and smart home [3]. Smart devices will continuously generate much data in real time. IoT applications need to process continuous data streams generated by batch devices, filter and pre-process the data, and obtain information from the data to respond after further logic analysis [4]. This means that IoT data need a faster processing method [5]. However, due to the rapid increase of users and the continuous access of devices,

✉ Tao Shen
shentao@tongji.edu.cn

Chan Gao
chan.gao9987@gmail.com

Dawei Xu
xudawei@techfantasy.com.cn

¹ College of Design and Innovation, Tongji University, Shanghai, China

² Architecture Department, Huzhou University, Huzhou, China

³ School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan

⁴ College of Computer Science, Qinghai Nationalities University, Xining, Qinghai, China

IoT at this stage has caused large calculation delays. Second, the network has low reliability and is vulnerable to network attacks [6]. Thus, edge computing is proposed and applied to solve these problems of IoT [7]. The edge computing has a powerful interconnected edge computing platform with optimal network security functions and the highest security level. This strategy can realize network edge operations on nearby distributed servers based on the actual situation of network distribution, build a fused network system, and achieve some core network functions such as computing, storage, application, and transmission [8]. Therefore, analysing and improving the efficiency of IoT data processing and reducing network delay from the perspective of MEC are crucial to the construction of a smart network city.

As an important task in the development of the times, image recognition has been extensively used in airports, mobile payments, security inspections, and law enforcement. At present, the use of MEC can satisfy the smart environmental protection, smart detection, and smart security [9]. However, the usage scenarios have become more and more complex with the continuous deepening of IoT technology, and the internet also faces increasingly greater challenges [10]. The traditional cloud computing is to directly upload images to the data centre for recognition, and remotely provide computing and storage functions. However, the distance among users in the cloud data centre is long, which will cause a long transmission delay and cannot satisfy users for image recognition. The real-time response to the demand brings huge pressure and leads to network congestion [11]. Image feature extraction is the most critical point for the image cognition. Yang et al. took the local binary feature extraction through rotation invariance based on the multi-channel convolution neural network (CNN), and the accuracy of which was higher by 1 time than the conventional binary extraction [12]. Ren et al. proposed a new method for deep learning image recognition with the latent features, the orthogonal matrix was adopted to preserve the data features of the original image, discriminated by the loss objective function; and the results showed that the method has good effectiveness and robustness [13]; Guo et al. proposed an efficient method of spectral space deep feature extraction and classification based on deep learning; it could make full use of the multi-scale spatial features obtained by the guided filter to improve the classification accuracy of the spatial image [14]. Based on the above researches, it is found that more and more scholars are beginning to apply deep learning technology for image feature extraction. Therefore, this article aims to combine deep learning and MEC method to improve image processing performance of IoT, reduce network delay of IoT, and truly meet user and construction needs.

An optimized and improved deep layered discriminant analysis algorithm is proposed in this study based on the

analysis of the classic feature extraction algorithm, which is the innovation of this study. It effectively integrates the deep learning algorithm with the layered discriminant algorithm, and achieves the intelligent real-time image detection through the MEC. The main contribution of this study is to adopt the hierarchical deep learning method to effectively and real-time extract the features of the image by optimizing the distance between adjacent neurons. The distributed method of MEC is applied to effectively reduce the transmission of cloud servers and increase the accuracy of image recognition. This study can provide a theoretical basis for improving the performance of the Internet of Things. To achieve the above solution, the objective and significance of this study are introduced in the first part. In the second part, the differences between the traditional IoT computing architecture and the MEC architecture are compared and analysed, and the shortcomings of the current IoT architecture are pointed out. Then the principle of deep learning HDA is proposed, based on which a real-time image recognition system is constructed under the MEC and deep learning. In the third part, the effectiveness of the algorithms proposed in this study is verified through model performance, parameter optimization, and IoT performance. Analysis is development based on the above results in the fourth section, and the main conclusions of this study are drawn in the fifth part.

2 Methodology

2.1 Computing framework of conventional IoT and mobile edge computing

The conventional IoT computing model is shown in Fig. 1. The user uploads the data directly to the base station,

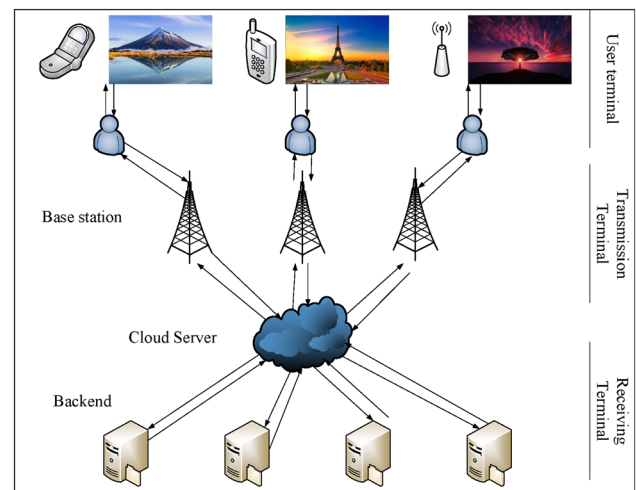


Fig. 1 Computing framework of conventional IoT

imports it into the cloud data centre for recognition, provides computing and storage functions remotely, and transmits the data back finally. However, the distance between the cloud data centre and user is too long, resulting in a long transmission delay. Thus, it cannot meet the user’s requirement of real-time response for image recognition. What’s more important that uploading the image recognition task to the core network will put tremendous pressure on broadband and cause network congestion.

