INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2021.15.85 Volume 15, 2021

Face Detection and Recognition Method Based on Improved Convolutional Neural Network

Zhengqiu Lu
School of Information &Media, Zhejiang Fashion Institute of Technology, Ningbo 315211,Zhejiang, China

Chunliang Zhou
School of Finance &Information, NingBo University of Finance & Economics, Ningbo 315175,Zhejiang, China

Xuyang
School of Information &Media, Zhejiang Fashion Institute of Technology, Ningbo 315211,Zhejiang, China

Weipeng Zhang
School of Digital Technology and Engineering, NingBo University of Finance & Economics, Ningbo 315175,Zhejiang, China

Received: January 4, 2021. Revised: July 11, 2021. Accepted: July 27, 2021. Published: July 30, 2021.

***Abstract*—with rapid development of deep learning technology, face recognition based on deep convolutional neural network becomes one of the main research methods. In order to solve the problems of information loss and equal treatment of each element in the input feature graph in the traditional pooling method of convolutional neural network, a face recognition algorithm based on convolutional neural network is proposed in this paper. First, MTCNN algorithm is used to detect the faces and do gray processing, and then a local weighted average pooling method based on local concern strategy is designed and a convolutional neural network based on VGG16 to recognize faces is constructed which is finally compared with common convolutional neural network. The experimental results show that this method has good face recognition accuracy in common face databases.**

***Keywords*— Face Recognition, Convolutional Neural Network, Local Weighted Average Pooling.**

I. INTRODUCTION

FACE recognition is one of the research hotspots in the fields of pattern recognition, image processing, artificial intelligence and cognitive science in recent years. Compared with fingerprint, iris and other biometric recognition technologies, face recognition is widely used in many fields such as criminal detection, security monitoring, high-speed railway ticket checking, and attendance check, due to its advantages of convenience, reliability and non-contact. With the rapid development of deep learning, face recognition technology especially based on deep convolutional neural network has become one of the mainstream research methods.

Face recognition is a biometric technology which is based on people's facial features information for identification. It searches and matches the extracted face image feature data with the feature template stored in the database, and then compares the face feature to be recognized with the face feature template by setting threshold, and finally judges the face of identity information according to similarity. Early in the late 1880s, British psychologist Galton published two papers on face recognition using a set of numbers to represent different human profile features. And then in 1965, Chan and Bledsoe published the earliest academic paper about automation face recognition (AFR) , marking the real start of face recognition research. Face recognition technology can be divided into traditional face recognition technology and face recognition technology based on deep learning [1].

Traditional face recognition technology mainly extracts the geometric relationship between facial features such as eyes, nose and mouth through the understanding of facial information, such as distance, area and angle, etc. These features can be combined into a vector to represent facial information, and finally combined with a high-performance classifier for recognition. Therefore, traditional face recognition technology can be further divided into the face recognition method based on prior knowledge, face recognition method based on statistics and face recognition method based

on sparse representation. Face recognition methods based on prior knowledge mainly cover geometric feature-based methods, template-based matching methods and elastic graph matching methods. In order to overcome the pose, expression and other problems of Face recognition technology in practical application, Face recognition methods based on local feature descriptors, such as LBP and Gabor Face, have appeared since 1998 [2]. At present, the research on this method mainly focuses on the proposed of new features and the fusion of different features. The face recognition methods based on statistics mainly include the methods such as subspace analysis and model-based, and the representative algorithms are PCA, Eigenface, LDA, and Fisher Face, etc. In recent years, Shan Shiguang and Wang Xiaoguang et al. proposed a specific human face subspace for the lack of Eigenface [3] and a unified subspace analysis combining three different subspace methods [4]. The method based on sparse representations is widely used in face recognition fields, and it mainly recovers the input face images by the signal decomposition which is based on sparse regularization constraints, among which SRC[5], MRR[6] and other methods have achieved good results**.**

However, the face recognition method of deep learning abandons the traditional manual feature extraction method and directly uses the original image autonomous learning to obtain highly abstract face features. At present, common deep learning models used in face recognition are mainly Deep Belief Network (DBN)[7] and Deep Convolutional Neural Network. Christian et al. [8] proposed the "Inception" deep convolutional neural network structure, which optimizes the utilization of computational resources within the network. This structure can not only maintain the sparse network structure, but also have the computing power of high-performance dense matrix, so as to greatly avoid the over-fitting situation that is easy to fall into in face training. Angshul et al. [9] proposed a supervised autoencoder algorithm based on class sparse, which effectively improved the face recognition rate.

