Computer Science and Information Systems 18(4):1395–1409 https://doi.org/10.2298/CSIS200922028F

Face Recognition Based on Full Convolutional Neural Network Based on Transfer Learning Model

Zhongkui Fan[1] and Ye-peng Guan[1, 2]

[1]School of Communication and Information Engineering,
Shanghai University, 200444 Shanghai, China
{fanzkui, ypguan}@shu.edu.cn
[2]Key Laboratory of Advanced Displays and System Application,
Ministry of Education, 200444 shanghai, China

Abstract: Deep learning has achieved a great success in face recognition (FR), however, little work has been done to apply deep learning for face photo-sketch recognition. This paper proposes an adaptive scale local binary pattern extraction method for optical face features. The extracted features are classified by Gaussian process. The most authoritative optical face test set LFW is used to train the trained model. Test, the test accuracy is 98.7%. Finally, the face features extracted by this method and the face features extracted from the convolutional neural network method are adapted to sketch faces through transfer learning, and the results of the adaptation are compared and analyzed. Finally, the paper tested the open-source sketch face data set CUHK Face Sketch database(CUFS) using the multimedia experiment of the Chinese University of Hong Kong. The test result was 97.4%. The result was compared with the test results of traditional sketch face recognition methods. It was found that the method recognized High efficiency, it is worth promoting.

Keywords: transfer learning, convolutional neural network, face recognition, adaptive scale, optical face features

1. Introduction

With the rapid development of Internet technology, information technology has been fully integrated into people's lives. How to accurately and effectively confirm identity has become an urgent issue in the field of information security. Face recognition [1] Compared with other biometric technologies, such as: vein recognition[2], voice recognition, fingerprint recognition, gene recognition, iris recognition, etc., because of its intuitive, convenient, non-contact and other excellent features, it makes it useful in people's daily life and society. In terms of security, such as access control, image retrieval, automatic login and face payment, and criminal investigation, it plays an important role. Therefore, it has been a research hotspot in the field of computer vision for decades. In recent years, with the development of deep learning the popularity of GPUs, and the open source of large-scale face databases, the maturity of optical face recognition technology has made it reach The practical application of standards has been applied to all walks of life, so that we have entered the era of face brushing.

There are various channels for obtaining faces. Under normal circumstances, it can be obtained through surveillance cameras or cameras to take pictures. However, in some cases, it is not possible to obtain face pictures through technical means. Only the sketches of human faces can be drawn by the artist based on the images of witnesses. This situation appeared more in the field of criminal investigation, and optical face recognition technology could not solve this problem, so sketch face recognition technology was produced [9]. This technology is a new type of face recognition technology developed on optical face recognition technology. It has very important application value in the field of criminal investigation. In recent years, it has begun to rise in the field of face recognition.

Sketch face recognition is a major branch of heterogeneous face recognition technology. Its role is to match a person's optical face with its corresponding sketch face. Its contribution is: in most criminal investigation cases, the police optical photos of the suspect were not available. At this time, the optical face recognition method has failed, but drawing a sketched face through the description of a witness is undoubtedly the most effective method to determine its identity. In view of this problem, the idea of extracting optical face features from CNN and then adapting them to sketch face features through transfer learning is proposed to solve the problem of insufficient sketch face training samples.

The proposed method achieved superior performance on CUFS [3] data-set and the contributions are summarized as follows:

(1) Adaptive scale feature extraction: An adaptive scale feature extraction method is proposed, which can adaptively adjust the feature extraction scale according to feature sensitivity.

(2) Establish a full convolutional neural network based on transfer learning model by analysing the role of each layer of AlexNet in image classification and using the VGGFace transfer learning model for face photo-sketch recognition.

2. Related Work

Sketch face recognition has been pioneered by Uhl and Lobo [4], to match sketch to photos, the proposed method uses Eigenface and Principle Component Analysis (PCA). Based on CUFS and CUFSF datasets, many state-of-the-art methods have been proposed by many researchers. Klare and Jain [5] proposed a method that extract the feature locally using a Scale Invariant Feature Transform (SIFT) descriptor. To improve the accuracy further, this method has been extended by fusing Multiscale Local Binary Pattern (MLBP) and the SIFT with Local Feature Discriminant Analysis (LFDA) [6]. Galoogahi and Sim [7] see that most of the research works use common features that are not meant for a cross-modality matching. Therefore a new face descriptor called Histogram of Averaged Oriented Gradients (HAOG) is proposed to extract modality-invariant features of salient facial components. The fact that facial shape is relatively invariant across modality, thus, Galoogahi and Sim [8] proposed a new face descriptor that is a shape-based and claimed to work on cross images. It is called Local Radon Binary Pattern (LRBP). The algorithm projects each non-overlapping patch on Radon space using Radon transform. Then, the features are extracted at every patch using Local

Binary Pattern (LBP). Recently, Difference of Gaussian Oriented Gradient Histogram (DoGOGH) has been demonstrated to be very effective in matching facial sketch to photo [9]. To cater for shape exaggerations effects, this method has been extended to Cascaded Static and Dynamic DoGOGH (C-DoGOGH) with intention to further improve the retrieval rate accuracy by catering the shape exaggerations effects [10]. It combines static and dynamic local featur extraction in a cascaded fashion.