MEC mainly solves the transmission delay of conventional mobile cloud computing. MEC uses edge servers near the data sources to provide computing and storage resources for applications, and processes the edge big data in a distributed manner. MEC is more suitable for real-time data analysis and intelligent processing, and is more effective than conventional IoT computing models. The MEC framework includes three parts: user, edge server, and cloud server. The specific structure of MEC is shown in Fig. 2.

2.2 Hierarchical discriminant analysis

HDA first constructs two adjacency graphs, and then minimizes the sum of intra-class distances, and then maximizes the sum of inter-class distance. The sum of intra-class distances and the sum of inter-class distances are hierarchically optimized, so the results of the optimization process will not be biased towards any side. Therefore, the sample points of the same class are compact, while the sample points of different classes are separated in the projection space. The specific calculation equation is as follows: for the sample X_i , it is assumed that its k same and different classes of adjacency sets are $P_k^+(X_i)$ and $P_k^-(X_i)$, respectively, then the neighbouring matrix F_{ij}^W can be expressed as below:

$$F_{ij}^W = \begin{cases} +1, & X_i \in P_k^+(X_j) \text{ or } X_j \in P_k^+(X_i) \\ 0 & \text{Otherwise} \end{cases}$$

The inter-class neighbouring matrix F_{ij}^B can be expressed as below:

$$F_{ij}^B = \begin{cases} +1, & X_i \in P_k^-(X_j) \text{ or } X_j \in P_k^-(X_i) \\ 0 & \text{Otherwise} \end{cases}$$

The sum of intra-class distances after projection $\theta(P_1)$ can be expressed as Eq. (3):

$$\theta(P_1) = \sum_{X_i \in P_k^+(X_j)} \sum_{X_j \in P_k^+(X_i)} \|P_1^T X_i - P_1^T X_j\|_{F_{ij}^W} = 2P_1^T X(D^W - F^W)X^T P_1.$$

In the above equation, D^W represents a diagonal matrix, F^W represents elements on the diagonal, T represents the number of feature points in all images, and P_1 is the number of feature points in the first image. First, the sum of intra-class distances is optimized so that samples of the same class can be gathered together in the projection space. That is to say, the smaller the distance among samples of the same class, the better. Therefore, the sum of distances among samples of the same class is minimized, and the objective function can be expressed as below:

$$\begin{cases} \min \theta(P_1) \\ \text{s.t. } P_1^T P_1 = I \end{cases}$$

The sum of inter-class distances after projection $\psi(P)$ can be calculated with Eq. (5):

$$\psi(P) = \sum_{X_i \in P_k^+(X_j)} \sum_{X_j \in P_k^-(X_i)} \|P_1^T X_i - P_1^T X_j\|^2 = 2P^T X(D^B - F^B)X^T P.$$

In the above equation, D^B represents a diagonal matrix and F^B represents an element on the diagonal. The second step is to optimize the sum of inter-class distances. In the projection space, the larger the distance among sample points of different classes, the better. Therefore, when optimizing the sum of inter-class distances, it has to be minimized. Its objective equation is given as below:

$$\begin{cases} \max \theta(P_1) \\ \text{s.t. } P_1^T P_1 = I \end{cases}$$

s.t. indicates that the equation meets a certain condition, and I is the feature extraction calculation breakpoint. After minimizing the sum of the distances among samples of the same class, the data set has already undergone feature extraction for once. Therefore, after the first step of feature extraction, the extracted feature extraction point $\psi(P)$ can be expressed as below:

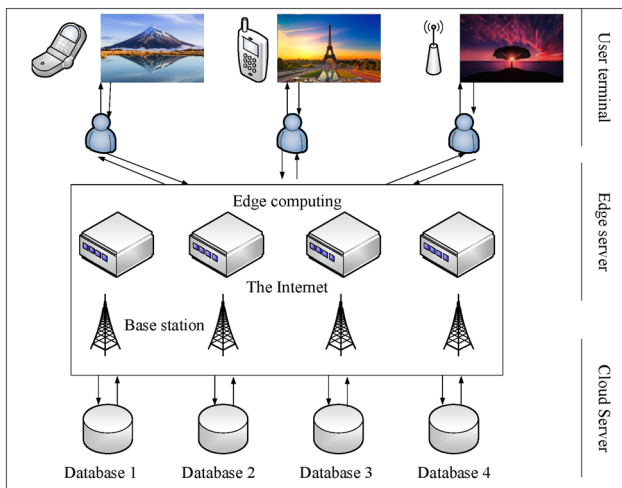


Fig. 2 Framework of mobile edge computing

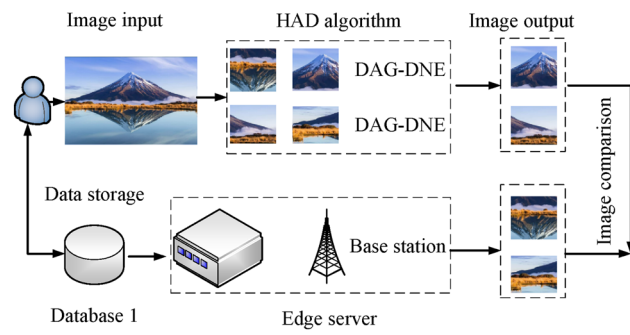


Fig. 3 Construction of real-time image recognition system based on MEC and deep learning

$$\psi(P) = 2\text{tr}\{P^T X_{\text{new}} M X_{\text{new}}^T P\}.$$

In the above equation, X_{new} is the image extraction point of the next layer, M is the characteristic constant, and tr represents the matrix calculation.