Lacopo et al. [10] proposed an algorithm to solve large face and pose change problems, which used multi-feature pose model and face images rendering to solve the problem of low accuracy caused by face pose change. Schroff et al.[11] proposed a FaceNet face recognition algorithm which get 99.63% accuracy in Label Faces in the Wild (LFW). This algorithm uses a triple loss function for network training and directly map face image to Euclidean space based on the character of distance in space represents the similarity of face images. Zhang Kaipeng et al. [12] proposed a deep cascading multi-task framework (MtCNN), which can detect and align faces through cascading three deep convolutional networks, and achieve higher accuracy of face alignment. Wen Yandong et al. [13] proposed the Center Loss Function, which can increase the distance between classes and have good generalization ability for new data. Liu Weiyang et al. [14] proposed Large Margin Softmax Loss function, which can effectively guide the network learning features of small in-class distance and large inter-class distance, and can avoid over fitting. DONG Lanfang [15] tuned the pre training model so that the pre training model can complete the current specific task, reduce the training time of the model and accelerate the convergence speed of the model. ABUZNEID[16] put forward a face recognition method based on traditional manual features and convolutional neural network, in which LBP graph is obtained by LBP Operator first and used as input of convolutional neural network. Deng proposed Arcface[17]loss, which normalized the feature vector to the hypersphere, and maximized the classification boundary in the angle space. Zhu Fuli[18] proposed a face detection method based on enhanced parallel cascade convolution neural network to solve the problem of low accuracy of small-scale, fuzzy and occluded face detection in complex scenes. Li Haoxuan[19] proposed darknet53 network, which uses a large number of bottleneck connections, and can reduce the dimension when the dimension of the feature graph is high.

II. MULTI-TASK CONVOLUTIONAL NEURAL NETWORK (MTCNN)

Traditional face detection method requires manual extraction of image features, while the detection method based on deep learning can automatically extract features by building convolutional neural network, which reduces the interference of human factors and has higher recognition accuracy.

MTCNN is a deep cascade multi-task framework, which can simultaneously complete face detection and face alignment, and with the advantages of simple network structure and fast recognition speed. It is composed of three sub-networks, namely Proposal NETWORK (P-NET), Refine NETWORK (R-NET) and Output NETWORK (O-NET), which can detect face and locate face key points simultaneously. However, due to the size of the original face image is different, the relatively small face needs to be enlarged and then detected, and the relatively large face needs to be reduced and then detected. In MTCNN, the original face image should be scaled to different scales to form an "image pyramid", and then the scaled image will be calculated through the neural network. In this way, faces can be detected at a uniform scale.

A. P-Net

P-NET is a full convolutional neural network. The input is a three-channel RGB image which the width and height of pixels is twelve. The output has three parts: whether there is a face in the input 12*12*3 image, the position of the face frame and five key points position, which refer to the left eye, right eye, nose, left corner of mouth and right corner of mouth.

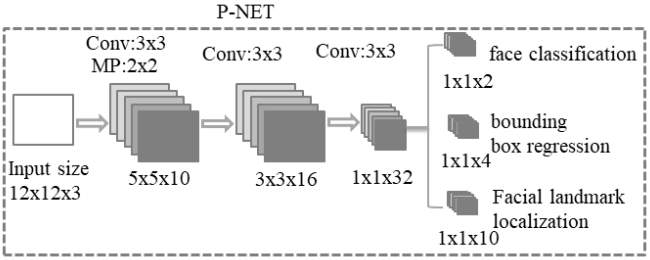

Fig. 1. P-Net

P-NET uses low precision to make sure that all faces can be detected to the greatest extent. In actual calculation, the input

image will be processed by multi-scale pyramid, that is, through the movement of the input layer, each 12*12 area in the original image will be detected once. Meanwhile, for the overlapping candidate boxes, non-maximum suppression (NMS) is used to screen [20], and finally the image will be sent to P-NET network for training.

The so-called non-maximum suppression refers to the overlapping area of two candidate boxes divided by the area composed of candidate boxes. Given the two candidate boxes are A and B, the overlap degree of them is: IOU= (A∩B)/(A∪B).

In the actual development, if there are several multiple candidate boxes, suppose there are six. Firstly, sort according to the classification probability of the classifier category, and set the corresponding threshold. The probability of face from small to large is sorted as A, B, C, D, E, F. the steps is as follows.

Step 1: Starts with the maximum probability candidate box F, judges whether the overlap degree IOU which is from A to E with F is greater than a certain set threshold. Here, the overlap degree refers to the degree to which two candidate boxes overlap;

Step 2: If the overlap degree exceeds the threshold between B and F or between D and F, discards B and D, and marks F as the first rectangle need to save ;

Step 3: Selects E as the highest probability from the remaining rectangular boxes A, C and E, and then judge the overlap degree between E and A or between E and C. If the overlap degree is greater than a certain threshold, discards A or C and marks E is the second rectangle need to save;

Step 4: Repeats this process until all the rectangles need to save will be found.

