Retraction Notice

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of systematic manipulation of the publication process, including compromised peer review. The Editor and publisher no longer have confidence in the results and conclusions of the article.

JF and JT agreed with retraction. XL and YW either did not respond directly or could not be reached.

Face recognition technology in classroom environment based on ResNet neural network

Jian Fang,[a,b] Xiaomei Lin,[a,*] Jia Tian,[b] and Yue Wu[b]
[a]Changchun University of Technology, Changchun, China
[b]Jilin Engineering Normal University, Changchun, China

Abstract. With the rapid development of electronic computers and information technology, face recognition is widely used in fields such as enterprises, entertainment, information security, and daily life. However, the current face recognition technology is still relatively poor in distinguishing facial features, resulting in a low accuracy of face recognition, which cannot meet increasing application requirements. For this reason, it is necessary to develop more accurate face recognition technology. Residual neural network (ResNet) neural network is a deep residual learning convolutional neural network, which can be used to solve the degeneration problem (i.e., after adding more layers to the neural network, the performance drops rapidly) of deep-learning neural networks. We aim to study the specific principles, calculation methods, and characteristics of the ResNet neural network, analyze the complexity of the classroom environment, propose the use of ResNet network for face feature extraction, and use additive angular margin loss for deep face recognition (ArcFace) as the loss function of the ResNet network to improve the ResNet network. Because the ArcFace function has the advantages of high performance, easy programming, low complexity and high training efficiency. We use this scheme to conduct field tests of face recognition on many students in the classroom. The test results show that this scheme enhances the discrimination of facial features and improves the accuracy of face recognition. In the case of different numbers of people, facial defects, and strong light exposure, the system can still detect and recognize faces stably. In the case of facial defects, the recognition accuracy rate is still 63.3%, and the recognition accuracy rate is still more than 60% under the illumination of strong light, and the recognition accuracy rate under correct conditions is more than 70%. © *2022 SPIE and IS&T* [DOI: 10.1117/1.JEI.31.5.051421]

Keywords: face recognition; ResNet neural network; residual learning; ArcFace; classroom environment.

Paper 220105SS received Mar. 18, 2022; accepted for publication Jul. 5, 2022; published online Sep. 14, 2022.

Retracted

1 Introduction

1.1 *Background Significance*

With the rapid development of society, there are increasing scenarios where face recognition is applied, such as in face payment, electronic information authentication, daily attendance, residential security, and other scenarios. Many schools have also begun to implement face recognition for student attendance. School attendance in class has always been one of the most important metrics for students. For a long time, class attendance was mainly done by humans. However, whether it is the teacher's roll call or paper attendance, it not only wastes class time but also fails to properly supervise students. In some previous studies, students can complete classroom attendance in certain areas of the education building through mobile facial recognition. However, this method also caused some students to perform virtual positioning, signing, etc., which could not prevent them from using a series of deceptive attendance behaviors in the classroom. The current improved design is to use a camera at the entrance of the classroom to recognize the faces of class students and complete class attendance. However, using this method

*Address all correspondence to Xiaomei Lin, linxiaomei@ccut.edu.cn

1017-9909/2022/$28.00 © 2022 SPIE and IS&T

alone will cause many complex problems in real life. For example, the check-in channel is congested, and changes in light cannot be correctly identified. Therefore, it is necessary to further improve the face recognition system in the classroom environment.

The rapid rise of face recognition technology based on the development of deep learning has developed rapidly in the field of image processing, and has achieved very good results in the field of image recognition. Before the advent of deep learning, support vector machine (SVM is a supervised learning model, usually used for pattern recognition, classification, and regression analysis) algorithm was the best way to realize image recognition. With the advent of convolutional neural networks (CNNs), weight sharing, computational complexity is greatly reduced, but the accuracy of image recognition is improved. CNN is a convolutional neural network. CNN can avoid the complicated feature extraction steps in machine learning in the past and directly input image data into the neural network. Due to the high accuracy of CNN, it has received widespread attention and quickly replaced the position of SVM. The current face recognition technology is divided into two parts: face feature extraction and face recognition based on features. The first is facial feature extraction. The RetinaFace face detection algorithm is used to obtain facial features, and the facial features are extracted in the neural network through the residual neural network (ResNet) residual network for deep learning, and finally face recognition is completed. But in fact, most of the commonly used face recognition methods use Softmax loss as the network layer, and fail to consider the structure of the loss function. The Softmax loss needs to consider whether the samples can be classified correctly, and the problem of expanding the distance between heterogeneous samples and isomorphic samples has a lot of room for optimization. Therefore, this paper points out the main characteristics of face information data cleaning during the change from Softmax Arcface to Arcface, improves the ResNet network structure, and finally makes the experimental model more suitable for learning facial features.

