A Survey of Face Recognition Methods based on Deep Learning

Beiwen Chen [a], Xiaofang Liao [b], Haolin Zhu [c], Zhuoxian Gong [d], Yingcong Li [e]

School of Computer Science, Guangdong University of Foreign Studies South China Business College, Guangzhou 510545, China

[a] 1427053342@qq.com, [b] 52626351@qq.com, [c] 1139520364@qq.com, [d] 2062875698@qq.com, [e] 2017787109@qq.com

Abstract. Since the 21st century, people have put forward higher requirements for information security technology in many fields such as society and economy. The requirement has promoted the development of biometric identification technology. Face recognition technology has become a research hotspot in biometric identification technology with its unique advantages. In recent years, with the development of deep learning, face recognition technology has made breakthroughs. Face recognition technology based on deep learning has been widely used in various fields such as finance, education, security, transportation, and new retail. In the process of face recognition technology becoming popular, some comprehensive literatures are urgently needed to summarize the methods of face recognition technology. Based on this, this paper first introduces the principle and existing problems of traditional face recognition methods, and then introduces in detail two typical face recognition methods based on deep learning—face recognition methods based on convolutional neural networks and face recognition methods based on deep belief networks. Finally, we provide an overview of common face datasets.

Keywords: Face Recognition; Convolutional Neural Network; Deep Belief Network; Face Dataset.

1. Introduction

In the era of big data, the information is redundant, the explosive growth of data has caused serious problems such as information leakage, the confidentiality mechanism is useless, and the information age is facing unprecedented challenges. The emergence of biometric identification technology can effectively solve this problem. Biometric identification technology is to identify people by collecting biometric samples of people[1]. Face recognition is the process of finding a matching face by comparing the face image detected from the static image or dynamic video with the face image in the database. Face recognition is usually used for identification, and it is a subject in the field of biometric identification[2]. Compared with other biometric identification technologies such as fingerprint recognition, iris recognition, and gait recognition, face recognition has the characteristics of high efficiency, convenience, and easy popularity, and it has become one of the most concerned research hotspots. At present, face recognition technology is widely used in government, banking, social welfare, military, security and defense, e-commerce and other fields [3].

At present, although some research results have been achieved in face recognition technology, the recognition rate and accuracy of face recognition have been affected due to the complexity of factors such as face and environment. However, traditional face recognition methods have been unable to solve the problem of face recognition under complex factors. In recent years, with the development of deep learning, there has been a major breakthrough in face recognition research. This paper first introduces the principle and existing problems of traditional face recognition methods, and then introduces in detail two typical face recognition methods based on deep learning—face recognition methods based on convolutional neural networks and face recognition methods based on deep belief networks. Finally, we provide an overview of common face datasets.

2. Traditional Face Recognition Methods

Typical traditional face recognition methods include face recognition method based on model, face recognition method based on locality preserving pro−jection, face recognition method based on geometric features, face recognition method based on algebraic features and face recognition method based on sparse representation, etc.

(1) The face recognition method based on model is to compare the extracted face image with all template records in the database, and then select the most similar template as the classification of the image. However, this method has certain defects. First, the number of pictures in the database is limited, and there is a situation where no matching template can be found. Secondly, the method has strict requirements on background, expression, light, angle and other conditions. Therefore, this method is not suitable for real-world scenarios.

(2) The face recognition method based on locality preserving pro−jection. In order to overcome the defects of nonlinear methods, Professor He proposed a face recognition algorithm based on manifold learning—locality preserving pro−jection. The algorithm can accurately obtain the local feature structure of the image while reducing the spatial dimension, but it cannot extract the inherent feature structure of the original image[4].

(3) The face recognition method based on geometric features is an intuitive and simple face recognition method. The face recognition system based on this method mainly extracts the main geometric feature points of the face, the continuous shape of the main organs of the face, the curvature of geometric features and other information for recognition[5]. However, the recognition rate and reliability of this method are low because the shape and position of the main facial organs cannot accurately reflect the nonlinear factors in the face image.

(4) Compared with the face recognition method based on geometric features, the face recognition method based on algebraic features has the advantage of being inclusive to a certain degree of human expression changes and lighting. The face recognition method based on algebraic features obtains features through principal component analysis, linear discriminant analysis, etc., and performs algebraic change processing on the grayscale features of the image, or by decomposing the matrix formed by the face image. It adopts a dimensionality reduction strategy, which has a strong effect in the face of linear structures, but the recognition effect is not ideal in the face of nonlinear structures[1].

(5) The Face recognition method based on sparse representation. With the development of sparse representation, John Wright et al. proposed to apply sparse representation to face recognition, and mentioned that if sparse representation is correctly applied to face recognition, feature selection will become simple. However, how to correctly calculate the sparse representation is the key to solving the problem, and this method cannot perform active learning and must be set manually[4].