B. R-Net

The R-NET structure is very similar to P-NET. The input of R-NET is a 24*24*3 image, representing a three-channel image with a pixel size of 24*24. R-Net is used to judge whether 24*24*3 images contain faces and predict the key point's location. In practical application, each P-NET output may be a face area is enlarged to the size of 24*24, and in the input of R-NET, each face will be for further judgment, so as to eliminate a lot of P-NET misjudgment, and improve the accuracy of face detection. The R-Net network structure is as follows.

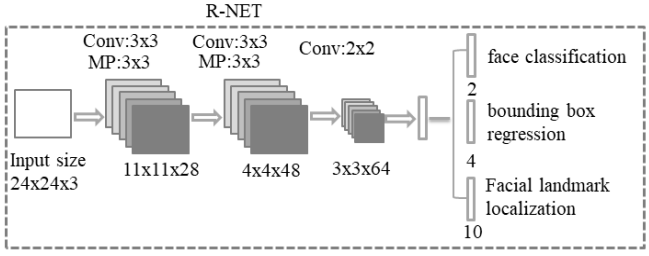

Fig. 2. R-Net

C. O-Net

On the basis of R-NET, the obtained region is enlarged to the size of 48*48, which is used as the input of O-NET network. Other structures of O-Net are similar to P-NET, except with more network channels and layers.

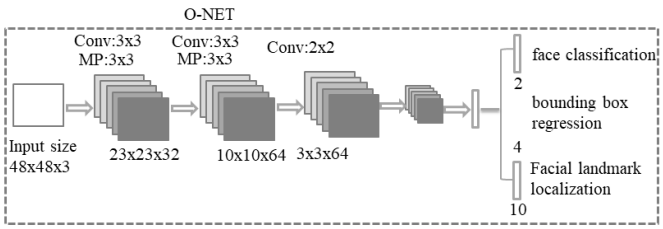

Fig. 3. O-Net

Therefore, the process of face image detection based on MTCNN is as follows:

First step, it preprocesses the input images through the multi-template and multi-scale image pyramid, adjusts all the images to the size of 12*12, and then sends the images to P-NET.

Second step, it generates candidate regions on the processed 12*12 images by P-NET sub-network. Due to the mis-selection and overlap of candidate regions, the non-maximum suppression (NMS) algorithm is adopted to screen candidate regions and the images extracted from the candidate regions are transmitted to R-NET.

Third step, it further adjusts the input candidate regions by the R-NET sub-network in order to make the generated region suggestions more accurate, and then sends the improved results to O-NET.

Final step, it generates the structure of face detection accurately from the candidate region by O-NET, marks the face key points and outputs and finally completes face detection

III. FACE RECOGNITION BASED ON CONVOLUTIONAL NEURAL NETWORK

A. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network, CNN is the most commonly used model structure in computer vision at present. Its training method is mainly to traverse the pixel points of the target image through the convolution Kernel, extract the image features, and predict the image. In the feedback training of a large amount of data, the network parameters are constantly adjusted to form a set of network with good recognition effect on the test image.

Compared with ordinary networks, convolutional neural network has some significant advantages. For example, as to pixels of images, it often has highly similar features in space and has close relationship between each other.

In general, CNN is composed of one or more convolution layers and fully connected layers at the top, as well as pooling layer and activation function.

(1) Convolution Layer

Convolutional layer is the most basic structure in the convolutional neural network. For an input face image, a point pixel of the face has a close relationship with its adjacent local pixels, but has a weak correlation with the distant face pixels. Instead of fully connected network operation, the convolution layer adopts the convolution method to collect image information, which has the local perception and weight sharing

characteristics. The convolutional layer usually appears after the input layer and the pooling layer in the convolutional neural network, and is used in conjunction with its activation function.

The convolution of the input image refers to sliding the convolution kernel on the original image, and convolution operation mainly includes multiplication and addition operations. In order to make sure that the edge information will not be lost in the operation process, the convolution operation is often accompanied by the way of full 0 filling. For example, if the input image size is 5*5, the size of the convolution kernel is 3*3, the convolution operation step size is one, and a 3*3 feature map will finally be formed after convolution operation with non-zero filling.