1.2 *Related Work*

Face recognition has always been a hot topic in the field of pattern recognition, in which feature extraction and classification play an important role. However, CNN and local binary pattern (LBP) (local binary mode, its initial function is to assist the local contrast of the image) can only extract one feature of a facial image and cannot choose the best classifier. Aiming at the problem of classifier parameter optimization, Wu and Jiang proposed two SVM structures optimized based on the artificial bee colony algorithm to classify the features of CNN and LBP, respectively. The clear experimental results in the Olivetti-Oracle Research Lab (ORL) (Olivetti Institute) and Face Recognition Technology (FERET, 1993, the FERET face database established by the US Department of Defense Advanced Research Projects Agency and the US Army Research and Experimental Project Team, used to evaluate the performance of face recognition algorithms) databases show the superiority of the method.[1] Facial expressions can reflect a person's emotions as well. If facial expressions can be recognized well, it will have a positive impact on understanding the reasoning of the mind and the thoughts of each other's minds. Kang stated that artificial neural networks are one of the tools of modern image text recognition systems, including handwritten images, and he presented the results of a computational experiment aimed at analyzing two artificial neural networks with different architectures and parameters (ANN) for the recognition quality of handwritten digits.[2] Ahdid et al. proposed two feature extraction methods for two-dimensional (2D) face recognition and three well-known classification techniques: Nerves (NN), K Nearest Neighbors (KNN), and SVM for comparison. To test its method and evaluate its performance, Ahdid et al. conducted a series of experiments using a 2D facial image database (ORL and Yale). The results show that, compared with Euclidean distance (ED), using graphic drawing (GD) to extract image features is computationally more effective.[3] The development of computer technology has led to the development of face recognition technology. Today, with the help of computer and network technology, face recognition technology has been successfully applied in many fields. Zhi and Liu establishes an effective face recognition model based on principal component analysis, genetic algorithm, and SVM. It uses principal component analysis to reduce the number of features, uses genetic algorithm to formulate search strategies, and uses SVMs to optimize and achieve classification. Zhi and Liu used the face database of the Institute of Engineering of the Chinese Academy of

Retracted

Sciences to conduct simulation experiments and found that face recognition can be implemented efficiently, with an accuracy rate of up to 99%.[4] Since the human face is a complex multi-dimensional visual model, it is difficult to develop a computational model to recognize it. Deshpande and Ravishankar proposed a face recognition method based on image features. The method proposed by Deshpande and Ravishankar is implemented in two phases. The first step is to use the Viola Jones algorithm to detect faces in the image. The next step is to use the fusion of principal component analysis and feedforward neural networks to recognize the faces detected in the image. The method proposed by Deshpande and Ravishankar uses BioID-Face-Database as the standard image database, and compares the performance of the method with existing methods. The results show that its method has better recognition accuracy.[5] Face recognition based on a small number of training samples is a challenging task. In daily applications, it may not be possible to obtain enough training samples, and most of the obtained training samples are in various lighting and postures. Insufficient training samples cannot effectively express various face situations, so it is a difficult task to improve the face recognition rate in the case of insufficient training samples. Zhao et al. used face pose pre-recognition (FPPR) model and double dictionary sparse representation classification (DD-SRC) for face recognition. The FPPR model is based on the geometric features of the face and machine learning, and the test samples are divided into full face and profile face. Different postures in a single dictionary affect each other, resulting in a low face recognition rate. DD-SRC contains two dictionaries, a full-face dictionary, and a contour dictionary, which can reduce interference. After FPPR, the samples are processed by DD-SRC, and the most similar one is found among the training samples. Zhao et al.'s experimental results show the performance of the algorithm on ORL and FERET databases, and also reflect the comparison with SRC (classification based on sparse expression), linear regression classification, and two phase test sample sparse representation (sparse representation of two-stage test samples).[6] The face symbol represents the unique characteristics of a human face. It can be used to reduce lengthy information and reduce computational complexity. However, the feature points extracted in each frame of the video are irregular and need to be aligned. Lin et al. proposed a new method based on facial landmarks and machine learning. Lin et al. aligns the feature data with the public coordinate system and uses the robust AdaBoost algorithm for classification. The results show that experiments using the public Honda/UCSD database have proved better performance than some advanced methods. Experiments in the Yale database show the sensitivity and specificity of the proposed method. The method proposed by Lin J can improve the recognition performance based on image sets.[7] Although, the above research can promote the development of face recognition technology to a certain extent, it has not been promoted because of the disadvantages of slow calculation efficiency and high development cost. And the content of this article has gathered the characteristics of high calculation accuracy, high efficiency, and low cost.

1.3 *Innovation*

(1) Using the outstanding performance of the ResNet neural network to extract features, it is easier to perform face recognition. (2) ResNet neural network can solve the degeneration phenomenon brought by deep-learning network, making face recognition more accurate. (3) Analyzing the complexity of the classroom environment and matching the characteristics of the ResNet neural network can more effectively solve the problems of face recognition in the classroom environment. (4) Comparing the traditional face recognition method with the method in this article, it can better reflect the advantages of the face recognition technology in this article.

2 Face Recognition Method Based on Residual Neural Network

2.1 *Neural Network Face Recognition*

Nowadays, face detection technology has been applied to various fields of society, such as security, entertainment, life, etc., cannot be separated from face detection technology. Table 1 shows the proportion of current face recognition applications in various fields (data comes from public

Retracted

Table 1 Usage of face recognition technology.

Field of use	Proportion (%)
Finance	20
Security	30
Attendance and access control	42
Other	8

data collation). It can be seen from Table 1 that face recognition currently accounts for the largest proportion of applications in attendance and access control. The face recognition in the classroom environment in this article is also a study on classroom attendance.[8]

The most prominent of the current facial recognition technology neural network algorithms is the CNN, and the ResNet neural network in this article is one of them. CNNs have powerful representation learning capabilities, and can better play their characteristics in face recognition.

1. The structure of the CNN. Its structure consists of multiple layers. In a complete network, when an image is input, it first passes through the convolutional layer, then puts the output of this layer into the pooling layer for processing, and finally the output is processed by the fully connected layer. This process has a batch normalization layer and an activation function.

2.1.1 *Convolutional layer*

The role of convolution is to extract features. This layer is also the core of the entire network. This layer usually performs a lot of calculations. At the same time, the amount of calculation on this layer is also the largest in the network. Among them, the convolution kernel of the convolution layer is used to extract features. The convolution kernel is actually a sliding window with some parameters, also called filters. The detailed process is shown in Fig. 1.