Few works have considered the use of deep learning for face photo-sketch synthesis and recognition, most notable being the approaches in [11, 12, 13, 14]. However, these systems generally use relatively shallow networks or are primarily trained using images residing in a single modality (typically face photos).

Finally, few works consider the use of multiple sketches per subject. Most relevant to this letter is the work done in [15], [16], but the number of subjects and sketches used were both limited since the latter were manually created by employing several artists or software operators, making the process costly and time-consuming. These problems are critical, especially in the time-sensitive nature of real-world criminal investigations.

3. Adaptive Scale Feature Extraction

3.1. Feature Extraction

Extracting facial features needs to meet two basic requirements: first, to find the optimal features so that it can distinguish different faces to the greatest extent. Second, the extracted feature dimensions should be as small as possible, so that the training and testing speed can be improved. Before convolutional neural networks, the mainstream facial feature extraction algorithms were SIFT, LBP, HOG, and Gabor [17]. Theoretically speaking, the larger the extracted feature dimensions, the higher the accuracy. Literature pointed out that the use of multi-scale extraction of face features can improve the accuracy of the algorithm to a certain extent, but it will significantly increase the amount of calculation and is not conducive to engineering implementation. Based on this, an adaptive scale feature extraction method is proposed to reduce feature dimensions and improve the computing speed [18, 19]. Feature localization is the first step of feature extraction. It specifies the location of feature extraction, uses the face database with labels as training samples, learns the gradient direction of each label position, and then determines feature extraction based on the learned gradient direction position. Given an image $d \in R_{m\times 1}$ with m pixels, which contains p manually labeled points, as shown in Figure 1 (a), h is a feature extractor (herein specifically referred to as AMLBP), which uses labeled the human face is used as a training sample, and let it be x_*. When a human face is detected, an initialization mark x_0 is given as an average mark, as shown in Fig. 1 (b). Face feature localization can be achieved by minimizing the expression (1).

$$f\left(x_0+\Delta x\right)=\left\|h\left[d\left(x_0+\Delta x\right)\right]-\phi_*\right\|_2^2 \tag{1}$$

Where $\phi_* = h\left[d\left(x_*\right)\right]$ represents the AMLBP feature value at the face mark in the template library, and Δx represents the iteration step size. Training starts at x_0 and converges at x_*. To use the gradient method to differentiate, Taylor (1) is Taylor-expanded at x_0:

$$f\left(x_0+\Delta x\right) \approx f\left(x_0\right)+J_f\left(x_0\right)^T+\frac{1}{2}\Delta x^T H\left(x_0\right)\Delta x \tag{2}$$

Where $J_f\left(x_0\right)$ and $H\left(x_0\right)$ are Jacobian and Hessian matrices. Find the partial derivative of Δx in Equation (2) and make its derivative zero, and then get the first update of x:

$$\Delta x_1 = -H^{-1}J_f = -2H^{-1}J_h^T\left(\phi_0-\phi_*\right) \tag{3}$$

In the first gradient iteration process, the above formula is regarded as the projection of $\Delta\phi=\phi_0-\phi_*$ on the matrix $R_0=-2H^{-1}J_h^T$. In order to avoid calculating the Hessian matrix and the Jacobian matrix, take R_0 as the gradient direction, and R_0 can be directly obtained by learning the linear regression between $\Delta x_* = x_* - x_0$ and $\Delta\phi_0$. After simplifying equation (3), we can get:

$$\Delta x_1 = R_0\phi_0 + b_0 \tag{4}$$

Where b_0 is the offset. For a specific image, use Newton's method to update along the gradient direction:

$$x_k = x_{k-1} - 2H^{-1}J_h^T\left(\phi_{k-1}-\phi_*\right) \tag{5}$$

Equation (5) uses the R_{k-1} and b_{k-1} learned in the previous step to determine the new iteration position x_k:

$$x_k = x_{k-1} + R_{k-1}\phi_{k-1} + b_{k-1} \tag{6}$$

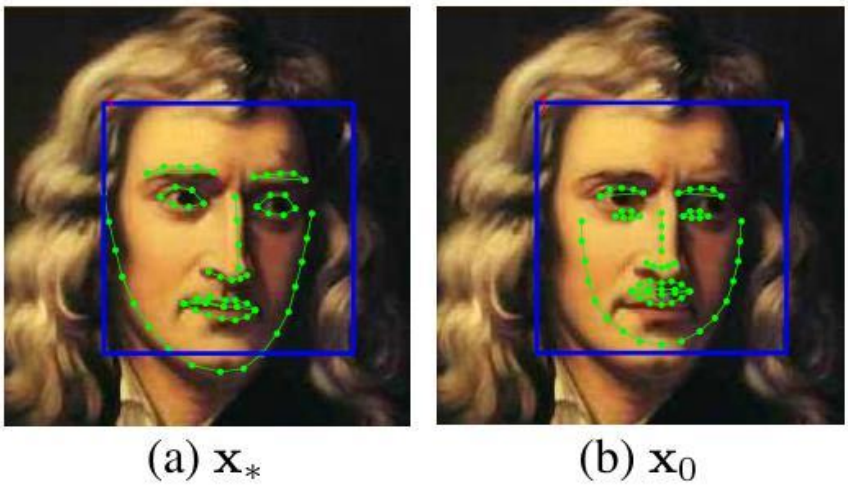

Fig. 1. a is a manually labelled face, b is a mark initialized using a face detector.

When a human face is detected, it can be positioned according to the gradient direction. The positioning effect is shown in Figure 2.

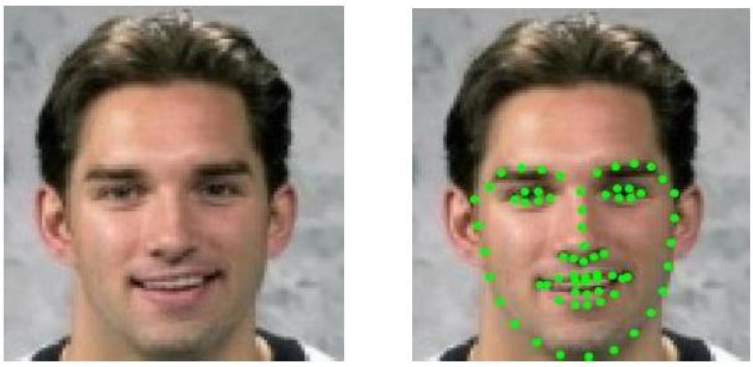

Fig. 2. a is the original face, b is the face after feature localization.

3.2. Adaptive Scale Feature Extraction

Taking into account the different amounts of information contained in different feature parts of the face, for example: the features contained in the eyes are the most sensitive and contribute the most to face recognition. Therefore, the feature extraction scale of this part should be the largest when feature extraction is performed; It is small, so its feature extraction scale should be appropriately reduced, so as to achieve the effect of reducing the feature dimension without reducing the amount of information. Based on

this, this paper proposes an adaptive scale feature extraction method, which can adaptively adjust the feature extraction scale according to feature sensitivity.

In order to obtain the sensitivity of the facial features, the facial features trained by Adaboost [20] and the features extracted by the adaptive scale are mapped, and then based on the error rate of each weak classifier (this is regarded as sensitivity) to determine each feature extraction scale. Figure 3 shows the experimental results, from which it can be analyzed that the classification effect is best when the feature extraction scale is greater than 4.

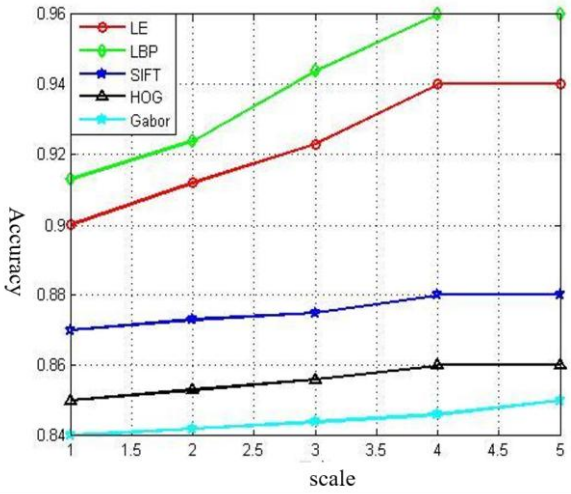

Fig. 3. Comparison of feature recognition results extracted at different scales.

Based on the experimental results in Figure 3, the maximum scale of adaptive feature extraction is set to 4, and the 200 facial features trained by the Adaboost algorithm are divided into four categories $K=\{k_1,k_2,k_3,k_4\}$, and the features are ranked from low to high according to the error rate. The features extracted by the adaptive scale are classified into four categories. The classification method is: use the Euclidean distance to search for features at the location of the feature to be classified, and the category of the feature closest to it is the category to which this feature belongs.

$$\min_{k(h(x))} D\left(p\{h(x)\}\right), p\{f(x)\}, k\{h(x)\} = k\{f(x)\} \tag{7}$$

Face classification is performed using a Gaussian process combined with a spectral mixed kernel function. The feature extraction scales are adaptive scale and single scale. The results are shown in Figure 4, which can be analyzed from the graph. The accuracy will also increase, but its feature dimensions will also increase, and the computing efficiency will decrease. The classification accuracy obtained using adaptive scale is the same as that of $k=4$ and $k=5$. The average accuracy is 2.6, which significantly reduces the feature dimension. Table 1 is the adaptive scale mapping, in which the feature mark numbers correspond to the face feature mark numbers in FIG. 5.