Convolution layer calculation is different from convolution in the general mathematical sense. The elements of the corresponding positions of two matrices are multiplied and added, and a bias is added at the end. The general mathematical expression is:

$$y^{j(r)} = f\left(b^{j(r)} + \sum_i k^{ij(r)} * x^{i(r)}\right) \quad (1)$$

Here, where symbol r represents the current layer, the symbol $x^{i(r)}$ represents the i-*th* input characteristic figure, symbol $k^{ij(r)}$ represents for the current layer's convolution kernels, the symbol $y^{j(r)}$ represents the j-*th* characteristic diagram, symbol $b^{j(r)}$ represents the bias corresponding to the convolution kernel. Activation function can make sure the nonlinear output of Convolution Layer and enhance the expression ability of whole neural network, namely the convolution results will do nonlinear mapping after convolution operation.

In convolution calculation, a common convolution kernel is used to participate in the calculation. It is a weight sharing method, which is different from the full connection method and it is a local connection which reduces the amount of learning parameters and calculation. Convolution layer can also set up multi-channel convolution calculation, and different convolution kernel sizes can obtain different scale characterization information. Generally the convolution kernel size is 1 * 1, 3 * 3, 5 * 5 or 7 * 7.

(2)Activation Function

Convolution Layer uses nonlinear function as the activation function in the process of convolution operation. The main function of activation function is to add nonlinear factors in order to solve the problem of insufficient expression ability of linear model. Common activation functions in the convolution operation include Sigmoid, ReLu and other functions, and the activation function value formula of the j-*th* neuron in layer l is as follows.

$$w_{j,k}^l = f\left(\sum_k w_{j,k}^l a_k^{l-1} + b_j^l\right) \quad (2)$$

Here, $w_{j,k}^l$ presents the weight of the *k-th* neuron of previous layer, b_j^l is bias term and $f(*)$ is activation function.

Sigmoid is an 'S' type activation function, and the general formula is as follows.

$$\sigma(x) = \frac{1}{1 + exp(-x)} \quad (3)$$

When x<-5 or x>5, the variation gradient of sigmoid function is very small, and the interval gradient can be approximately zero. After calculating the error gradient, the error needs to be back propagated, but it is difficult to transfer the gradient error to the front layer network and result in the difficulty of network training.

ReLu function is the most active activation function which is applied in the hidden layer of convolution operation. Compared with Sigmoid and other functions, ReLu does not have gradient saturation problem while the input data is positive, which makes the ReLu have faster compute speed. If the input data is greater than 0 or less than zero, ReLu does not do any complex computing.

$$ReLu(0, x) = \begin{cases} x, if\ x \geq 0 \\ 0, if\ x < 0 \end{cases} \quad (4)$$

(3)Pooling Layer

Pooling Layer is an important component of convolutional neural network, and it exists in pairs with convolution layer. The features is firstly extracted in convolution layer and then input into pooling layer in CNN network. The main purpose of introducing the Pooling Layer is to reduce dimensions, obtain lower dimensional representation information, improve the generalization ability of the model, and remove the noise and redundancy that may appear in the image.

According to the principle of relative invariance of each feature position of the image, pooling layer uses the aggregation statistical method to process the image information, and reduce the dimension of the feature object. Its main purpose is to replace all image features with the obtained image summary features, and retain the effective feature information.

The formula for the pooling layer is as follows.

$$X_j^l = f\left(down\left(X_j^{l-1}\right)\right) \quad (5)$$

Here, where symbol X_j^{l-1} represents the output of former layer on the convolutional network, and $f(*)$ represents the excitation function of the pooling layer . $down(*)$ is also a down sampling function, which is called the pooling function. Pooling methods mainly include average pooling, max pooling and stochastic pooling.

The input of maximum pool calculation is a two-dimensional matrix with the size of 4 * 4, and then finds the maximum value of the region through a 2 * 2 filter. Given moving distance of the sliding window is the step size, and its value is two, its output is a 2*2 characteristic matrix.

The steps of average pooling calculation are similar to max pooling calculation. It only needs to take the average value of the number in the sliding window. If the filter size is two or three, the output result is a characteristic graph which is compressed to 1 / 4 or 1 / 9 of the original one.

(4)Output Layer

Output layer usually consists of a full connection layer and a non-linear loss function Softmax.

Full connection layer is generally set at the end of CNN to

get high purification feature, which can be seen as highly condensing a graph into a specific value for the convenience of handing it to the final classifier. The general calculation formula is.

$$y_i = f(\sum_i x_i * \omega_{i,j} + b_i) \quad (6)$$

Here, where x_i and y_i respectively represent the input and output of this layer, and ω and b represent weights and offsets.

Convolution neural network can extract advanced features for classification, and can obtain satisfactory recognition accuracy with simple classification. It needs to use Softmax function to solve the multi category problem, and the loss function can realize the classification task of convolutional neural network. The Softmax model is an extension of logistic regression model, and probability distribution formula used in Softmax model is as follows.

$$p(Z_j) = \frac{e^{Z_j}}{\sum_{i=1}^{n} e^{Z_j}} \quad (7)$$

Where, Z_j is the vector value calculated by Softmax function using weight product plus bias calculation and then mapped to set (0, 1), with total value of these vector is one. $p(Z_j)$ is the prediction target result and the maximum $p(Z_j)$ value is selected while outputting.