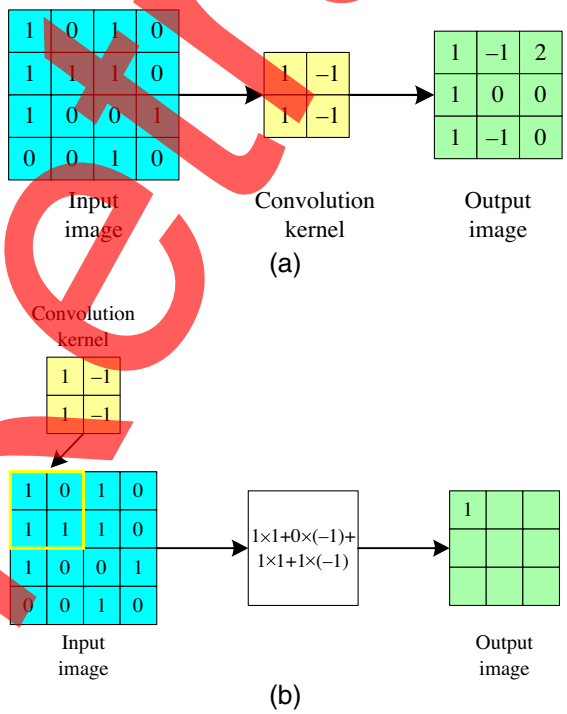

Fig. 1 (a) Feature extraction process by convolution kernel. (b) Convolutional layer for calculation.

Retracted

Denoting the input graph as M, J as the convolution kernel, and the output graph as O, then the calculation formula is

$$O = \sum_{x=1}^{a} \sum_{y=1}^{b} M_{x,y} J_{x,y}. \tag{1}$$

In the formula: x and y represent the coordinates and a and b represent the length and width of the figure, respectively.

However, the above flowchart is only a 2D simple algorithm. In the convolution process of a three-dimensional (3D) image, adding a convolution kernel will generate a feature map. If the step size is set, when the convolution kernel slides, some positions of the convolution kernel may not slide, and some edge information will be lost at this time. This problem can be solved by setting the fill value. Assuming that the input form of a 3D graphic is $L \times W \times D$, i.e., L, W, and D are length, width, and depth, respectively; the filling value is T; and J is the size of the convolution kernel, then:

$$W_O = \frac{(W + 2T - J)}{B} + 1, \tag{2}$$

$$L_O = \frac{(L + 2T - J)}{B} + 1, \tag{3}$$

$$D = S. \tag{4}$$

In the formula: B represents the moving step size, and S represents the number of convolution kernels.

2.1.2 *Pooling layer*

The pooling layer has no parameters, and the calculation method is basically the same as that of the convolutional layer, so the calculation process is relatively simple. Generally, there are two ways of average pooling and maximum pooling. The maximum pooling calculation method is to take the maximum value of a specific area in the input image, and the average pooling is to add the values of the specific area and divide by the number of blocks, as shown in Fig. 2.

2.1.3 *Fully connected layer*

After convolution and pooling, the input image enters the last layer for processing, that is, the fully connected layer, so the fully connected layer actually plays the role of connection. This layer can be expressed as

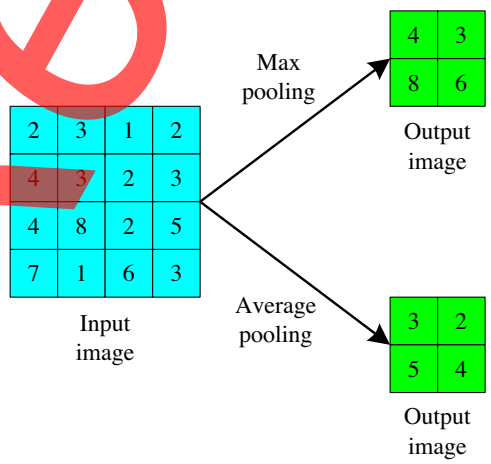

Fig. 2 Pooling layer pooling method.

Retracted

$$\mathrm{TL}(x) = f(W^N x + v). \tag{5}$$

In the formula: $TL(x)$ represents the output value of the fully connected layer, f represents the activation function, W^N represents the graphics matrix, and v represents the displacement offset.

The activation function is a function that runs on the neurons of the artificial neural network and is responsible for mapping the input of the neuron to the output. Common activation functions are sigmoid function, hyperbolic function and ReLu function. The sigmoid function is expressed as follows:

$$f(x) = \frac{1}{e^{-x} + 1}. \tag{6}$$

The hyperbolic function is expressed as follows:

$$f(x) = \frac{1 - e^{-2x}}{e^{-2x} + 1}. \tag{7}$$

The ReLu function is as follows:

$$f(x) = \max\{0, x\}. \tag{8}$$

2.1.4 *Batch normalization layer*

The batch normalization layer is added to speed up the training convergence speed and accuracy of the neural network. The batch normalization layer normalizes the data of each layer of the network to make the new distribution more in line with the actual distribution of the data. The expression is

$$O_k = \frac{I_k - e(I_k)}{\sqrt{\Delta(I_k)}}. \tag{9}$$

In the formula, O_k represents the output of this layer, I_k represents the input of this layer, $e(I_k)$ represents the expected prediction, and $\Delta(I_k)$ represents the variance of the input.

At the end, the calculation accuracy formula is used to evaluate the accuracy of recognition, and the formula is as follows:

$$\overline{G} = \frac{\sum_{x=1}^{m} G_k}{m}. \tag{10}$$

In the formula: $\overline{G}$ represents the average accuracy of the total calculation sample, m represents the sample type, and $\sum_{x=1}^{m} G_k$ represents the accuracy of the total sample

$$G = \frac{\mathrm{RS}}{\mathrm{RS} + \mathrm{ES}}. \tag{11}$$

In the formula: G is the calculation accuracy, RS is the sample that is calculated correctly, and ES is the sample that is calculated incorrectly.