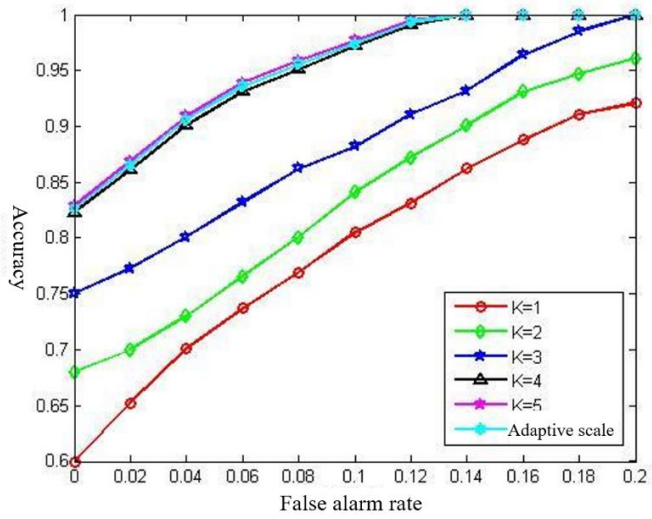

Fig. 4. Feature extraction effect of adaptive scale and single scale.

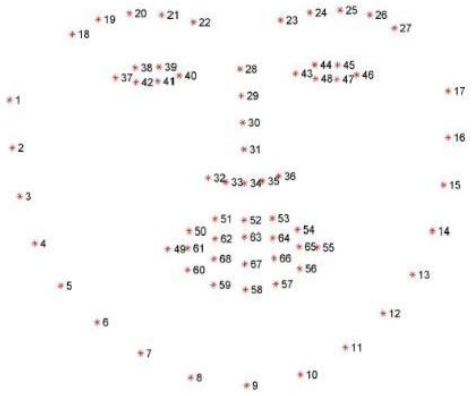

Fig. 5. Face feature labeling.

Table 1. Feature extraction scale mapping

Number of feature points	Feature label	scale
17	1-17	1
10	18-27	3
4	28-31	4
5	32-36	3
12	37-48	4
20	49-68	2

4. Convolutional Neural Network

The research on Convolutional Neural Networks (CNN) began in 1962. After Hubel and Wiesel proposed the receptive field by studying the visual cortex of cats, in 1984, Fukushima proposed the concept of cognitive machines based on receptive fields. Machine is the first application of receptive field on neural network, which can be regarded as the first implementation of convolutional neural network. After many scholars in the later period of innovation, convolutional neural network has become the most representative network in deep learning. Convolutional neural networks have achieved great success in the image field, and CNN models have achieved excellent results in visual competitions based on the ImageNet [21] dataset over the years. Compared with traditional vision algorithms, CNN has the advantage of directly extracting features from the original image without pre-processing the image [22], thereby avoiding the problem of image information loss during the pre-processing stage.

Figure 6 shows the convolutional neural network structure. The network structure has 5 convolutional layers and 3 fully-linked layers. The final output layer uses the Softmax function to output 1000 classes. Equation 8 is the Softmax principle, with K output categories. The calculation process is:

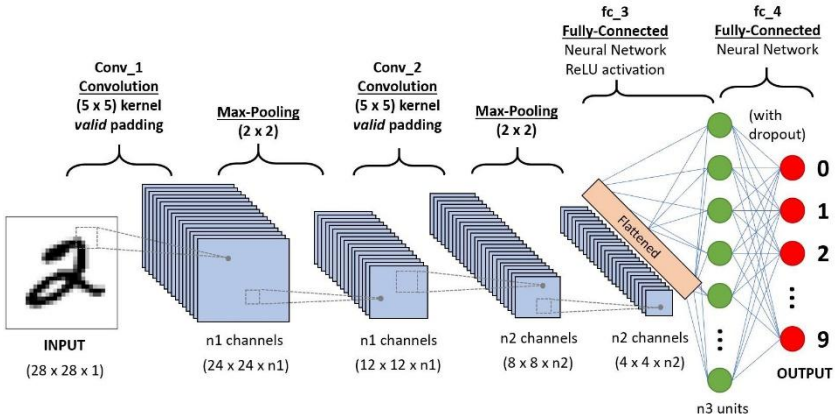

Fig. 6. Development stages of convolutional neural networks.

$$soft\max\left(a_i\right)=\frac{\exp\left(a_i\right)}{\sum_j \exp\left(a_j\right)},\ j=0,1,2,...,k-1 \tag{8}$$

Softmax maps the output of multiple neurons to the [0,1] interval when performing multi-classification, so it can be understood as a probability. The output of Softmax is equivalent to the probability distribution of the picture classified into each category. This function is a monotonically increasing function. The larger the input value, the greater the probability that the input image belongs to the category label.

$$b_{x,y}^{i} = a_{x,y}^{i} / \left(k + a \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} \left(a_{x,y}^{j} \right)^2 \right) \tag{9}$$

In the formula, $a_{x,y}^{i}$ is the i-th convolution of the feature at the (x, y) position in the input feature, and then the result is obtained through ReLU. $b_{x,y}^{i}$ is the corresponding normalized result. N is the total number of convolutions, and $k = 2, n = 5, \alpha = 10^{-4}, \beta = 0.75$ is the hyperparameter.