B. FACE RECOGNITION IMPLEMENTION

Face recognition process based on convolutional neural network includes five stages, and they are face collection, face detection, gray-scale processing, face recognition modeling and face recognition, which are shown in Fig. 4.

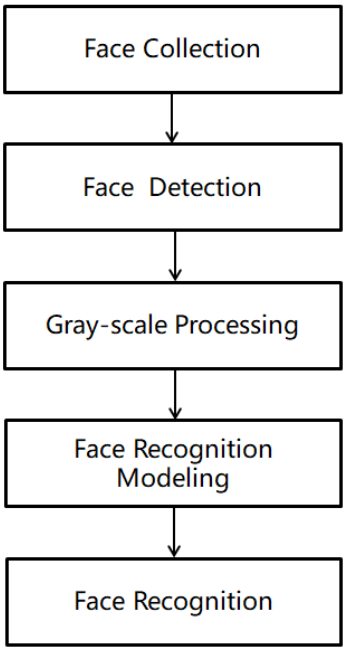

Fig. 4. Overall process of face recognition

(1)Face Collection

Face data collection comes from two parts, one of which is collected from face databases. CMU_ PIE face database is created by Carnegie Mellon University of United States, with a total of 3332 faces. The images with pose and illumination change are also collected under strict control, and this face database has gradually become an important test set in the field of face recognition. CASIA-Face database contains 2500 Asian face images from 500 people, of which 100 are selected in this paper. ORL face database was created by AT & T Laboratory of Cambridge University. There are 400 facial images, and images of some volunteers include the changes of posture, expression and facial accessories. LFW is a face database completed by the Computer Vision Laboratory of Amherst University of Massachusetts, and mainly used to study the face recognition under unrestricted conditions. It is a common test set for face recognition at present, of which the face images are all from natural scenes in life.

Other face images are collected from our own computer camera. Face acquisition module needs to call the camera to take photos. In this paper, OpenCV is used to call the computer camera to take pictures, and then the taken pictures are stored in the certain directory.

Firstly, initializes the parameters, including the images number taken for each face, the time interval between two images captured by the camera and the saved location of face image. Secondly, calls the computer camera and lets the camera warm up for one second to prevent underexposure. Then, reads the captured face image circularly and saves it to the specified file path. Finally, releases the camera after shooting.

(2)Face Detection

The purpose of face detection is to detect the face's specific location from the image, and in this paper, MTCNN algorithm is used for face detection.

Firstly, reads each face image circularly, and then calls the face detection interface of MTCNN to detect the face image. If there is a face in the image, retains the face image. In this paper, sets the NMS value to 0.7, that is, if the NMS value is more than 0.7, it is considered to be a positive face. Secondly, further judges the face information, extracts the face location information and stores it in the variable of faceboxes. If there is a negative value in the faceboxes by sum method of numpy lib, deletes the face image. Finally, intercepts the face information of each face image and prepares for the next gray processing.

(3)Gray-scale Processing

General color images are composed of RGB three color channels, and each color channel has a value range of 0-255. The larger the value, the darker the corresponding color. Therefore, the general way to store images in the computer is a three-dimensional m*n matrix, each dimension represents a color channel, and m and n respectively represent the width and height of the image.

Gray-scale processing is to remove the color in the image information and retain only the brightness information, the reason for this is that extracting the characteristic value from color images are greatly influenced by light, and color image has large amount of calculation in the subsequent process. After converted into gray image, there are still some characteristic distributed in the overall and local that can reflect the image color and brightness. And this paper mainly focus on the

recognition of face contour features, the color requirements are relatively low. So the image gray-scale processing, not only does not affect the effect of face detection, but also reduce the computation amount of this algorithm and improve the efficiency of face recognition.

The method of gray-scale processing uses weighted summation of three channels including R G and B in the color image. The formula is as follows.

$$Gray_{i,j} = 0.299 * R_{i,j} + 0.587 * G_{i,j} + 0.114 * B_{i,j} \quad (8)$$

Here, where R, G and B represent three color channels, and i, j represent the pixels positions. The face effect after the first three steps is shown in Fig. 5.

Fig. 5. Gray processing effect

(4)Face Recognition Modeling

In this paper, the face recognition model based on VGG16 network is adopted. The structure of the whole neural network consists of four convolution layers, five ReLu layers, two pooling layers, three dropout layers, two full connection layers, one flatten layer and one classification layer Softmax with a total of eighteen layers.