Now compared with the previous face detection technology, there has been a qualitative leap. But under the influence of facial expression changes, shadows, and light factors, the discrimination of face detection is easily affected by some functions. And through facial feature information extraction, calculation, classification, and identification, the influence of the above factors can be reduced, the accuracy of face detection can be improved, and the efficiency of face recognition can be improved.

Retracted

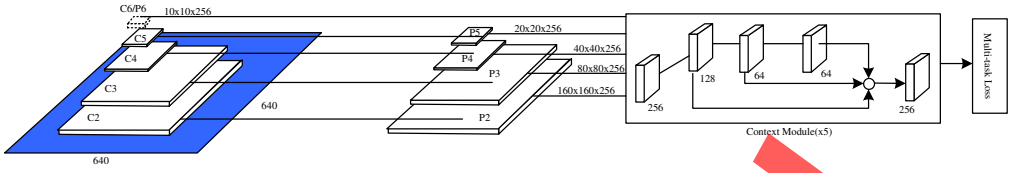

Fig. 3 RetinaFace network pyramid diagram.

2.2 *Face Feature Extraction*

2.2.1 *RetinaFace face detection algorithm*

The RetinaFace algorithm is the most powerful open source face detection algorithm. The RetinaFace algorithm uses a feature pyramid network structure, as shown in Fig. 3. The most prominent feature of the RetinaFace algorithm is the unique design building module based on the upper and lower structure of the feature pyramid.[9] The algorithm will directly derive the category probability and coordinate position of the target, so that the final detection result can be obtained after only one detection, and the detection speed is very fast.

2.2.2 *RetinaFace feature extraction network*

The RetinaFace network implements sampling prediction in each different sampling layer and compares them by predicting the offset of the anchors through the same anchors. And RetinaFace network has a unique key point detection, RetinaFace network structure can be divided into four branches. The first branch structure is to detect whether the target has face information, and the second branch is to draw the offset of the face frame. The third branch is to detect the key point offset on the face, and the fourth branch is dense point regression and self-supervision.[10] Multi-stage prediction is used for down-sampling prediction. Specifically, there are five kinds of sampling, as shown in Table 2. In addition, the environment module (4, 8, 16, 32, 64) and anchors (160, 160) (80, 80), (40, 40), and (20, 20), (10, 10) are included. Each point has three kinds of anchors, so the total anchors are 102,300 points, and about three-quarters of the anchors come from P2.

There are anchors with relatively accurate scales on the pyramid levels (P2) to (P6), and usually small anchors under the plane are used to detect faces that are difficult to detect. However, doing so requires more and more time and space, which will also increase the risk of false reports. Then sort the negative anchors according to the lost values, and filter out the anchors with the most losses, so that the ratio of negative samples to positive samples is at least 3:1. Crop square patches are randomly implemented in the original image and adjusted to a 640×640 image to generate a more powerful training set. Random cropping can also be carried out, through random horizontal flipping with half probability and light and shadow, color extraction, increase training data, and build a powerful network structure. The RetinaFace network can only be used to detect human faces, which greatly reduces the impact caused by facial expressions, object occlusion, light, and shadow, and improves the accuracy of face detection. However, most of the face recognition methods in the experiment use Softmax loss as the

Table 2 RetinaFace sampling analysis chart.

Anchors	Context module	Offset
P2 (160 × 160 × 256)	4	16, 20.16, 25.40
P3 (80 × 80 × 256)	8	32, 40.32, 50.80
P4 (40 × 40 × 256)	16	64, 80.63, 101.59
P5 (20 × 20 × 256)	32	128, 161.26, 203.19
P6 (10 × 10 × 256)	64	256, 322.54, 406.37

Retracted

network layer, failing to consider the structure of the loss function, and failing to consider the situation that Softmax cannot successfully classify samples. As a result, data such as heterogeneous samples and homogeneous samples are lost, so the ResNet neural network is introduced to improve it.[11,12]

2.3 *Residual Neural Network*

After the RetinaFace detection algorithm extracts the face features, because RetinaFace itself only detects the face, it cannot achieve the face recognition function. It is necessary to introduce the ResNet neural network to recognize the face information after extracting the facial features, and complete the remaining part of the face recognition experiment. To improve the accuracy of face recognition, the easy optimization feature of the ResNet network is borrowed. Using ResNet residual network built-in residual block jump connection method, the problem of gradient loss caused by increasing the depth of the deep neural network is solved. And the structure of ResNet greatly speeds up the training speed of neural networks. The specific manifestation is that the neural network of the first layer can learn the remaining output of the previous network, instead of learning the entire output.[13,14]

Figure 4 shows a schematic diagram of the first-layer algorithm structure of the ResNet network, where: $F(x)$ represents the residual and $H(x)$ represents the mapping output (network output). According to the schematic diagram:

$$F(x) = H(x) - x. \tag{12}$$

According to the multilayer neural network, any function can be fitted theoretically, so some layers can be used to fit the function. The method of fitting is simpler for directly fitting the network output or directly fitting the residual function, and it is easier to learn.[15]

2.3.1 *Residual neural network structure*

Figure 5 shows a schematic diagram of the two residual structures given in this article. The ResNet network structure will use two residual modules. The residual network structure in Fig. 5(a) has fewer layers and is a residual module that connects two $3 * 3$ convolutional networks together. The residual network structure in Fig. 5(b) has more layers of residual modules that connect three $1 * 1$, $3 * 3$, and $1 * 1$ convolutional networks.

Figure 5(b) residual error structure should be used in the deep network, which can reduce the network parameters and the amount of calculation. Similarly, input a feature matrix with a channel of 256. If the left-side residual structure is used, ~1,170,648 parameters are required, but if the right-side residual structure is used, only 69,632 parameters are required. When the deep network construction is obvious, it is more appropriate to use the residual structure on the right.

