Research on Face Recognition Technology Based on Deep Learning

Zhou Chun-rong
Office of Science and Technology Administration
Chongqing Vocational College of Transportation
Chongqing, China
wzxsly134564@163.com

***Abstract*—The traditional face recognition technology requires a lot of computation time for human face feature extraction, and the recognition accuracy is limited by the extracted face features. Aiming at the above problems, this paper studies face recognition technology based on deep learning. For the image to be recognized, the HOG feature is used to detect the face target in the image. In order to avoid the influence of facial expression and posture, face alignment processing is performed. After extracting the face recognition features in the image, the neural network is used to realize face recognition. Through comparison experiments with traditional face recognition technology, it is verified that the recognition technology using deep learning improves the recognition efficiency by about 1.5 times and has better recognition adaptability.**

***Keywords*—Deep learning, Face recognition, Feature extraction, Neural network**

I INTRODUCTION

Face recognition is a complex classification problem. Face recognition is a supervised classification problem. It refers to training samples after labeling the personal identity information of the face image in advance, and then classifying the face image to be recognized into a specific personal identity information set with a label. Face recognition technology is one of the important technologies of information security, computer network information verification and e-commerce verification. With the rapid development of the Internet, many biological technologies have been applied in the process of information security, computer network login and e-commerce. Among them, face recognition is one of the important parts in the field of biometrics, and plays an irreplaceable role in information security. With the continuous development of camera phone products in recent years, the way to obtain facial images has become more and more simple. Face recognition is developed from image recognition, and the recognition effects of different algorithms vary greatly [1].

Although the principle of face recognition technology based on KNN algorithm is simple, the calculation complexity is small, and the calculation speed is fast, but KNN algorithm is a classification algorithm that can not be iterated, and each calculation needs a lot of calculation time and complexity. In addition, its anti-interference performance also has great deficiencies. Face recognition technology based on Naive Bayes algorithm is very difficult to calculate, but the complete probability can not represent the relationship between face features, which has obvious disadvantages in practical application [2].

Deep learning is a relatively new branch of machine learning and artificial intelligence research, and it is also one of the most popular scientific research trends now [3]. In the past few years, the principles and algorithms of deep learning have changed how people process, model, and interpret data. Deep learning has been proven to be very good at discovering complex structures in high-dimensional data, so it is suitable for many fields in science, business, and government. Now most statistical models that accurately reason and predict user behavior are based on deep learning algorithms. Face analysis has always been a hot topic in the field of computer vision. In recent years, with the development of deep learning technology, the field of face analysis has also made great progress. Based on the above analysis, this paper will study face recognition technology based on deep learning.

II RESEARCH ON FACE RECOGNITION TECHNOLOGY BASED ON DEEP LEARNING

A. Using HOG features to realize face detection

Face detection is to determine the position, size and pose of the face in a given image, so as to avoid the influence of these factors on the accuracy of face detection. In this paper, hog feature is used to detect face. The core idea of directional gradient histogram (HOG) feature is to calculate on a cell unit with uniform size and dense grid, and use overlapping local contrast normalization technology to improve the description ability. HOG feature can describe the image local difference information well and is not easily interfered by noise. It is one of the best features to describe the edge and shape in the field of target detection and tracking.

The steps of hog feature extraction are as follows [4]:

(1)Normalized gamma. The gamma transform is used to normalize the image, which can effectively reduce the local shadow and illumination changes of the image.

(2)Sliding window settings. The segmented image is a number of sliding windows (block), and the block is used to slide the HOG feature of the face on the whole image.