2.3 Construction of real-time image recognition system based on mobile edge computing and deep learning

In Fig. 3, the constructed real-time network recognition system based on MEC and deep learning is mainly composed of five parts: data input, feature extraction, data comparison, edge computing, and data collection, which constitute the real-time image processing IoT. In the feature extraction, the proposed HDA algorithm firstly minimizes the sum of intra-class distances, and then maximizes the sum of inter-class distances. Since the sum of intra-class distances and inter-class distances is optimized hierarchically, such calculation method is not biased to either side. Thus, the algorithm has a smaller load and runs faster in the projection space. In terms of edge computing, the layered image recognition edge computing method proposed in this article is more suitable for real-time data analysis and processing, and is more secure and effective than the conventional cloud computing methods. This method can solve the low efficiency of IoT data processing and reduce network delay.

The deep learning of this article refers to the convolutional neural network (CNN), which generates a hierarchical structure containing several levels by continuous convolution and down-sampling operations on the input image. CNN first transmits the sample information through forward propagation layer by layer to obtain the sample predicted value, and then transmits the error between the predicted value and the true value layer by layer to each layer using the back propagation (BP) algorithm. Finally, the partial derivative of the error cost function to the parameters of each layer is calculated, and the weight of each layer is adjusted and updated using the stochastic gradient descent method.

The network output can be more accurate by modifying the weight of each layer.

It is a network structure containing multiple levels, each level contains a large number of feature maps, and each feature map has many independent neurons. The CNN model for building a hierarchical structure is to adjust the original image to a 224×224 3-channel red, green, and blue (RGB) image as the input of the model, and then the input image is processed layer by layer to extract the depth semantic information of the image. 64 convolution kernels with a size of 11×11 are selected, and the traversal step size is set to 4. A convolution operation is performed on an input image of $224 \times 224 \times 3$ to obtain a feature map on the convolution output layer. The subsequent operations are as above, and the output results can be obtained by convolution layer by layer.

2.4 Source and training of data set

Data set source: the most important thing for image recognition in daily life is character feature recognition and text feature recognition. To verify the effectiveness of the algorithm proposed in this article, 7 different public data sets are adopted, including Yale face database (Yale), University of Manchester Institute of Science and Technology (UMIST), Olivetti research lab (ORL), and extended Yale face database B (YaleB) for character feature recognition, and ISOLET, BinAlpha, and USPS for text feature recognition.

As shown in Table 1, in the data sets for character recognition feature, Yale contains 165 different personal images processed with different lightings, and their facial expressions are different; UMIST is composed of 564 images of 20 people, focusing on different races, genders, and appearance characteristics; ORL focuses on different facial expressions without light; YaleB focuses on different postures and 64 lighting conditions. In the data sets of text recognition, the ISOLET contains 1560 common spoken letter images, and the BinAlpha collects different handwriting methods, including 1404 different styles. The USPS is similar to BinAlpha,

Table 1 Samples and classification numbers of different data sets

Classification of data set	Algorithm	Number of samples	Category	Dimension
Character recognition	Yale	165	15	1024
	UMIST	574	20	1024
	ORL	400	40	1024
	YaleB	2414	38	1024
Text recognition	ISOLET	1560	26	617
	BinAlpha	1404	36	320
	USPS	9298	10	256

but the data involves more handwriting samples, including 9298 types. Figure 4 shows partial sample data.

Data set training: all data sets are converted to the same resolution to ensure the consistency of input parameters, and then the data is reduced to 100 dimensions with principal component analysis (PCA) so that it can effectively remove the noise of data and reduce the complexity of system operation. Among them, 80% is used as the training data set, and the other 20% is undertaken as the test data set.

2.5 Indicators for system performance analysis

Recognition performance analysis: the commonly used feature extraction algorithm is selected for comparison, and Luma-dependent nonlinear enhancement (LDNE), modified firefly algorithm (MFA), directed acyclic graph-deoxyribonucleic acid (DAG-DNA), 1 high alcohol-drinking (HDA-1), 2 high-alcohol-drinking (HDA-2), fully-reuse convolutional neural network (FRCNN) are utilized for analysing the performance of the image recognition system [15]. The analysis is based on three dimensions: accuracy, recall, and mean average precision (mAP).

Accuracy refers to the proportion of accurately recognizing the image with the feature extraction algorithm, and it can be calculated with below equation:

$$\text{Accuracy} = \frac{A}{A + B}. \quad (8)$$

In the above equation, A is the number of images accurately and correctly recognized by the system with the feature extraction algorithm, and B is the number of images accurately and incorrectly recognized by the system with the feature extraction algorithm. The range of accuracy is $[0, 1]$. The larger the value, the higher the accuracy of the model.

For a given data set, recall refers to the ratio of the number of correctly classified samples in the correctly recognized

samples to the total number of samples. The calculation equation is as follows:

$$\text{Recall} = \frac{C}{C + D}. \quad (9)$$

In the above equation, C is the number of images accurately recognized and classified with the feature extraction algorithm, and B refers to the number of images inaccurately recognized and classified with the feature extraction algorithm. The range of the recall is $[0, 1]$. The larger the value, the more correctly recognized and classified samples of the model.