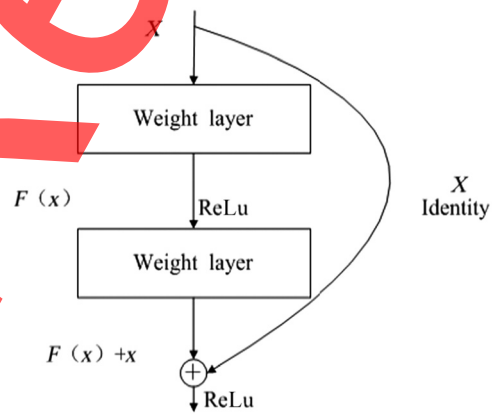

Fig. 4 Schematic diagram of the first layer of neural network.

Retracted

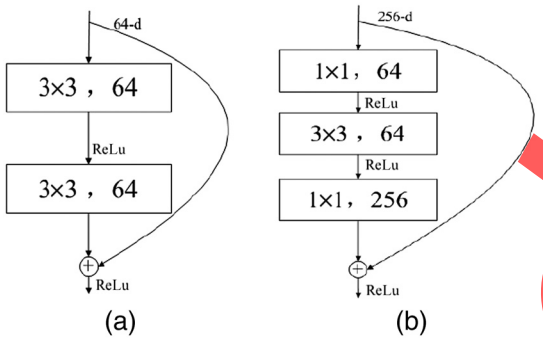

Fig. 5 (a) Schematic diagram of the (a) two- and (b) three-layer residual network structures.

The residual unit can be expressed as

$$y_1 = h(x_l) + F(x_l, w_1), \tag{13}$$

$$x_{l+1} = f(y_l). \tag{14}$$

Among them, x_1 represents the input of the L residual unit, and x_{1+1} is the output of the L residual unit. It is worthy to note that each residual unit usually contains a multi-layer structure. F is the residual function, $h(x_1) = x_1$ is the identity mapping, and F is the ReLU activation function. According to the above formula, we can get the learning characteristics of shallow 1 to deep L as follows:

$$\frac{\partial \text{loss}}{\partial x_l} = \frac{\partial \text{loss}}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} = \frac{\partial \text{loss}}{\partial x_L} \cdot \left(1 + \frac{\partial}{\partial x_L} \sum_{i=1}^{L-1} F(x_i, w_i)\right). \tag{15}$$

The first factor represents the gradient of the loss function to L. The 1 in parentheses indicates that the short-circuit mechanism can propagate the gradient without loss. The remaining gradients except for the L gradient need to pass through the weights layer, and the gradient is not directly transmitted. The residual gradient is not coincidentally −1, even if it is relatively small, the existence of 1 will not cause the gradient to disappear, so the residual learning will be easier.[16,17]

The ResNet neural network has different network layers. The more commonly used ones are 50-layer, 101-layer, and 152-layer, which are formed by stacking residual modules, as shown in Table 3.

Table 3 ResNet network layer.

Layer	ResNet50	ResNet101	ResNet152
Conv1	—	Convolutional layer: 7 × 7, 64, step size 2	—
	—	Maximum pooling layer: 3 × 3, step size 2	—
Conv2_x	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
Conv3_x	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$
Conv4_x	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
Conv5_x	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
Output	—	Average pooling layer, 1000-d fc, softmax	—

Retracted

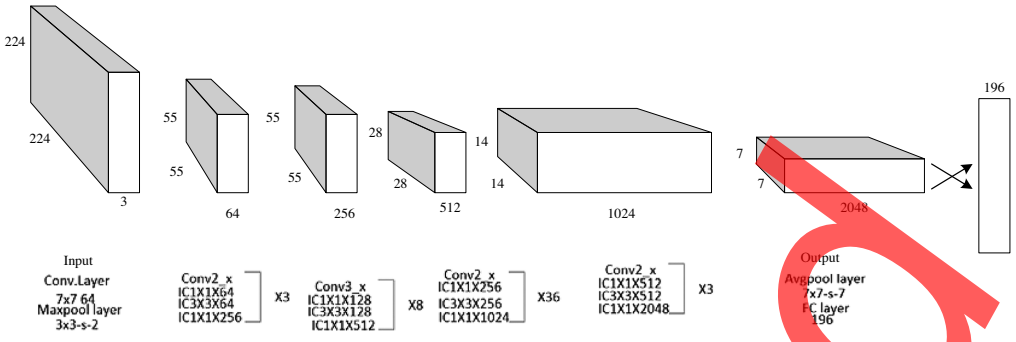

Fig. 6 ResNet input and output network structure diagram.

ResNet is composed of residual network units, and its core lies in identity mapping. The input and output relationship of the residual unit can be obtained by Eq. (10)

$$H(x) = F(x) + x. \tag{16}$$

The network structure is shown in Fig. 6, which is composed of residual blocks into four parts: $[1 \times 164, 3 \times 364, 1 \times 1256] \times 3[1 \times 1128, 3 \times 3128, 1 \times 1512] \times 8$, $[1 \times 1256, 3 \times 3256, 1 \times 11024] \times 36$ and $[1 \times 1512, 3 \times 3512, 1 \times 12048] \times 3$. Finally, the activation function of the Conv5_3 layer is set to Tanh. The 7×7 average pool of the output layer and Softmax are fully connected to the classification layer. Among them, Conv3_1, Conv4_1, and Conv5_1 are the down-sampling layers.

2.3.2 *Residual neural network using ArcFace loss function*

Facial feature extraction usually uses Softmax loss as the network layer. Experiments show that Softmax loss considers whether it can classify samples correctly, and expands the distance between heterogeneous samples and narrow similar samples. The distance between the classes of the problem has a lot of room for optimization. Therefore, the change process from Softmax Arcface to Arcface is discussed, and the main characteristics of data cleaning are pointed out, and the Resnet network structure is improved to make it more suitable for learning facial features. The flow of additive angular margin loss for deep face recognition (ArcFace) from input to output is shown in Fig. 7.