(3)Calculate the gradient. The block is evenly divided into four units, and the blocks slide by overlapping two cell units. The horizontal and vertical gradients of pixel (x, y), gradient $I_x(x, y)$ in horizontal direction and gradient $I_y(x, y)$ in vertical direction are calculated.

$$\begin{cases} I_x(x, y) = I(x+1, y) - I(x-1, y) \\ I_y(x, y) = I(x, y+1) - I(x, y-1) \end{cases} \quad (1)$$

The gradient amplitude $m(x, y)$ of pixel (x, y) is obtained as follows:

$$m(x,y)=\sqrt{\left[I_x(x,y)\right]^2+\left[I_y(x,y)\right]^2}^2 \quad (2)$$

The gradient direction $\theta(x,y)$ of pixel (x,y) is as follows:

$$\theta(x,y)=\arctan\left[\frac{I_y(x,y)}{I_x(x,y)}\right] \quad (3)$$

(4)Accumulate and calculate the gradient direction of the space cell. Divide the gradient direction into m evenly. m is usually 9 directions, if the gradient direction is positive or negative, it will be 360. Divide evenly into m intervals, otherwise divide 180° into m intervals evenly. Perform a weight-based cumulative calculation on the gradient amplitudes of all points in the same gradient direction on the same cell to obtain the gradient histogram of the cell.

(5)Perform normalized comparison within overlapping cell blocks. Normalize multiple cell gradient histograms in each block into a histogram to represent the HOG feature of the current block.

(6)Collect hogs for all blocks on the detection window. The hog feature of the whole image is extracted by sliding the block window.

(7)Output hog feature.

After extracting hog features, the matched face detector is used to detect the face. First of all, face detector training is needed for face matching using hog features. The training is divided into two stages: preprocessing and training [5].

Preprocessing stage: 1) Input the collected face image training set and test set, and mark the specific face boundary of each picture; 2) Image up-sampling, the purpose is to enlarge the smaller face to make the detection accurate Increase the degree, adjust the corresponding bounding box while upsampling; 3) Mirror the face images in the training set to expand the number of samples.

Training stage: 1) define the scanner to scan the image and extract hog features; 2) set the window size of the scanner to detect the face in different sizes through image pyramid transformation; 3) define the trainer to train the face detector, which is initialized by the scanner; 4) train to generate and perfect the face detector when there is an unconfirmed face boundary 5) test and get the returned detection results. After face detection, face alignment is needed.

B. Face alignment processing

There is a strong correlation between face shapes, so face shapes can be converted into low dimensional shape parameters to store face shape information, and low dimensional shape parameters can be used for iterative operation. Principal component analysis (PCA) is used to reduce the dimension of all face shapes in the training set, and the shape eigenvalues and shape eigenvectors are obtained to form the face shape space. The expression of face shape in face shape space is as follows [6]:

$$S=T_{X_t,Y_t,c,\theta}\left(\overline{S}+P\phi\right) \quad (4)$$

In the formula, $T_{X_t,Y_t,c,\theta}$ is the transformation amount of shape rotation, scaling and translation, $\overline{S}$ is the average face shape, ϕ is the face shape space parameter, and P is the coefficient of the face shape space parameter. According to formula (4), the face shape can be restored by inputting the low-dimensional shape parameter ϕ, and different face shapes can be obtained by inputting different ϕ. As long as the training samples have diversity of face shapes, a powerful face shape space can be formed. Face alignment can be regarded as the process of finding an optimal shape parameter. In order to find the optimal face shape parameters, combined with the joint regression method, the face shape parameters are updated iteratively until the shape parameters converge. The iterative formula is as follows [7]:

$$\phi^t=\phi^{t-1}+R^t\left(I,\overline{S}+P\phi^{-1}\right) \quad (5)$$

Among them, ϕ^t is the face shape parameter of the t iteration and $R^t\left(I,\overline{S}+P\phi^{-1}\right)$ is the regression quantity of the shape parameter related to the face feature and the current shape.

Given N training sample $\left\{I_i,\hat{S}_i\right\}_{i=1}^N$, in order to improve the effectiveness of the algorithm, each training sample is assigned several initial face shapes. These initial shapes are obtained by the average face shape through shrinkage, rotation and translation transformation.