The mAP is to solve the single-point limitation of accuracy and recall. For individual cases, each category can obtain an average precision (AP) through the area under receiver operating characteristic (ROC), and the average AP is called the mAP, which can be calculated with the below equation:

$$\text{mAP} = \frac{\sum_{q=1}^Q \text{AP}}{Q}. \quad (10)$$

In the above Eq. (10), AP refers to the calculated area size after drawing the ROC curve according to any category in the target recognition task, Q is the number of all samples, and q is the sample value.

IoT performance analysis: it mainly analyses the response time and occupancy rate of broadband. Among them, the broadband response time mainly analyses mobile internet long term evolution (LTE) and wireless fidelity (WIFI). For the performance analysis of IoT, the analysis results obtained based on the MEC strategy are compared with the results not using the MEC. The general image feature extraction and HDA are adopted for feature extraction of images.

3 Results and discussion

3.1 Comparison on performance of image recognition system based on MEC and deep learning

First of all, the algorithm proposed in this study is compared with the commonly used image recognition algorithms and is further on different data sets to verify the effectiveness of proposed mobile edge computing and deep learning image recognition. The accuracy, recall, and mAP of the model are analysed from two perspectives: character feature extraction and text feature extraction. Figure 5a–d shows the comparison results on the accuracies of algorithms under different character feature recognition data sets. The figures show that the accuracies of the algorithms continuously increase with the increasing number of projections. Among them, the



Fig. 4 Partial face and text sample data

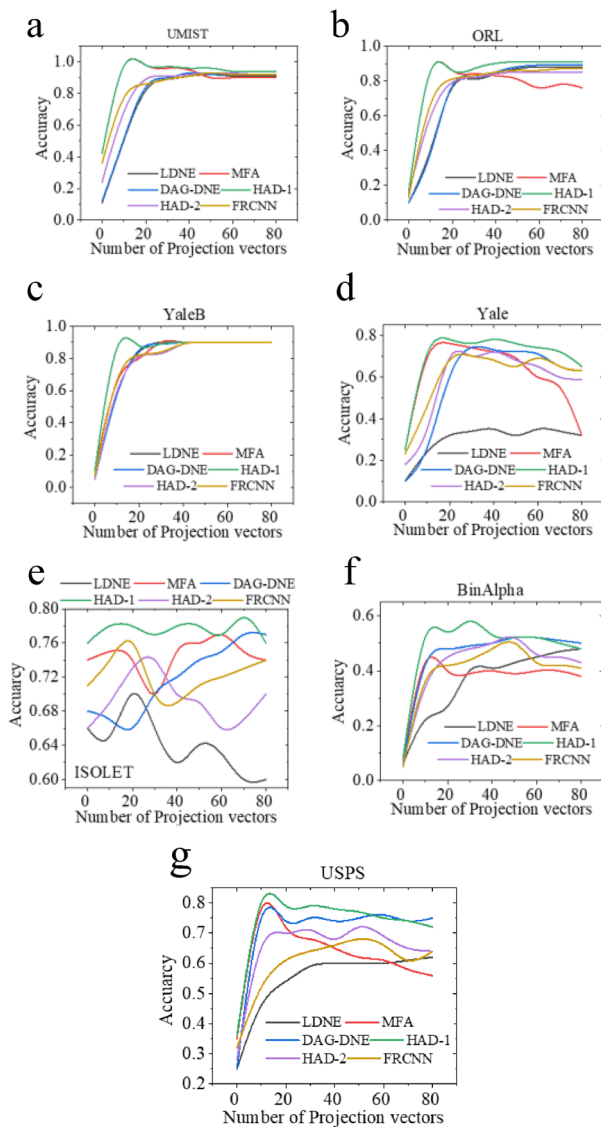


Fig. 5 Comparison of accuracies of various algorithms in different character feature recognition data sets

overall accuracy of the Yale is up to 78%, and the accuracy of other character feature recognition data sets can reach about 90%. The performance of HDA-1 on different data sets is higher, up to 94%, than other image feature extraction algorithms. The accuracy of FRCNN algorithm under the supervised learning also reaches 91% when the projection parameter is higher than 30. The accuracy of MFA on the UMIST, ORL, and Yale continues to decline with the increase of projection vector. LDNE has the worst performance in the Yale, because the data set involves more character expressions. The above results reveal that all algorithms have good recognition effect for different lighting conditions, character characteristics, and character postures, but have poor recognition effect for human expressions. The

HDA-1 under unsupervised learning has the best performance, and the image recognition accuracy basically maintained at 92% on different data sets.

Figure 5e–g shows the comparison results on accuracies of algorithms under different text feature recognition data sets. The figures disclose that the accuracies of the algorithms continuously increase with the increasing number of projections, which is similar to the case in character feature recognition data sets. Differences in different data sets are large. The accuracy in the BinAlpha when the projection vector is 30 maintains at a certain level; and the accuracy in the USPS is maintained at a certain level when the projection vector is 20. The accuracy of text recognition is relatively lower than that of the character feature recognition. Among them, the performance of HDA-1 algorithm on different data sets is higher, up to 79%. The accuracy of FRCNN algorithm under supervised learning is basically inferior to the accuracy of unsupervised learning. The LDNE has the worst performance in the ISOLET. The above results suggest that all algorithms have poor recognition of original text and handwriting. Among them, HDA-1 under unsupervised learning has the best performance, and the accuracy of image recognition is basically maintained at 70% on different data sets.