Assuming that the sample classification number is n, the dimension of the input data x is d, and the dimension of the model weight w is DXN. The first step is to normalize the sample x and the weight w, and obtain a $1\times n$-dimensional fully connected output after conversion. The second step is to use a network connection to output TargetLogit. The third step is to multiply the obtained Target Logit by the normalized parameter s, and finally Prob is calculated by the Softmax probability.[18]

At present, the networks with better results in extracting facial features include SphereFace, CosineFace, and ArcFace. In ArchFace, the classification limit is directly maximized in the angular space, while CosineFace maximizes the classification limit in the cosine space. However, the loss functions of SphereFace, CosineFace, and ArcFace are all modified on the basis of the traditional Softmax loss. The loss function formula is as follows:

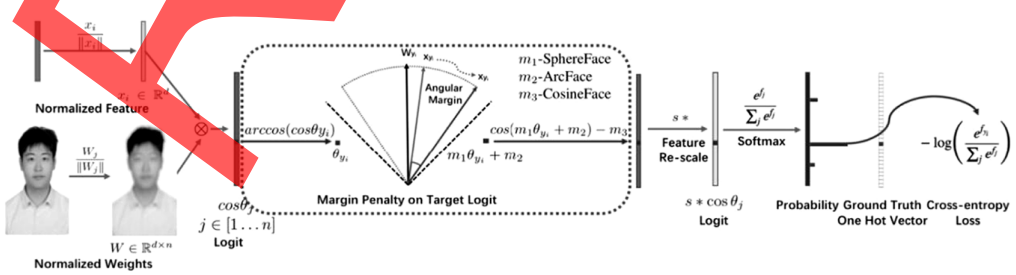

Fig. 7 ArcFace input and output flow chart.

Retracted

$$L = -\frac{1}{m}\sum_{i=1}^{m} \log \frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{w_{y_i}^T x_i + b_{y_i}}}. \tag{17}$$

This is the traditional Softmax formula. $w_{y_i}^T x_i + b_{y_i}$ represents the output of the fully connected layer. In the process of reducing the loss of L_s, the proportion of $w_{y_i}^T x_i + b_{y_i}$ must be increased so that more samples of this type fall within the decision boundary of this type.

From Softmax Loss to ArcFace Loss, scholars have made many improvements to the loss function. The ArcFace loss function is as follows:

$$L = -\frac{1}{m}\sum_{i=1}^{m} \log \frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}}. \tag{18}$$

Among:

$$w_j = \frac{w_j}{\|w_j\|}, \quad x_i = \frac{x_i}{\|x_i\|}, \quad \cos\theta_j = w_J^T x_i. \tag{19}$$

Angular margin in ArcFace is to directly maximize the classification limit in the angular space, it is the minimum distance of the hyper spherical surface, not the distance between the two feature points directly connected. The experimental results show that the characteristics of the designed loss function network learning are more obvious, which greatly improves the recognition accuracy.[19,20]

3 Face Recognition Test in Classroom Environment Based on Residual Neural Network

3.1 *Using Arcface to Recognize the Detected Face*

To perform face recognition, one first needs to extract the features of the face image, improve the robustness of network learning coding, and design a good loss function; the use of ArcFace loss function can effectively improve the robustness (refers to the control system maintains some other performance characteristics under a certain structure or size parameter perturbation) of learning coding. After the feature extraction is completed, import the trained face data model, and then compare. This paper adopts the method of comparing the distance of cosine similarity, and the cosine value of the angle between two vectors in the vector space is used as a measure of the difference between two individuals. The closer the value is to 1, the more accurate the face recognition is. The cosine similarity value measures the similarity of two variables in various directions (attributes). If the similarity exceeds a certain value, the recognition is successful, and a face recognition experiment is completed.[21]

The formula is

$$w_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| * |N(v)|}}. \tag{20}$$

After the similarity is obtained, the threshold can be set for face recognition. Then get the effect map of the face recognition experiment of ArcFace, compare the two detected faces, it can get the similarity of the face map, and compare it with the picture library, so as to get the face recognition result in the single-person recognition experiment. For the complete face recognition process, first use the ResNet network to train the local face, then use the RetinaFace algorithm to extract the face features, and then use ArcFace to compare the detected face area with the local face.[22–25]

Figure 8 shows the actual effect of face recognition with or without occlusion obtained in this experiment. The English letters indicate the abbreviation of the recognized person's name, and the numbers indicate the cosine value of the test. Figure 8(a) shows face recognition under normal conditions, and the cosine value is about 0.716, which has a high degree of recognition.

Retracted

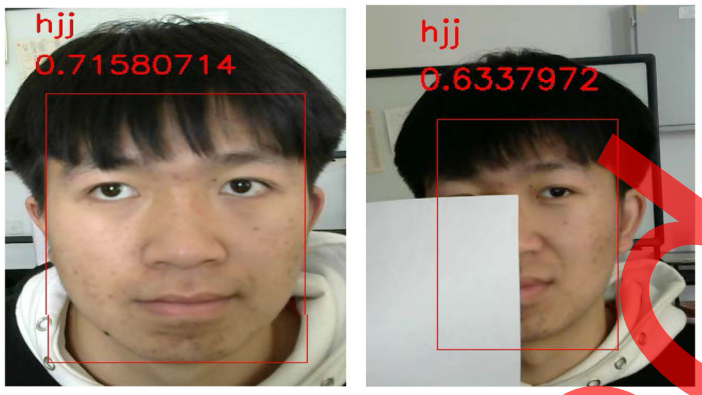

Fig. 8 Face recognition (a) without occlusion and (b) when occluded.