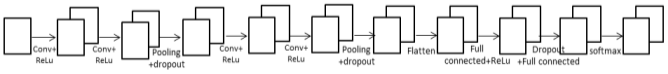

Fig. 6. Improved Convolutional Neural Networks

The input variable is the image data detected by MTCNN, processed by gray-scale into a single channel, and then compressed to an image with width and height of sixty-four. Therefore, the dimension of each image is 64*64*1, and this network can accept several images. The dimension of the input layer of the whole network is N*64*64*1, where N is the number of images.

In the first convolution operation, the input image is convolved by thirty-two convolution kernel of 3*3*1, and after convolution, ReLu function is used as the nonlinear mapping function. The second convolution operation is the same as the first one.

Each 2*2 region is done pooling after the two convolution operations are completed. Since maximum pooling operation only extracts the maximum value of the pooling area, and the element information and residual information are lost. While the average pooling operation directly calculates the average value of elements within the pooling area, and ignores the difference in contribution intensity among elements. So the local weighted average pooling method based on the local concern strategy is adopted in this paper, which add learnable pooling weights to each element in pooling area to get the purpose of showing effective information and suppress the invalid information reasonably, and the maximum pooling and average pooling can be regarded as two special cases of this method.

The specific construction process of local weighted average pooling method for the layer k is as follows：

Step 1: Constructs the element set in the local pooling region of the input feature graph. According to the feature graph of the convolution output at the $k-1$ layer above, constructs the elements set of local region which are need to be done pooling.

$$Y = \{y_1^{k-1}, y_2^{k-1} \dots y_i^{k-1}\}, i = \{1,2, \dots n\} \quad (9)$$

Step 2: Determines the initial weight matrix of the elements in the locally pooling region. Generates random initial weight in the range of [0, 1] by truncated normal distribution function $X \sim N(\mu, \sigma^2)$, and the corresponding random initial weight matrix is:

$$U = \{u_1^k, u_2^k, \dots, u_i^k\}, 0 \le u_i \le 1 \quad (10)$$

Step 3: Constructs weighted pooling kernel vector. Get the final pooling weight u_i^k after adding Softmax function to initial weight u_i^k, so proportion of the elements with strong discriminating ability is larger and the elements with weak discriminating ability is smaller.

The corresponding weighted pooled kernel vector is expressed as:

$$B = \{\beta_1^k, \beta_2^k, \dots \beta_i^k\}, i = \{1,2, \dots n\} \quad (11)$$

Step 4: Establishes a local weighted average pooling representation. Multiply the generated weighted pooling kernel vector with elements set, and gets the *j-th* pooled abstract character representation after calculation of weighted average pooling. Finally, add dropout layer and do regularization to prevent over-fitting.

In the third convolution operation, convolute the feature map generated from second convolution layer by sixty-four convolution kernel with 3*3*32. The fourth convolution operation is the same as the third. Then do the second pooling operation and regularization by dropout.

After the convolutional pooling operation is full connection layer, it includes the flatten layer, the dense layer, the activation function layer and the dropout layer.

(5)Face Recognition

Face recognition is the last stage, including face model training and face result testing. In face model training, firstly reads the gray processed face images, splits the face data set by setting validation_ split value and then gets the test data set and verification data set respectively. Then starts the model training, sets relevant parameters, including batch_size, epochs and other parameter values during the training process and saves the training results into the 'cnn_model.h5' file. Finally, imports the trained model and puts the sample data into the model for prediction. If the test result reaches the expected threshold, it indicates that the training has been completed. In this paper, the initial threshold is 0.95.

While testing face results, firstly generates a video stream with a certain number of frames by calling the camera, and captures the picture of each frame. Then, detects the face part

by MTCNN algorithm, and does gray processing and size processing. Finally, recognizes the face part by established model. In this paper, a face rectangle will be drawn with the cv2 object of OpenCV library according to four coordinate face points, and the specific face object will be displayed with text.

IV. EXPERIMENTS

The experimental environment of development and training used in this paper shows in Table. 1.

Table. 1. Experiment environment and configure

Experiment environment	Experiment configure
Operation system	Windows10
Memory	16G
Programming language	Python3.67
Face collection	OpenCV
Experiment platform	GeForce GTX 1060 GPU

Public face database such as CMU_PIE are selected for training and testing in the experiment, and there are sixty-eight people in total and forty-nine faces with different angles and shapes per person. At the same time, eight faces are collected through the camera, with a hundred faces for each person. So there are four thousand one hundred and thirty-two faces in this experiment, which make sure the diversity of face data collected.

In order to verify the performance of the improved convolutional neural network, the proposed convolutional neural network in this paper is compared with the common convolutional neural network, where its model includes three convolutional layers, three pooling layers and one full connection layer.