Figure 6a–d shows comparison results in recall of different algorithms under different character recognition data sets. The results are similar to accuracy. The recall in the Yale data set fluctuates greatly, and recalls of all algorithms reach stable values when the projection vector is 40. The unsupervised learning method is obviously better than the supervised learning FRCNN. HDA-1 has the best performance, with a maximum recall rate of 97%. LDNE has the worst performance in the unsupervised because the algorithm is assigned with the initial weight so that the system runs slower and the load is heavier, thereby lowering the overall image recognition rate. The above results suggest that it is advisable to extract features from the image by optimizing the sum of the intra-class and inter-class distances.

Figure 6e–g shows the comparison results in recall of various algorithms under different text feature recognition data sets. The results are similar to those of the accuracy, and the recall fluctuates greatly on each data set. Compared with the supervised learning method, the unsupervised learning method has better performance. HDA-1 has the best performance, and the recall rate is maintained at an average of 60%. Among them, LDNE in unsupervised learning has improved performance compared with character recognition. It may be because the image recognition of text converts more text into numbers, the way to assign weights is better. HDA-1 with the optimized sum of intra-class and inter-class distances has better performance than the algorithm without optimization. The above results indicate that the optimized sum of intra-class and inter-class distances can

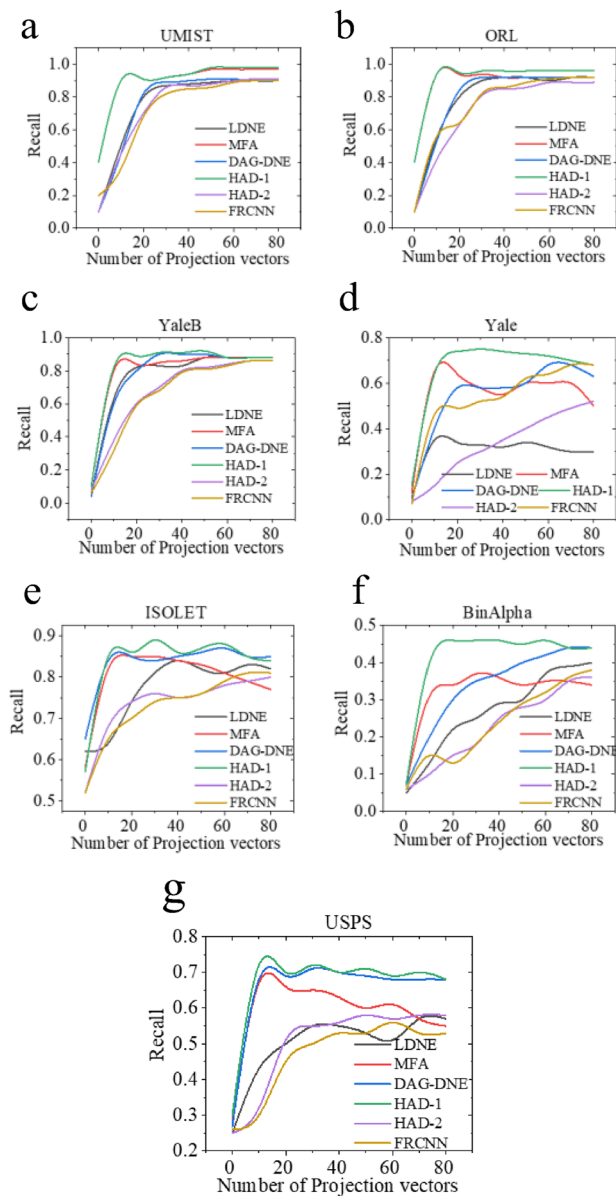


Fig. 6 Comparison on recall of various algorithms under different character and text feature recognition data sets

significantly improve the correct classification ability of image recognition.

Figure 7a–d illustrates the comparison result in mAP of various algorithms under different character recognition data sets. Being similar to the accuracy results, the Yale data set has large fluctuations and unsupervised learning method is better than the supervised learning FRCNN. HDA-1 has the best performance, and the mAP is maintained at 85%. The performance is greatly improved by 15.3% than HDA-2, 12.4% than LDE, and 8.9% than DAG-DNA. The above results also further prove the effectiveness of proposed HDA in the recognition and classification of character images.