The network parameters are constantly adjusted when training the network. The changes in the network parameters at each layer will cause the input feature distribution of the latter layer to change. This phenomenon is called internal covariance change. However, it is necessary to adapt the parameters of each layer to the input when learning. Feature distribution. To this end, the data needs to be normalized. The purpose of normalization is to make the data mean 0 and unitized variance. The expression is as follows:

$$\hat{x}^{(k)} = \frac{x^{(k)} - E\left(x^{(k)} \right)}{\sqrt{Var\left[x^{(k)} \right]}} \tag{10}$$

However, this method will reduce the expressive ability of the convolution layer. For example: using sigmoid as the activation function. This method will limit the data to around 0 mean. Then, only the linear part of the activation function is used and the function of the non-linear part is not used. Makes the network expression ability poor. Therefore, the normalized data needs to be further processed to maintain the expressive power of the model. Its formula is as follows:

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \tag{11}$$

In theory, all the data must be processed every time, but the amount of data processed by the convolutional neural network is huge. Each time all the data is processed will significantly increase the amount of calculation, so the data is processed in batches. Let each batch of input data be: $B = \{x_1, x_2, ..., x_m\}$. The output is: $y_i = BN_{\gamma,\beta}(x_i)$.

$$\begin{aligned} \mu_B &= \frac{1}{m} \sum_{i=1}^{m} x_i \qquad \sigma_B^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2 \\ \hat{x}_i &= \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \qquad y_i = \gamma x_i + \beta = BN_{\gamma,\beta}(x_i) \end{aligned} \tag{12}$$

In order to extract signal features more comprehensively, the research direction of deep learning from 2014 has mainly focused on building deeper network structures, but increasing the model will reduce the network computing efficiency. Making use of computing performance has become the focus of research.

5. Face Recognition For Deep Learning Transfer Components

Since AlexNet won the championship in the ImageNet competition in 2012, deep learning has received unprecedented high attention in the field of machine learning. In recent years, a large number of papers on deep learning have emerged, but until now, deep learning is still a black box. You can't feel it, and you can't explain and deduce it with theory. Because CNN has a good hierarchical structure, this article analyzes its mobility by using CNN's feature extraction rules. The regularity of CNN's face feature extraction is shown in Figure 7.

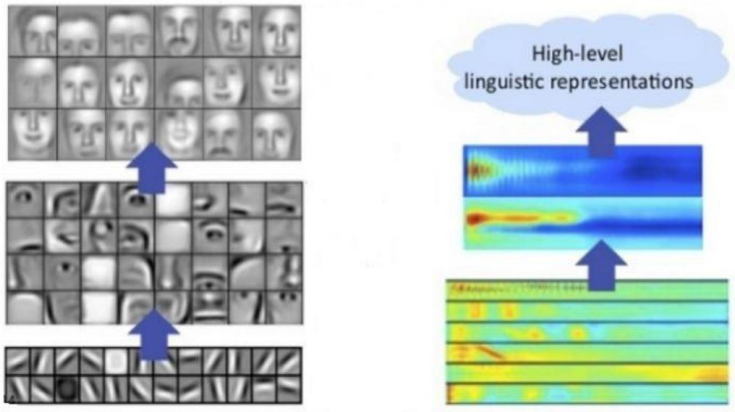

Fig. 7. Neural network for facial feature extraction.

This article trains the AlexNet network on the Pytorch framework, using the ImageNet training set, which has a total of 1000 classes. In the experiment, it was divided into A and B, each of 500 types. The AlexNet network has a total of 8 layers. Except that the 8th layer is a classification layer, this article analyzes the layers 1 to 7 layer by layer. The analysis method is: take the data of category B as the reference standard, fix the first *n* layers of network A, then initialize the remaining 8-*n* layers, and then classify B. This process is called AnB. The corresponding BnB is to fix the first n layers of the trained B network, initialize the remaining 8-n layers, and then classify the type B data. This process is called BnB. AnB and BnB network structure is shown in Figure 8.