Calculate the parameter regression R^t according to the sample characteristics and the current face shape, so that the alignment error of all samples is minimized, namely [8]:

$$R^t=\arg\min_R\sum_{i=1}^N\left\|P\left(\delta\phi_i^{t-1}-R\left(I,\overline{S}+P\phi^{-1}\right)\right)\right\| \quad (6)$$
$$\delta\phi_i^{t-1}=\hat{\phi}_i-\phi_i^{t-1}$$

Among them, ϕ_i^{t-1} is the $t-1$ shape parameter of sample i, $\hat{\phi}$ is the true shape parameter of sample i, and $\delta\phi_i^{t-1}$ is the $t-1$ alignment parameter error of sample i. In order to make the parameter regression quantity R^t more effective, a secondary shape parameter regression algorithm is designed. The parameter regressor R^t is subdivided into k weak parameter regressors r, and each weak regressor r is a low-dimensional shape parameter increment set $\left\{\Delta\phi^1,\Delta\phi^2,\cdots,\Delta\phi^{2F}\right\}$. In the process of calculating k weak regressors, several features that are more relevant to this iteration are selected from the same feature pool. Since the features selected in each iteration process are weak features, a strong feature is formed after k iterations , Making the parameter regression R^t more effective. After aligning the faces in the image, the image is further processed to extract the face recognition features in the image to facilitate face recognition.

C. Extracting face recognition features from images

Usually, the image is collected in YUV mode, and YUV is converted to RGB mode by mode conversion. The image conversion function of YUV to RGB mode is shown in formula (7) [9-10].

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1.347028 \\ 1 & -0.322467 & -0.673154 \\ 1 & 1.765912 & 0 \end{pmatrix} \begin{pmatrix} Y \\ U-128 \\ V-128 \end{pmatrix} \quad (7)$$

After the image mode is converted, the image is preprocessed. When people collect images, they will be affected by light, human posture, and even the background during image collection, resulting in errors in automatic recognition. The preprocessing of the collected images can eliminate the influence to the greatest extent and improve the accuracy of automatic recognition.

After geometric normalization of the collected image, as shown in the figure below, the eyes and nose of the face are taken as three feature points to calibrate the feature points. The intersection point of the horizontal line between the two eyes and the longitudinal extension line of the nose was taken as the coordinate midpoint o, and the distance between the two eyes was D. Take o as the geometric normalization center reference point, cut d left and right, and take 0.7d and 2D regions in the vertical direction to cut the collected images. The face region of the original image is copied to the standard size image to realize the geometric normalization of the image.

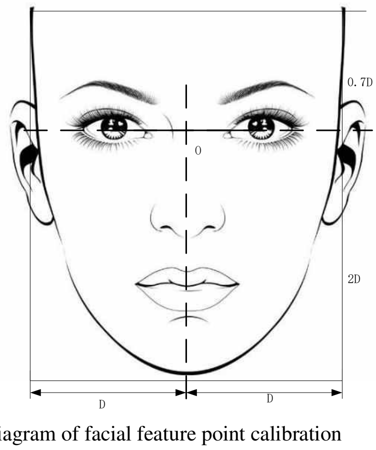

Fig.1. Schematic diagram of facial feature point calibration

Because RGB can not reflect the morphological features of objects in the image, it will affect the feature extraction of images. Therefore, it is necessary to convert the image into gray image. From the optical point of view, the brightness is quantized and divided into 0-255 levels, and the image is grayed according to formula (8).

$$G = 0.39 \times R + 0.50 \times G + 0.11 \times B \quad (8)$$

The noise in the image transmission process is processed by mean filtering. Choose a window with a size of m×m, and slide the window sequentially from the first pixel in the image, and calculate the average value of pixels in the original image window corresponding to pixel (x, y) according to formula (9) for each window sliding [11-12].

$$f(x, y) = \frac{1}{m}\left(\sum_{i=-1}^{1}\sum_{j=-1}^{1} f(x+i, y+j)\right) \quad (9)$$

In formula (9), m is the window size. $f(x, y)$ is the average value of pixels in the window for every (i, j) unit sliding of the window.