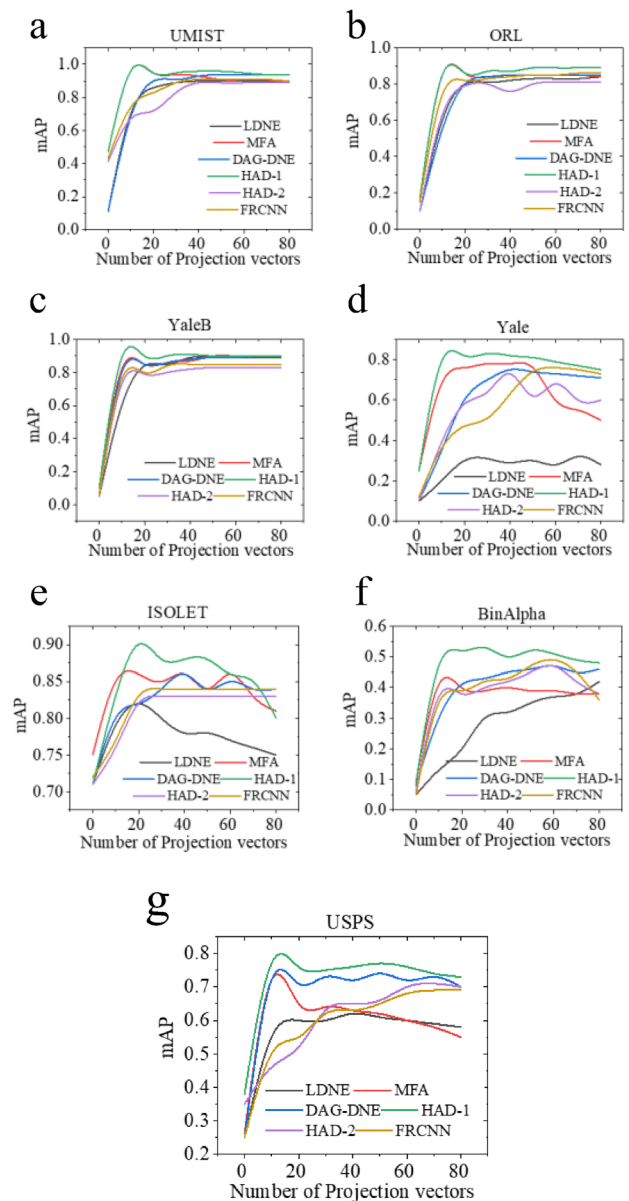


Fig. 7 Comparison on mAP of various algorithms under different character and text feature recognition data sets

Figure 7e–g illustrates the comparison results in mAP of the algorithms under different text feature recognition data sets. Being similar to the accuracy, the mAP on different data sets fluctuates greatly. Although the unsupervised learning method is better than the supervised learning FRCNN, its performance in the character feature recognition is improved. The best performance is realized in the HDA-1. The Map is maintained at around 75%. The performance is greatly improved by 5.4% than HDA-2, 7.8% higher than LDE, and 2.6% higher than DAG-DNE. The above results further prove the effectiveness of the proposed HDA algorithm in the image recognition and classification of text.

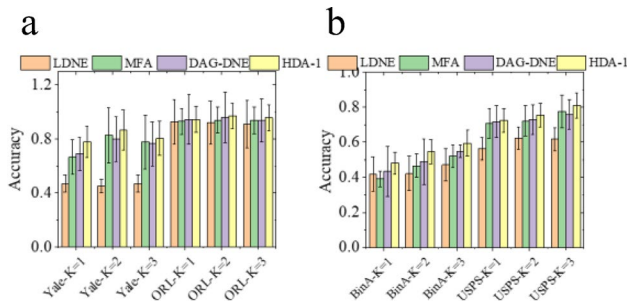


Fig. 8 Optimal parameter determinations of various algorithms under different data sets

3.2 Determination of optimal parameters for the image recognition system based on MEC and deep learning

Next, the parameters of the model affect the accuracy of the image recognition effect of the model under different data sets in the model. Therefore, four algorithms with better performance are selected and the differences in model performance at different adjacent distances are analysed. The specific results are shown in Fig. 8. The optimal parameter determination results of different algorithms under different data sets are given in Fig. 8. If the distance parameter K of two adjacent feature extraction points changes, and the difference in accuracies of various algorithms under different data sets is more obvious. Performance of the weak supervised learning method without distance optimization is much different from that with distance optimization, so FRCNN and HDA-2 are deleted. When the distance between the two feature extraction points becomes long, accuracy of the algorithm will first increase and then decrease. Under the character recognition data set Yale and ORL, LDNE has the lowest accuracy, which is maintained at an average of 65%; and the HDA-1 algorithm has the highest accuracy rate. The difference is even greater when $K=1$, and the highest accuracy can reach 98.25%. In the text feature recognition data set BinAlpha and USPS, LDNE has the lowest accuracy, maintaining an average of 55%; and HDA-1 has the highest accuracy. When $K=1$, the accuracy is maintained at 62%, and the difference is relatively significant. The above results show that when the distance parameter K of two adjacent feature extraction points is 2, HDA-1 of the performance has a better performance, and the highest accuracy rate can be 98.25%.

3.3 Analysis of IoT performance of image recognition system based on MEC and deep learning

Finally, the above algorithm is applied to the actual IoT system. The performance differences of proposed HAD, the

general feature extraction algorithm in the MEC, and the general feature extraction algorithm in the mobile cloud computing are analysed from the perspective of broadband and unlimited network response time. The response time of each framework under LTE and WIFI is given in Fig. 9. It reveals that the response time of different images under different system frameworks varies greatly. The response time of the WIFI and LTE is different from each other greatly by twice. The response time of WIFI is less than 1000 ms, and the recognition response time of LTE is 2000 ms. Based on comparison results of different IoT strategies, it can be known in Image 5 of the LTE that the average response time of the strategy not using MEC but using the general image feature extraction method is 2700 ms, which is different by 35.18% with the strategy using the MEC and the general image feature extraction method and by 37.03% with the strategy using the MEC and the HDA image feature extraction method. In Image 5 under the WIFI, the average response time of the strategy not using MEC but using the general image feature extraction is 720 ms, which is different by 29.16% with the strategy using MEC and the general image feature extraction and by 30.55% with the strategy using the MEC and the HDA image feature extraction method. The above results reveal that the strategy using MEC and HDA need less time on network performance of the IoT.