Firstly, we split the face data randomly for model train, with 70% of the data are used for training and 30% for testing. Then set epoch as 700 and batch size as 120, and the accuracy and loss of these two networks model on training set are shown in Fig.7 and Fig.8.

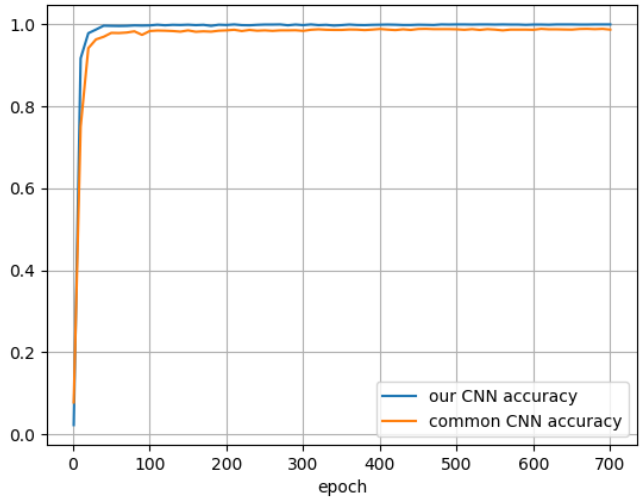

Fig. 7. Train accuracy

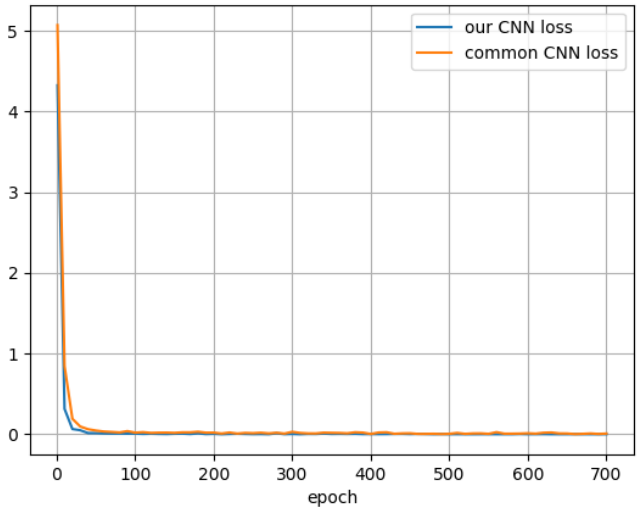

Fig. 8. Train loss

It can be seen from Fig. 7 and Fig. 8 that the convolutional neural network model proposed in this paper has good accuracy in the training stage and can converge faster than the traditional convolutional neural network in the loss.

Then the two networks were trained in the same experimental environment, and we respectively count the accuracy where epoch values are 10, 100, 200, 300, 400, 500, 600, 700, 1000, and 1500. The accuracy of the two network models in the test set are shown in the fig. 9. It can be seen from the figure that the proposed convolutional neural network algorithm in this paper has better effect than the common convolutional neural network.

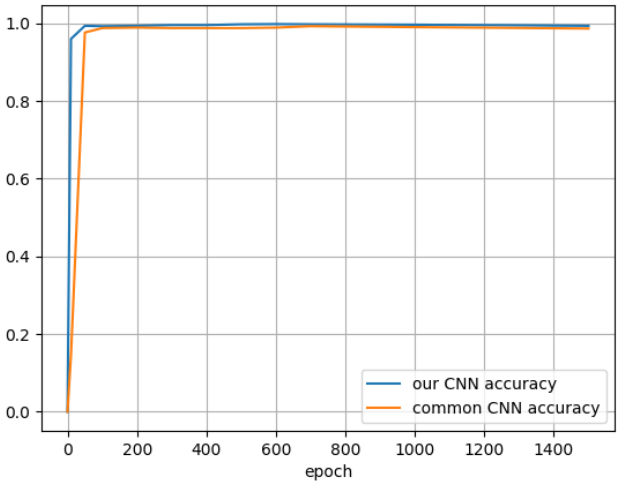

Fig. 9. Test accuracy

In addition, in order to verify the generalization of the algorithm in this paper, we do experiments on the face databases such as ORL, Yale, CASIA-Face and LFW, and the experimental results in these common face databases are high accuracy which is showed in the Table. 2.

Table. 2 Algorithm accuracy in different face databases

Face databases	Accuracy
CMU_PIE	99.79%
ORL	99.50%

Yale	98.79%
CASIA-Face	98.25%
LFW	97.63%

Here, the improved convolutional neural network gets the highest accuracy of 99.79% in CMU_PIE face database, secondly is ORL, Yale and CASIA-Face, and the accuracy in LFW is a little low relatively because each person's face is less even only one face in this face database. So the collection number of single face has great influence on accuracy of face classification recognition. In addition, the accuracy is 99.76% in self-made face database with the face is collected by OpenCV.