This paper uses the LBP feature vector extraction method to extract the features of the face image. Taking the image in the training sample set as a sample, the image is divided into M regions, the coordinate of the center point of the region is (x_c, y_c), and the coordinate of the point adjacent to the center pixel is $(x_n, y_n), n \in N$. The calculation formula of pixel (x_n, y_n) coordinate is as follows:

$$\begin{cases} x_n = x_c + r\cos\left(\dfrac{2\pi n}{N}\right) \\ y_n = y_c - r\sin\left(\dfrac{2\pi n}{N}\right) \end{cases} \quad (10)$$

In formula (10), r is the radius of the circular feature sampling area, n is the number of sampling points in the sampling area, and N is the number of circular sampling areas in the sample image. Calculate the pixel gray value of each sampling area according to formula (11).

$$F(x, y) \approx [1-x]\begin{bmatrix} F(0,0) & F(0,1) \\ F(1,0) & F(1,1) \end{bmatrix}\begin{bmatrix} 1-y \\ y \end{bmatrix} \quad (11)$$

Using the calculated gray value of the central pixel of the area as the threshold value, follow the calculation process shown in the figure below to solve a string of binary unsigned numbers, that is, the LBP feature value of the sampling area.

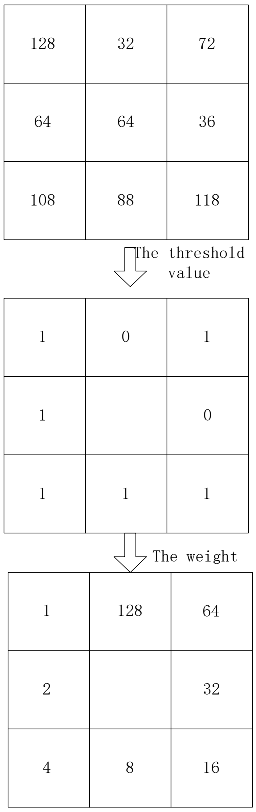

Fig.2. Schematic diagram of LBP eigenvalue calculation

The steps in the above figure can be expressed as: compare the threshold value with the gray value of the pixels around the center pixel. If the gray value is equal to the threshold, the pixel is marked as 1; if the threshold is greater than the gray value of the surrounding pixels , The pixel position is marked as 0. By analogy, until all the surrounding pixels are marked, the 0 and 1 numbers after the mark are arranged in a certain order, and the LBP characteristic value of the sampling area can be obtained. The binary sequence represents the characteristics of the corresponding block. After normalizing the histograms of all blocks in the sample image, the face recognition feature vector of the sample image is obtained. After extracting the face recognition feature vector, the neural network is used to complete the recognition process.

D. Face recognition based on Neural Network

This paper chooses deep residual network for face recognition. ResNet introduces a special residual learning unit between output and input, adding the input and output to form a residual structure. The ResNet network model can construct different levels of network structure, which solves the problem of the disappearance of gradients caused by increasing depth.

In the neural network, the convolution layer convolutes the input network image according to the following formula:

$$S(i,j)=(X*W)(i,j)=\sum_{m}\sum_{n}x(i+m,j+n)w(m,n) \qquad (12)$$

In the above formula, X is the two-dimensional image input to the neural network; W is the convolution kernel, which is a two-dimensional matrix. If X is a multi-dimensional tensor, the corresponding W is also a multi-dimensional tensor. It can be seen from the formula (12) that the convolution operation on the image is actually the matrix multiplication operation on the matrix and the convolution kernel of different regions of the input image.

Since the dimension of the feature graph after convolution layer does not decrease, the dimension compression of the input feature graph can be carried out through pooling layer.