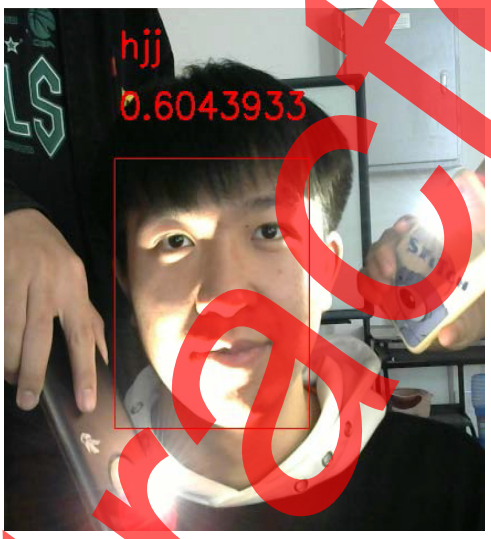

Fig. 9 Face recognition with changing light intensity.

Figure 8(b) shows face recognition under occlusion, the cosine value is about 0.638. The relative recognition degree will be lower, but it will not affect the correct recognition.

Figure 9 shows face recognition under the condition of changing the light intensity. It can be seen that the recognized cosine value is about 0.604, which is a bit lower than the previous two recognition effects, but it can also be correctly recognized.

The single-person recognition experiment cannot fully represent the system, so this paper has conducted a test of multi-person detection. Figure 10 shows the face recognition test in the case of multiple people. In Fig. 10(a), a face recognition test was performed on 32 students in the class. In Fig. 10(a), it can be seen that all of the 32 students tested were detected. In Fig. 10(b), we can see that all the detected students have been correctly identified.

3.2 Experimental Results and Analysis

After many experiments, it is proved that the face recognition technology described in this article can stably extract facial features and compare them based on facial features. In the single-person recognition experiment and the multi-person recognition experiment, it was found that the system was running stably, and there was no recognition error or recognition failure. But the single-person experiment can only reflect part of the experimental results, so this face recognition system is tested in other environments. Due to the improved ArcFace loss function, the system can operate stably in a complex environment. After changing the environmental

Retracted

(a)

(b)

Fig. 10 Face recognition (a) with multiple people and (b) accuracy in the case of multiple people.

Table 4 Analysis of the accuracy of face detection experiments.

Test scenario	The number of tests changed		Whether have facial occlusion		Changes in light intensity	
Test project	Single	Multi-player	No	Yes	Normal light	Blaze
Test times	200	200	200	200	200	200
Number of successes	143	141	143	126	143	121
Accuracy rate (%)	71.5	70.6	71.5	63.3	71.5	60.4

information, the face recognition system still operates normally, and the recognition accuracy rate is high. To further understand the actual situation of the system, this paper conducted 200 sets of face recognition tests on different test people, with or without facial defects, and different light intensities. The test results are shown in Table 4.[26,27]

It can be seen from Table 4 that the system can still detect and recognize faces stably when the number of people is different, the face is illuminated by defects, or strong light. In the case of facial defects, the recognition accuracy rate is still 63.3%, and the recognition accuracy rate is still more than 60% under the illumination of strong light, and the recognition accuracy rate under correct conditions is more than 70%.

4 Discussion

The face recognition in this paper takes advantage of the efficient feature extraction of the improved ResNet network, and combines the ArcFace loss function to make up for the large

Retracted

recognition error and low detection effect of traditional recognition. However, the experiment in this article still has some shortcomings:

1. In the classroom environment, the combination of image enhancement algorithm and target detection algorithm is actually higher than the simple ArcFace frame recognition rate, which reduces the detection miss rate.
2. The improved algorithm is definitely more efficient than the original algorithm, but the system also has some defects. For example, a non-human target that is prone to misidentification may be selected. Therefore, in future research, we will try to improve the algorithm to reduce the possibility of error, integrate the face recognition channel and access control system of the student dormitory built before the school, and obtain complete student face database data. Obtaining the prerequisites of the school education management system, such as first obtaining the timetable data of each class and the detailed list of class students, which can improve and perfect the entire system and do better for future research.

5 Conclusions

Through many experimental tests, the face recognition technology based on ResNet neural network, in the case of facial defects, the recognition accuracy rate is still 63.3%, and the recognition accuracy rate is still more than 60% under the illumination of strong light, and the recognition accuracy rate under correct conditions is more than 70%. It can realize face recognition efficiently and accurately, and achieve good facial feature extraction. Then the facial feature information is compared with the face trained locally on the ResNet network, and the comparison result is accurately obtained to achieve the purpose of face recognition, and the recognition rate is high and the recognition speed is fast. The face recognition system solves the problem of loss of information due to the Softmax loss function, so that the experimental model achieves a high-efficiency recognition state. And the face recognition technology in this article has great value for continuous development. For example, adding facial expression recognition, age recognition and other image processing fields in the current development stage on the basis of face recognition. Therefore, the face recognition studied in this paper lays the foundation for future development.

Retracted

Acknowledgments

This work was supported by the Science and Technology Development Plan Project of Jilin Province, the project name is Educational Robot Patent Information Analysis and Strategic Research (Grant No. 20190802025ZG). This work was also supported by Education Robot Innovation team of Jilin Engineering Normal University.