The broadband occupancy ratios under different IoT solution frameworks are given in Fig. 10 below. It indicates that the amount of data transmitted using the HDA is smaller because HDA requires less feature information under the same recognition rate, the overall occupation of broadband is relatively small. Thus, the calculation is reduced, the processing time is shorter, and the bandwidth occupancy is less. The application of MEC in image recognition can effectively reduce the transmission traffic uploaded to the core network, thus greatly reducing the bandwidth occupancy of the core

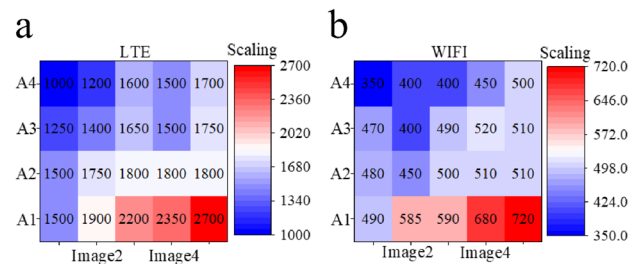


Fig. 9 Response time of various frameworks under the LTE and WIFI. Note: the algorithm not using MEC but using the general image feature extraction is called A1, the algorithm not using MEC but using the HDA image feature extraction method is called A2, the algorithm using the MEC and the general image feature extraction is called A3, and the algorithm using MEC and HDA image feature extraction method is called A4

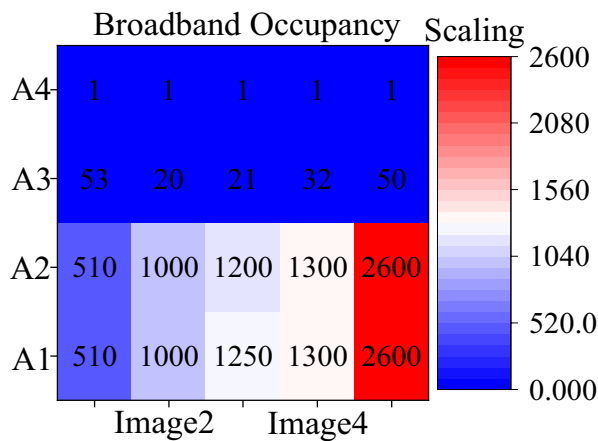


Fig. 10 Broadband occupancies under different IoT solutions

network. The above results all suggest that the MEC and HDA can occupy less bandwidth and have larger outputs in the IoT network performance.

4 Discussion

Firstly, images are performed with feature extraction by using various algorithms, including LDNE, MFA, DAG-DNA, HDA-1, HDA-2, and FRCNN. Of which, the accuracy, recall, and mAP of the MFA are weaker than those of HDA-1 (Figs. 5, 6, 7) in the recognition of text and character features. The reason may be that MFA is designed with two graphs to find the projection matrix, which represents the density of the image and the dispersion degree of the image, and the algorithm mainly selects the dimension of the subspace, so it requires strong empirical assumptions, leading to the loss of some important discriminative information [16]. Performance of LDNE algorithm is the worst (Figs. 5, 6, 7). LDNE assigns weights to inter-class and intra-class adjacent parts through two adjacent graphs to establish the relationship to maintain the balance. Thus, performance is naturally worse compared with other algorithms [17]. The DAG-DNA algorithm is significantly better than MFA and LDNE (Figs. 5, 6, 7). It constructs two adjacent graphs and uses the sum of intra-class distances and the sum of inter-class distances to process, which greatly improves the operating efficiency of algorithm; but such algorithm has too large distance in the subspace, resulting in insufficient optimization [18]. On the basis of the DAG-DNA, HDA-2 removes the sum of the inter-class distances, leading to a significant reduction in the performance of the algorithm (Figs. 5, 6, 7), which is consistent with reports in related literature [19]. The proposed HDA algorithm first minimizes the sum of intra-class distances, and then maximizes the sum of inter-class distances. The sums of intra-class distances

and inter-class distances are optimized in hierarchically, it is not biased to either party, so it has the best performance in the projection space, which is consistent with previous reports [20].

Second, the MEC is adopted to reduce the pressure of core network and shorten the response time, which has been shown in results under this article (Figs. 9, 10). The conventional cloud computing mode is to directly upload images to the cloud data centre for identification, and provide computing and storage functions remotely. However, the distance between cloud data centre and user is relatively long, resulting in longer transmission delay, so it cannot meet the user's real-time response demand for image recognition [21]. More importantly, uploading the image recognition task to the core network will put tremendous pressure on broadband and cause network congestion. The proposed hierarchical image recognition increases the way of edge computing, which has been proved in many articles [22, 23]. Edge computing is more suitable for real-time data analysis and processing, and is more secure and effective than the original conventional cloud computing [24]. Finally, an image recognition system based on MEC and deep learning is formed through the fusion of MEC and hierarchical algorithms, which greatly improves the overall performance of IoT, has high accuracy, can realize real-time image recognition and classification quickly and accurately, and achieve the up and down transmission through IoT.

5 Conclusion

Issues in real-time image recognition in the context of IoT are analysed. Using the HDA under unsupervised learning, the information of image uploaded by user can be extracted in the way of hierarchical features, and accuracy of image recognition is as high as 98% after optimizing the parameters. MEC implements distributed computing through edge servers and base stations to improve the performance of traditional wireless stations, effectively reduce the pressure of core network, shorten the response time, and greatly reduce the broadband occupancy ratio. Compared with the method not using MEC and HDA, the performance is improved by 37.03%. This research can achieve rapid processing of real-time images in the context of IoT, and the high accuracy of image recognition can provide theoretical ideas for related research in this field. However, due to the limited time, this article still has many shortcomings: firstly, the adopted supervised learning algorithm is still less than the experiment so that the advantages and performance of unsupervised learning can't be fully utilized; secondly, the current image recognition in the MEC environment is a kind of shallow learning feature extraction, and deep learning training takes longer time and requires larger data. In this article,

advantages of deep learning are not utilized fully to further improve the accuracy of the algorithm. In future, in-depth research can be performed to better apply it in real life and truly realize the life of the IoT.