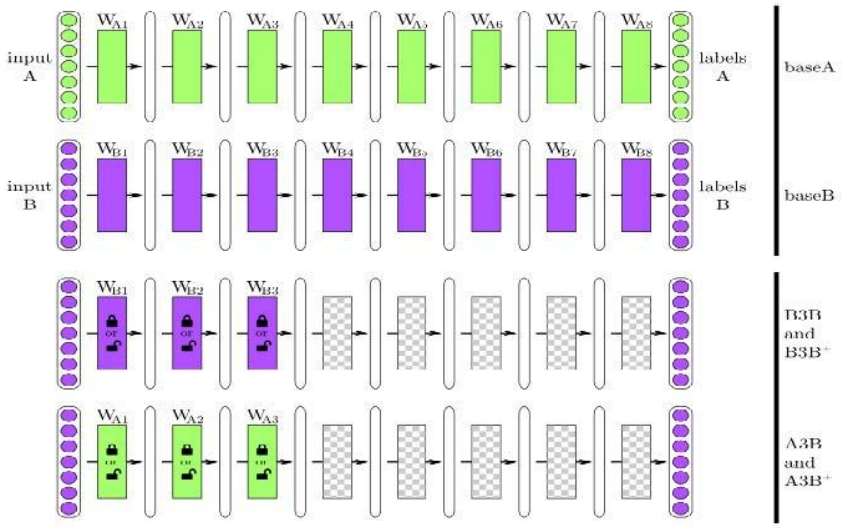

Fig. 8 AnB and BnB network structure.

Figure 9 shows the results of AnB and BnB experiments. From the figure, it can be analyzed that for BnB, the first 3 layers of the trained model are tested and there will be no loss of model accuracy. At the 4th and 5th layers, the accuracy is reduced but It's not too bad. The accuracy has obviously improved by the sixth and seventh layers. The reason is that the fourth and fifth layers are in the back part of the network. The extracted features are relatively specific, so the accuracy will decrease when the training samples change. Layers 6 and 7 learn again on the basis of abstract features to improve accuracy, which is exactly the effect required for transfer learning. For BnB + (BnB plus fine-tune), the whole result is not changed, which shows that fine-tune can promote the model well. The migration effect of AnB and AnB + is more convincing, because this is the migration of two different network-trained models. For AnB, the first 3 layers of network A are migrated to network B, but its accuracy is not affected. This proves once again that the first three layers of the network have learned abstract features, and the accuracy has begun to decline when it migrates to the fourth to fifth layers, which shows that the deeper the network, the more specific the extracted features. However, the accuracy of the 6th to 7th layers has decreased after a slight improvement. This is because the features of the 6th to 7th layers are not updated, the learning ability is poor, and the features extracted from these layers are the most specific. With the addition of fine-tune to the AnB + network, all layers performed very well and even exceeded baseB.

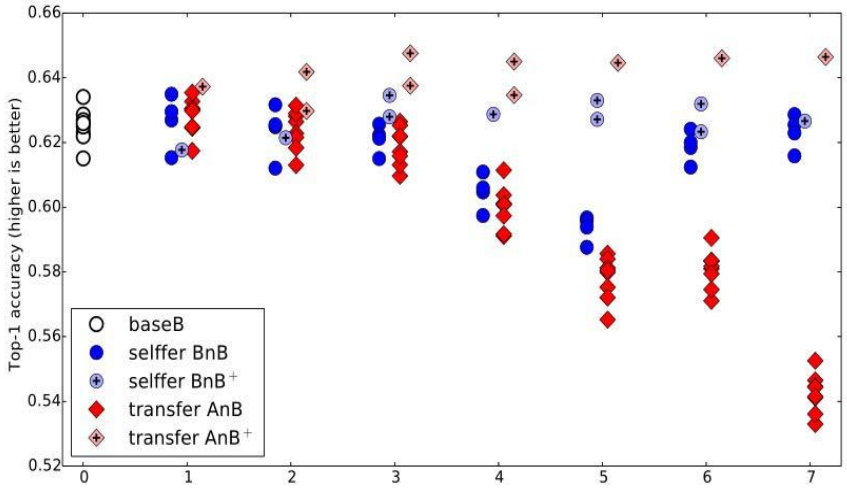

Fig. 9. AnB and BnB experimental results.

In order to exclude the accidental grouping of ImageNet data, that is, the data in the two groups are similar, for example, there are dogs in class A and dogs in class B. Then this will cause B to get better results when migrating A, Re-classify to ensure that there are no similar pictures in groups A and B, and repeat the above experiment to achieve the same results as the previous one. Figure 10 shows the experimental results.

It can be concluded from Figure 10 that the deeper the CNN network, the more specific the features learned, resulting in the performance of the model decreasing with the deepening of the network layer where the features are migrated, so the previous abstract feature layer is more suitable for migration. In addition, adding fine-tune to the transfer learning overcomes the differences between the data and can significantly improve the accuracy and help network migration.