V. CONCLUSION

According to the research of face detection and face recognition, this paper proposes a face recognition model based on convolution neural network. It collects face data with OpenCV, does face detection and gray-scale processing with MTCNN, and finally does face recognition with improved convolution neural network. Meanwhile, aiming at the existing situation that the information is loss in maximum pooling layer in convolution neural network, it designs a local weighted average pooling method based on local concern strategy under the largest pool and average pool methods. Finally, this paper does a large number of face character data test and the experimental results show that the proposed algorithm has high accuracy in different face databases.

ACKNOWLEDGMENT

This work was supported in part by Foundation of Zhejiang Fashion Institute of Technology(2019-2A-003), Zhejiang Philosophy and Social Science Project(NO.21 NDJC167YB), and the Public Welfare project of Ningbo City (202002N3139 , 2019C10051).

REFERENCES

[1]Jing Chenkai, Song Tao, Zhuang Lei, "A survey of face recognition technology based on deep convolutional neural networks", *Computer Applications and Software*, Vol. 35,no. 1, pp. 223-231, 2018.

[2] C Dong, X Cao, W Fang, S Jian, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification", in *Proc. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 2013, pp. 3025-3032.

[3] S.Shan, W.Gao, D.Zhao, "Face identification from a single example image based on face-specific subspace(FSS)", in *Proc. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, 2002, pp. 2125-2128.

[4] X.Wang, X.Tang, "A unified framework for subspace face recognition", *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 26, no. 9, pp. 1222-1228, 2004.

[5] J. Wright, A. Y. Yang, A. Ganesh, et al, "Robust face recognition via sparse representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, no. 2, pp. 210-227,2009.

[6] M.Yang, L.Zhang, D.Zhang, " Efficient misalignment-robust representation for real-Time face recognition", in *Proc. Proceedings of European Conference on Computer Vision*, Florence,2012, pp. 850- 863.

[7] G.E.Hinton, R.R.Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science*, Vol. 313, no. 5786, pp. 504-507, 2006.

[8] Szegedy C, "Going deeper with convolutions", in *Proc. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015, pp.1-9.

[9] A Majumdar,R Singh,M Vatsa, "Face Verification via Class Sparsity Based Supervised Encoding", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, no. 6, pp. 1273-1280, 2017.

[10] Iacopo M, Stephen R, Gérard G, et al, "Pose-aware face recognition in the wild", *IEEE Signal Processing Letters*, Vol. 23, no. 8,pp. 4838-4846,2016.

[11] Schroff F, Kalenichenko D, James P, "Facenet: A unified embedding for face recognition and clustering", in *Proc. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015, pp. 815–823.

[12] Zhang K. P, Zhang Z. P, Li Z. F, et al, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks", *IEEE Signal Processing Letters*, Vol. 23, no.10, pp. 1499-1503, 2016.

[13] Wen Y. D, Zhang K. P, Li Z. F, et al, "A Discriminative Feature Learning Approach for Deep Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 47, no. 9, 2016, pp. 499-515.

[14] Liu W. Y, Wen Y. D, Yu Z. D, et al, "Large-Margin Softmax Loss for Convolutional Neural Networks", in *Proc. Proceedings of the 33rd International Conference on Machine Learning*, New York, 2016, pp.507-516.

[15] DONG Lanfang, ZHANG Junting, "Research on face age and gender classification based on deep learning and random forest", *Computer Engineering*, Vol. 44, no.5, pp. 246-251, 2018.

[16] ABUZNEID M A , MAHMOOD A, "Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network ", *IEEE Access* , Vol. 6, no.3, pp. 20641-20651, 2018.

[17] Deng J, Guo J, Xue N, et al, "Arcface: Additive angular margin loss for deep face recognition", in *Proc. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019, pp. 4690-4699.

[18] Zhu Fuli,Yang Lei,Ji Bo, "Face detection method based on enhanced parallel cascaded convolutional neural network", *Computer Applications and Software*, Vol. 37, no.11, pp. 101-105, 2020.

[19] Li Haoxuan, Wu Dongdong, "Multi-Face Real-time Detection in Natural Scene Based on Deep Learning", *Journal of test and measurement technology*, Vol. 34, no.1, pp. 41-47, 2020.

[20]Hou Zhiqiang, Liu Xiaoyi, Yu Wangsheng etal. "Improved algorithm of Faster R-CNN based on double threshold-non-maximum suppression", *Opto-Electronic Engineering*, Vol. 46, no.12, pp. 82-92, 2019.