When the neural network processes the input image, it is necessary to introduce a loss function to increase the diversity of the neural network's recognition of human faces, so as to improve the accuracy of the neural network model in recognizing the difference of different human faces. This paper chooses the cross-entropy loss function, and the specific form is as follows:

$$L=-\frac{1}{K}\sum_{i=1}^{M}\frac{e^{W_{y_i}^{T}x_i+b_i}}{\sum_{j=1}^{m}e^{W_{y_i}^{T}x_i+b_i}} \qquad (13)$$

In the above formula, W is the weight matrix; W_j is the j column of the weight matrix; b is the bias term; x_i is the i training sample; M is the number of samples. After the neural network is constructed, the TensorFlow framework is used to realize the training process of the network, and realize the adjustment of the loss function, and realize the data interaction by defining the relevant parameters during the training process. So far, the research on face recognition technology based on deep learning has been completed.

III TESTING EXPERIMENT

The face recognition technology based on deep learning is studied above. In this section, simulation experiments will be conducted to verify the feasibility and effectiveness of the technology [13-14].

A. Experimental content

This experiment was carried out in a computer simulation platform, and the experimental data of this experiment was collected using the data monitoring software in the computer simulation platform. Load the two face recognition technologies into a computer simulation platform with exactly the same configuration, and use the prepared face data set as the recognition object of the two face recognition technologies. The face recognition technology based on deep learning studied in this paper is compared with the face recognition technology based on statistical features, and the performance of the two methods is evaluated by comparing the experimental data of the corresponding contrast parameters of the two recognition technologies.

The comparison index of this experiment is the precision and recall of the two face recognition technologies for face data set recognition, as well as the time-consuming of face data recognition. By comparing the above experimental indicators, processing and analyzing the experimental data, the performance of the two face recognition technologies is compared, and the experimental verification is completed.

B. *Experimental data set and preparation*

The following face data sets were used for the experiment

1) In ORL face data sets with small change of camera angle, slight pose and expression changes and single image background;

2) IFD face collection, the total amount of data is 1680 photos, each of which belongs to 120 people, each with 14 photos, with different angles but consistent backgrounds. The changing angle of the face in the IFD picture is much larger, and the facial expressions of the two different pictures of the same person have also changed a lot;

3) Video face dataset. The images of video frames are downloaded from the video website. There are more than 300 people in the data set, which are face images collected in different scenes. It can represent scenes in real life. The size of the face image data set is about 5.4gb.

Call the random.shuffle() method in Sklearn to refresh the data set, and select 20%, 25%, and 30% of the data set in the disordered order after refreshing as the test data set after model training, and the rest as the experiment The training set is used to train the face recognition technology in the experiment. After training, use 20%, 25%, 30%, and 35% data to verify the effective recognition effects of different face recognition technologies.

C. *Experimental result*

The time-consuming situation of the two face recognition technologies to recognize the face data set is shown in the following table.

TABLE I FACE RECOGNITION TECHNOLOGY TIME-CONSUMING/S

Experimental data set	Test data size	Text recognition technology	Contrast recognition technology
ORL	20%	6.68	15.17
ORL	25%	7.03	15.93
ORL	30%	5.25	18.61
ORL	35%	6.14	15.09
IFD	20%	6.89	15.52
IFD	25%	7.12	14.54
IFD	30%	5.93	12.79
IFD	35%	7.55	12.62
Video screenshot	20%	5.56	18.98
Video screenshot	25%	4.29	17.86
Video screenshot	30%	5.63	18.25
Video screenshot	35%	7.48	18.8

Analysis of the above table shows that the face recognition technology studied in this paper takes less time to recognize different data sets than the comparison method to recognize the data sets. In this experiment, the average recognition time of the data set is calculated. The average recognition time of the technology in this paper is 6.29s, and the average recognition time of the contrast technology is 16.08s. The above data shows that the efficiency of the method in this paper is improved by about 1.5 times.

The specific data of precision and recall of face recognition technology for face dataset recognition are shown in the table below.