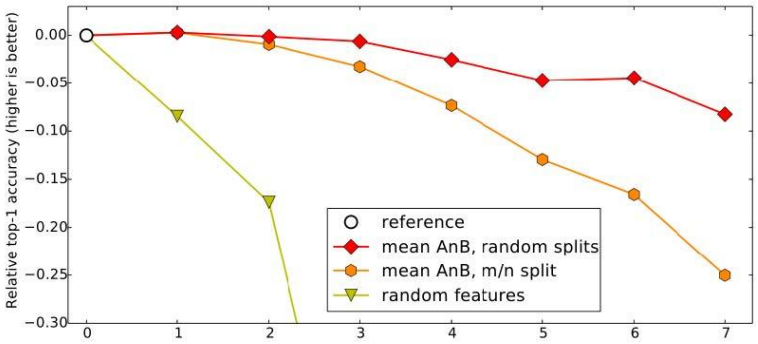

Fig. 10. AnB and BnB experimental results after removing similarity of samples.

Through experiments comparing the effects of network migration in the middle of the three frames, it was found that the optical face features extracted on the VGGFace framework and the sketch face features are best integrated. Therefore, this article divides

the VGG16 network into the first, middle, and last three parts. The first part is the first to fourth layers of the network, the middle part is the fifth to tenth layers of the network, and the second part is the eleventh to sixteenth layers of the network. The facial features correspond to sketch faces, and finally they are adapted using JDA. Finally, the sketch faces and corresponding optical faces are used to train and test the model.

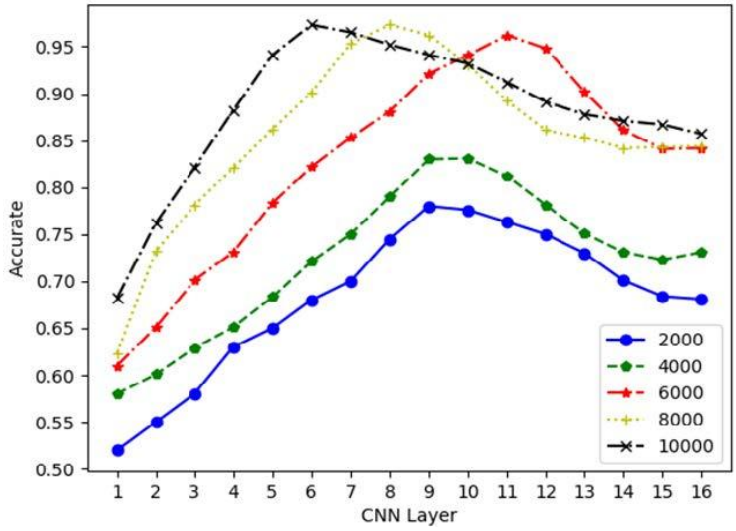

Fig. 11. The effect of the number of sketch face samples on the accuracy of the model.

From the experimental results shown in Figure 11, it can be seen that the features extracted in the middle part of the network are best suited for sketching the face. It is verified again that the face features extracted in the middle part of the CNN are most suitable for migration. The main role of this figure is that it proves that By extracting the optical face features from the CNN and adapting them to the sketch faces, it can effectively reduce the training samples of the sketch faces. Comparing Fig. 11, it can be seen that when the optical face features are not extracted from the CNN, 10,000 sketch faces are trained with the highest accuracy. It can reach 78.2%. After extracting the optical face features from CNN to match the sketch face, the accuracy of 2,000 sketch faces can reach 75.1%, and the accuracy can reach more than 82.3% when the number of sketch faces is 4,000. When the number of sketched faces reaches 6,000, the accuracy can reach 95.2%. When the number of sketched faces continues to increase to 8,000 and 10,000, the accuracy increase is not obvious, and the highest can only reach 97.4%, which indicates the success of this paper. The problem that the accuracy of the sketch face cannot be increased due to insufficient training samples for the sketch face is solved.

6. Conclusion

In this paper, an adaptive scale feature extraction method is proposed to build the sketch face training sample. Optical face features are extracted from the central network layer of CNN. Good results have been achieved through JDA [23] and sketch face adaptation. The performance of this method is analyzed through test results. It is found that the use of JDA + CNN can make sketch face recognition accuracy. It reaches about 97.4%, and the accuracy has been slightly improved compared with the traditional sketch face recognition algorithm. Its outstanding performance is reflected in the fact that it can effectively reduce the sketch face training samples. By analyzing the role of each layer of the convolutional neural network, reveal the basic principle of convolutional neural network, point out the direction for transfer learning.

Acknowledgment. This research work was supported in part by National Key R&D Program of China (2019YFC1520500,2020YFC1523004).