References

1. Qi, F., Zhu, X., Mang, G., et al.: UAV network and IoT in the sky for future smart cities. *IEEE Netw.* **33**(2), 96–101 (2019)
2. Dai, H.-N., Zheng, Z., Zhang, Y.: Blockchain for Internet of Things: a survey. *IEEE Internet Things J.* **6**(5), 8076–8094 (2019)
3. Wang, Z.L.: Entropy theory of distributed energy for internet of things. *Nano Energy* **58**, 669–672 (2019)
4. Knieps, G.: Internet of Things, big data and the economics of networked vehicles. *Telecommun. Policy* **43**(2), 171–181 (2019)
5. Ahmad, K., Mohammad, O., Atieh, M., et al.: Enhanced performance and faster response using new IoT litetechnique. *Int. Arab J. Inf. Technol.* **16**(3A), 548–556 (2019)
6. Rahbari, D., Nickray, M.: Low-latency and energy-efficient scheduling in fog-based IoT applications. *Turk. J. Electr. Eng. Comput. Sci.* **27**(2), 1406–1427 (2019)
7. Khan, W.Z., Ahmed, E., Hakak, S., et al.: Edge computing: a survey. *Future Gener. Comput. Syst.* **97**, 219–2135 (2019)
8. Zhou, Z., Chen, X., Li, E., et al.: Edge intelligence: paving the last mile of artificial intelligence with edge computing. *Proc. IEEE* **107**(8), 1738–1762 (2019)
9. Aggarwal, S.L.P.: Data augmentation in dermatology image recognition using machine learning. *Skin Res. Technol.* **25**(6), 815–820 (2019)
10. Lahasan, B., Lutfi, S.L., San-Segundo, R.: A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression. *Artif. Intell. Rev.* **52**(2), 949–979 (2019)
11. Miao, Y., Wu, G., Li, M., et al.: Intelligent task prediction and computation offloading based on mobile-edge cloud computing. *Future Gener. Comput. Syst.* **102**, 925–931 (2020)
12. Yang, A., Yang, X., Wu, W., et al.: Research on feature extraction of tumor image based on convolutional neural network. *IEEE access* **7**, 24204–24213 (2019)
13. Ren, Z., Sun, Q., Wu, B., et al.: Learning latent low-rank and sparse embedding for robust image feature extraction. *IEEE Trans. Image Process.* **29**(1), 2094–2107 (2019)
14. Guo, Y., Cao, H., Bai, J., et al.: High efficient deep feature extraction and classification of spectral-spatial hyperspectral image using cross domain convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **12**(1), 345–356 (2019)
15. Ye, M., Ji, C., Chen, H., et al.: Residual deep PCA-based feature extraction for hyperspectral image classification. *Neural Comput. Appl.* 1–14 (2019)
16. Al-Khaleefa, A.S., Ahmad, M.R., Isa, A.M., et al.: MFA-OSELM algorithm for WIFI-based indoor positioning system. *Information* **10**(4), 146–153 (2019)
17. Cheng, C., Zhou, H., Chai, X., et al.: Adoption of image surface parameters under moving edge computing in the construction of mountain fire warning method. *PLoS ONE* **15**(5), 0232433–0232441 (2020)
18. Ding, C., Wang, S.: Appropriate points choosing for subspace learning over image classification. *J. Supercomput.* **75**(2), 688–703 (2019)
19. Liu, X., Zhao, H.: Hierarchical feature extraction based on discriminant analysis. *Appl. Intell.* **49**(7), 2780–2792 (2019)
20. Xu, X., Li, Y., Jin, Y.: Hierarchical discriminant feature learning for cross-modal face recognition. *Multimed. Tools Appl.* 1–20 (2019)
21. Xu, X., Zhang, X., Gao, H., et al.: BeCome: blockchain-enabled computation offloading for IoT in mobile edge computing. *IEEE Trans. Ind. Inf.* **16**(6), 4187–4195 (2019)
22. Li, J.Q., Han, Y.Q.: A hybrid multi-objective artificial bee colony algorithm for flexible task scheduling problems in cloud computing system. *Clust. Comput.* 1–17 (2019)
23. Abdullahi, M., Ngadi, M.A., Dishing, S.I., et al.: An efficient symbiotic organisms search algorithm with chaotic optimization strategy for multi-objective task scheduling problems in cloud computing environment. *J. Netw. Comput. Appl.* **133**, 60–74 (2019)
24. Wang, S., Zhao, Y., Xu, J., et al.: Edge server placement in mobile edge computing. *J. Parallel Distrib. Comput.* **127**, 160–168 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Tao Shen is a post-doctor in College of Design and Innovation, Tongji University. His research interest is mainly in the areas of computer-aided design and innovation design. He has published several research papers in scholarly journals in the above research areas and has participated in several conferences



Chan Gao is a lecturer in the Architecture Department of Huzhou University in Zhejiang province since 2015 and started doctoral study in Japan Advanced Institute of Science and Technology science 2020. Her research interest includes deep learning, computer-aided design, computer simulation and other scientific research work



Dawei Xu is a researcher in College of Computer Science, Qinghai Nationalities University. His research interests lie in the area of algorithm and visualization, ranging from theory to design and implementation. He has published several research papers in the above areas.