TABLE II COMPARISON OF RECOGNITION RECALL AND PRECISION OF FACE RECOGNITION TECHNOLOGY

Experimental data set	Test data size	Text recognition technology		Contrast recognition technology	
		Precision	Recall rate	Precision	Recall rate
ORL	20%	96.2	96.4	87.4	89.7
ORL	25%	97.1	95.6	89.4	90.8
ORL	30%	93.5	98.7	88.4	89.4
ORL	35%	94.8	95.5	87.6	88.7
IFD	20%	97.3	97.4	89.5	89.5
IFD	25%	94.1	97.0	87.5	90.5
IFD	30%	95.1	98.4	87.6	87.3
IFD	35%	94.2	95.8	89.7	90.1
Video screenshot	20%	93.9	97.2	88.4	88.2
Video screenshot	25%	94.2	95.8	89.8	89.1
Video screenshot	30%	94.4	95.5	88.5	90.5
Video screenshot	35%	93.7	95.6	89.9	89.7

Analyzing the above figure, we can see that the accuracy and recall of face recognition by the recognition technology studied in this paper are both higher than the comparison method. According to the definition of precision and recall, the face recognition technology studied in this paper can recognize face data more comprehensively, that is, the face recognition technology studied in this paper is more adaptable. In summary, the face recognition technology based on deep learning studied in this paper can effectively improve the efficiency of face recognition and ensure the stability of face recognition.

IV CONCLUDING REMARKS

Face recognition technology has become the preferred method of social identification information, with greater use value and potential. In this paper, face recognition technology based on deep learning is studied, and the effectiveness of the method is verified by simulation experiments. In the future research, it is necessary to increase the experimental training of occlusion and light change to improve the robustness of recognition.

REFERENCES

[1] HU Yuping. Cross-Age Face Recognition Algorithm Based on HOG-CSLBP and Deep Learning [J]. Journal of Southwest China Normal University(Natural Science Edition),2020,45(03):115-120.

[2] ZHU Yongzhi, SU Xiaoyun. Face Recognition Technology based on Multi-Task Convolutional Neural Network [J]. Communications Technology,2020,53(03):718-723.

[3] LI Huiyong, HE Zhuofeng. Face Recognition Technology for Specific Foreigners [J]. Modern Informationn Technology, 2019, 3(23):71-73.

[4] YANG Han. Research on face recognition algorithm based on deep learning [J]. Electronic Design Engineering,2019,27(19):155-159.

[5] LIU Liang. Research on face recognition technology based on PCA and LDA improved algorithm [J]. Wireless Internet Technology, 2019,16(17):110-111.

[6] ZHENG Minjie. Face Recognition Technology in Video Monitoring System [J]. Applications of IC,2019,36(09):37-39.

[7] ZHANG Jiwei, NIU Shaozhang, CAO Zhiyi, et al. Face Recognition via Deep Learning and Constraint Sparse Representation [J]. Transactions of Beijing Institute of Technology, 2019, 39(03): 255-261.

[8] ZHENG Jian, WANG Zhiming, ZHANG Ning. Improved Face Recognition Method Based on Deep Learning [J]. Computer and Modernization,2018(12):90-95.

[9] HU Jun. Age-Invariant Face Recognition Based on Deep Learning [J]. Modern Computer,2018(32):35-37.

[10] Zheng P, Shuai L, Arun S, Khan M. Visual attention feature (VAF) : A novel strategy for visual tracking based on cloud platform in intelligent surveillance systems. Journal of Parallel and Distributed Computing, 2018, 120, 182-194.

[11] G. Sun, Z. Song, J. Liu, et al. ″Feature Selection Method based on Maximum Information Coefficient and Approximate Markov Blanket[J].″ Zidonghua Xuebao/acta Automatica Sinica, 2017, 43(5):795-805.

[12] Fu W, Liu S, Srivastava G. Optimization of big data scheduling in social networks[J]. Entropy, 2019, 21(9): 902.

[13] G. Sun, F. Lang, Y. Xue. ″Chinese Chunking Method based on Conditional Random Fields and Semantic Classes[J].″ Journal of Harbin Institute of Technology, 2011, 43(7):135-139.

[14] Mengye L, Shuai L. Nucleosome positioning based on generalized relative entropy, Soft Computing, 2019, 23: 9175-9188