References

1. Masi, Iacopo, et al. Deep face recognition: A survey.2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE, 471-478. (2018)
2. Yang, Lu, et al. "Finger vein recognition with anatomy structure analysis." IEEE Transactions on Circuits and Systems for Video Technology, 1892-1905. (2017)
3. X. Tang and X. Wang, Face sketch synthesis and recognition, Ninth IEEE International Conference on Computer Vision, 687–694, (2003).
4. R. U. Jr and N. d. V. Lobo, A framework for recognizing a facial image from a police sketch, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 586–593, (1996).
5. B. Klare and A. K. Jain, "Sketch to Photo Matching: A Feature-based Approach, " Proc. SPIE Conference on Biometric Technology for Human Identification VII, (2010)
6. B. Klare, Z. Li, and A. K. Jain, Matching forensic sketches to mug shot photos, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 3, 639–646,(2011).
7. H. K. Galoogahi and T. Sim, Inter-modality face sketch recognition, IEEE International Conference on Multimedia and Expo, 224–229,(2012).
8. H. Kiani Galoogahi and T. Sim, Face sketch recognition by Local Radon Binary Pattern: LRBP, Proceedings - International Conference on Image Processing, ICIP, 1837–1840, (2012).
9. S. Setumin and S. A. Suandi, "Difference of gaussian oriented gradient histogram for face sketch to photo matching, " IEEE Access, vol. 6,39344–39352, (2018).
10. S. Setumin and S. A. Suandi, Cascaded static and dynamic local feature extractions for face sketch to photo matching, IEEE Access, vol. 7, 27135–27145, (2019).
11. P. Mittal, M. Vatsa, and R. Singh, Composite sketch recognition via deep network – A transfer learning approach, in Proc. Int. Conf. Biometrics, 251–256. (2015).
12. L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang, End-to-end photo-sketch generation via fully convolutional representation learning, in Proc. ACM Int. Conf. Multimedia Retrieval, New York, NY, USA. 627–634.(2015)
13. S. Saxena and J. Verbeek, Heterogeneous face recognition with CNNs, in Proc. Eur. Conf. Comput. Vis. 483–491. (2016)

14. D. Zhang, L. Lin, T. Chen, X. Wu, W. Tan, and E. Izquierdo, Content adaptive sketch portrait generation by decompositional representation learning, IEEE Trans. Image Process., vol. 26, no. 1, 328–339, (2017) .
15. C. Peng, N. Wang, X. Gao, and J. Li, Face Recognition from Multiple Stylistic Sketches: Scenarios, Datasets, and Evaluation. Cham, Germany:Springer, 3–18. (2016)
16. X. Gao, N. Wang, D. Tao, and X. Li,Face sketch-photo synthesis and retrieval using sparse representation IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 8, 1213–1226, (2018).
17. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." computer vision and pattern recognition: 886-893. (2005)
18. Zhang, C., Zhao, X., Cai, M., Wang, D., Cao, L.: A New Model for Predicting the Attributes of Suspects. Computer Science and Information Systems, Vol. 17, No. 3, 705-715. (2020),.
19. Chen, H., Dai, Y., Gao, H., Han, D., Li, S.: Classification and Analysis of MOOCs Learner's State: The Study of Hidden Markov Model. Computer Science and Information Systems, Vol. 16, No. 3, 849–865. (2019),
20. Shuai Di, Honggang Zhang, Chun-Guang Li, Xue Mei, Danil Prokhorov, & Haibin Ling. (2018) 'Cross-domain traffic scene understanding: a dense correspondence-based transfer learning approach.' IEEE Transactions on Intelligent Transportation Systems, Vol. 19 No.3, pp.45-757. (2017)
21. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., & Ma, S., et al.. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252. (2015)
22. Zhang, D., et al., The generative adversarial networks and its application in machine vision. Enterprise Information Systems,: p. 1-21. (2019)
23. Chen, D., Ren, S., Wei, Y., Cao, X., & Sun, J.. Joint Cascade Face Detection and Alignment. European Conference on Computer Vision. 109-122. (2014)

Zhongkui Fan receive the BS degree in Computer engineering from Huanghe science and technology College, and the MS degree in Computer Engineering from Jiangxi University of Science and Technology. He is currently working toward the PhD degree in the School of Communication and Information Engineering with Shanghai University. His research interests include computer vision and machine learning.

Ye-peng Guan was born in Xiaogan, Hubei Province, China, in 1967. He received the B.S. and M.S. degrees in physical geography from the Central South University, Changsha, China, in 1990, 2006, respectively, and the Ph.D. degree in geodetection and information technology from the Central South University, Changsha, China, in 2000. From 2001 to 2002, he did his first postdoctoral research at Southeast University in electronic science and technology. From 2003 to 2004, he did his second postdoctoral research at Zhejiang University in communication engineering, and he had been an Assistant Professor with the Department of Information and Electronics Engineering, Zhejiang University. Since 2007, he has been a Professor with School of Communication and Information Engineering, Shanghai University. He is the author of more than 120 articles, and more than 20 patents. His research interests include intelligent information perception, digital image processing, computer vision, and security surveillance and guard.

Received: September 22, 2020; Accepted: April 04, 2021.