References

1. Y. Wu and M. Jiang, "Face recognition system based on CNN and LBP features for classifier optimization and fusion," *J. China Univ. Posts Telecommun.* **25**(01), 41–51 (2018).
2. D.-S. Kang, "The research of face expression recognition based on CNN using tensorflow," *J. Adv. Inf. Technol. Convergence* **7**(1), 55–63 (2017).
3. R. Ahdid, S. Safi, and B. Manaut, "Euclidean and geodesic distance between a facial feature points in two-dimensional face recognition system," *Int. Arab J. Inf. Technol.* **14**(4), 565–571 (2017).
4. H. Zhi and S. Liu, "Face recognition based on genetic algorithm," *J. Vis. Commun. Image Represent.* **58**(Jan.), 495–502 (2019).
5. N. T. Deshpande and S. Ravishankar, "Face detection and recognition using Viola-Jones algorithm and fusion of LDA and ANN," *Adv. Comput. Sci. Technol.* **10**(5), 1173–1189 (2017).
6. J. Zhao et al., "Pre-detection and dual-dictionary sparse representation based face recognition algorithm in non-sufficient training samples," *J. Syst. Eng. Electron.* **29**(01), 196–202 (2018).

7. J. Lin, L. Xiao, and W. Tao, "Face recognition for video surveillance with aligned facial landmarks learning," *Technol. Health Care Official J. Eur. Soc. Eng. Med.* **26**(C), 1–10 (2018).
8. T. Ibrayev et al., "On-chip face recognition system design with memristive hierarchical temporal memory," *J. Intell. Fuzzy Syst.: Appl. Eng. Technol.* **34**(3), 1393–1402 (2018).
9. E. S. M. El-Alfy et al., "A novel approach for face recognition using fused GMDH-based networks," *Int. Arab J. Inf. Technol.* **15**(3), 369–377 (2018).
10. N. Reddy, M. Rao, and C. Satyanarayana, "A novel face recognition system by the combination of multiple feature descriptors," *Int. Arab J. Inf. Technol.* **16**(4), 669–676 (2019).
11. Y. F. Liu et al., "Panoramic face recognition," *IEEE Trans. Circuits Syst. Video Technol.* **28**(8), 1864–1874 (2018).
12. D. K. Jain et al., "GAN-Poser: an improvised bidirectional GAN model for human motion prediction," *Neural Comput. Appl.* **32**(18), 14579–14591 (2020).
13. W. Elloumi, C. Cauchois, and C. Pasqual, "Will face recognition revolutionise the shopping experience?" *Biom. Technol. Today* **2021**(3), 8–11 (2021).
14. D. K. Jain et al., "Deep neural learning techniques with long short-term memory for gesture recognition," *Neural Comput. Appl.* **32**(20), 16073–16089 (2020).
15. S. Adam, "Facing biometrics' future: technology challenge seeks new face-recognition algorithms," *C4ISR: J. Net-Centric Warfare* **16**(6), 30 (2017).
16. H. Matsumura, T. Taketomi, and H. Kato, "Impact of facial contour compensation on self-recognition in face-swapping technology," *Multimedia Tools Appl.* **80**(3), 1–22 (2021).
17. M. Du, "Mobile payment recognition technology based on face detection algorithm," *Concurrency Comput. Pract. Exp.* **30**(22), e4655 (2018).
18. S. Xue, "Face database security information verification based on recognition technology," *Int. J. Netw. Secur.* **21**(4), 601–606 (2019).
19. A. LoSardo, "Faceoff: the fight for privacy in American public schools in the wake of facial recognition technology," *Seton Hall Legislative J.* **44**(2), 6 (2019).
20. T. Sakulchit, B. Kuzeljevic, and R. D. Goldman, "Evaluation of digital face recognition technology for pain assessment in young children," *Clin. J. Pain* **35**(1), 18–22 (2019).
21. C. Rui, S. Lima, and Á. Rocha, "Application of face recognition technology based on CA algorithm in intelligent residential property management," *J. Intell. Fuzzy Syst.* **35**(3), 2909–2919 (2018).
22. C. Liu, "A survey of virtual sample generation technology for face recognition," *Comput. Rev.* **60**(5), 220–221 (2019).
23. Y. Zhang et al., "Secure and efficient outsourcing of PCA-based face recognition," *IEEE Trans. Inf. Forensics Secur.* **15**, 1683–1695 (2020).
24. M. Sumithra et al., "Enhancement of cloud user data access security entrusted to AI face recognition techniques," *J. Cogn. Hum.-Comput. Interact.* **2**(2), 60–64 (2022).
25. Z. Yang et al., "Enhanced deep discrete hashing with semantic-visual similarity for image retrieval," *Inf. Process. Manage.* **58**(5), 102648 (2021).
26. J. Lunter, "Everyday biometrics: can face replace fingerprint recognition?" *Biom. Technol. Today* **2021**(4), 7–10 (2021).
27. Y. Wu and J. Liu, "Research on college gymnastics teaching model based on multimedia image and image texture feature analysis," *Discov. Internet Things* **1**, 15 (2021).

Jian Fang received his MS degree in control engineering from Jilin University, P.R. China. He is currently the dean of the electrical engineering school of Jilin Engineering Normal University. He is studying for a PhD in the School of Mechanical and Electrical Engineering at Changchun University of Technology. His research interests include artificial intelligence, educational robotics, and automation.

Xiaomei Lin earned a doctorate in engineering from Donghua University. She is currently a professor and doctoral tutor at Changchun University of Technology. In recent years, she has closely focused on key scientific issues in national strategic needs and common key technologies in the development of key industries in Jilin Province, carrying out innovative scientific research and technological development in the technical fields of online detection, material composition analysis, and laser induction.

Retracted

Jia Tian received her master's degree from Daqing Petroleum Institute. Now, she works at the College of Electrical Engineering of Jilin Engineering Normal University. Her research interests include artificial intelligence, educational robotics, and automation.

Yue Wu received her master's degree from Changchun University of Technology, P.R. China. Now, she works at the College of Automotive Engineering of Jilin Engineering Normal University. Her research interests include artificial intelligence, educational robotics, and automation.

Retracted