The traditional face recognition method has certain defects: some traditional face recognition methods must be helped by artificially set features, but the artificially set features often bring errors to the feature extraction and recognition process; Without human intervention, traditional face recognition methods often cannot automatically extract useful features from the original image; the features generated by traditional face recognition methods are identified as shallow features, and more in-depth high-semantic features cannot be obtained from the original image and deep features. Due to the influence of nonlinear factors and the constraints of the complexity of face recognition itself, these defects in traditional face recognition methods greatly reduce the accuracy of face recognition, so better methods are needed to solve such problems.

3. Face Recognition Method based on Deep Learning

Deep learning processes messages by simulating the human brain nervous system, and can deal with complex problems in face recognition. In the era of massive face data, face recognition based on deep learning has achieved the best results in terms of speed and accuracy, which is of great significance to the theoretical research of deep learning and the practical application of face

recognition[6]. The process of face recognition method based on deep learning mainly includes the steps of face preprocessing (face detection, face alignment, standardization, data enhancement, etc.), feature learning, feature comparison, etc. The key to face recognition is features learning. How to extract strong discriminative and robust features is the focus of face recognition research[7].

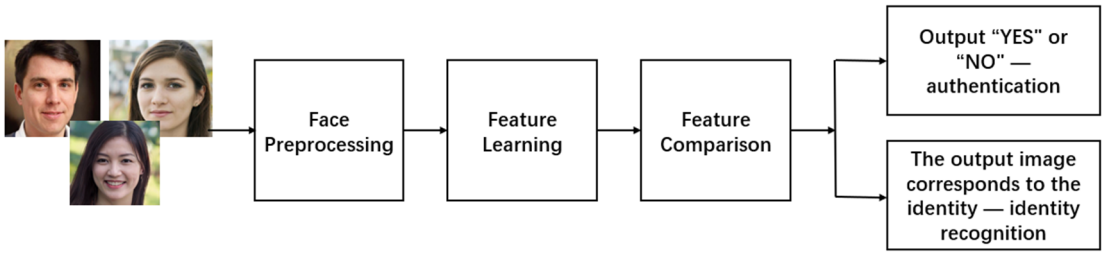

Fig 1. The training flow chart of face recognition model based on deep learning

At present, there are many deep learning models, among which the most important and typical two are convolutional neural network and deep belief network. Next, we will focus on the research and application of these two models in face recognition.

3.1 Face Recognition Method based on Convolutional Neural Networks

The convolutional neural network is derived from a multi-layer forward network, and automatically learns features through operations such as multiple convolutions, excitation functions, and pooling. The convolutional neural network is the first multi-layer network structure that has been successfully trained in the true sense. The more complex the data, the deeper the network structure[1]. Convolutional neural networks enable end-to-end learning, making training much simpler. Through forward and backpropagation, gradient descent, chain derivation feedback error signal, and finally adjusting the weights of each layer, a stronger generalization ability is achieved. The convolutional neural network does not need to preprocess the image, and can directly use the pixels of the image as the input of the network, which greatly reduces the complexity of data reconstruction in the process of feature extraction and classification in traditional face recognition methods. When multi-dimensional images are used as input, their performance is more obvious, which makes convolutional neural networks widely used in classification fields such as image recognition.

SUN[8] et al. proposed to apply DeepID to the face recognition neighborhood. The structure of DeepID is similar to that of a convolutional neural network, except that the hidden layer of the penultimate layer of DeepID is connected to the convolutional layer of the previous layer and the pooling layer before this convolutional layer. This connection makes it possible to consider both local and global features. The accuracy of this method on the LFW dataset reaches 97.45%. Baidu[9] proposed a method that combines muti-patch deep CNN and deep metric learning. The main idea of this method is to extract the features of the non-passing areas of the face through the convolutional neural network, and then reduce the extracted features to 128 dimensions through metric learning for identification. Finally, the accuracy of the method on the LFW dataset is 99.77%. SCHROFF[10] et al. proposed a method based on deep convolutional neural networks, which directly maps face images to Euclidean space, and the similarity of face images is represented by the distance in this space. Once the Euclidean space is formed, face recognition can be easily performed. The accuracy of this method on the LFW dataset is 99.63%.

3.2 Face Recognition Method based on Deep Belief Network

Deep belief network is a relatively widely studied deep learning model. Figure 2 is the structure of a deep belief network. The deep belief network is composed of a series of restricted Boltzmann machine units. The restricted Boltzmann machine is a bipartite graph model in which the nodes in the layers are independent of each other, and the joint probability distribution of the nodes in the layers satisfies the Boltzmann distribution[11]. The hidden layer unit and visible layer of a restricted

Boltzmann machine are connected to each other, there is no connection within the layer, and the hidden layer unit can obtain the high-order correlation of the input visible unit. Restricted Boltzmann machines fit the data well and are often used as nonlinear classifiers. An important advantage of deep learning compared with traditional neural networks is that it largely solves the training speed and accuracy problems of low-level neural networks. Before the global learning of the multi-layer model, the deep belief network will decompose the neural network into a cascade of multiple restricted Boltzmann machines, and then train them layer by layer[12]. In order to accurately describe the feature structure, the deep belief network can learn the abstract features of each layer from the bottom up. Different from the algebraic feature method, its feature extraction does not need to be manually selected, but is completely completed by automatic learning.

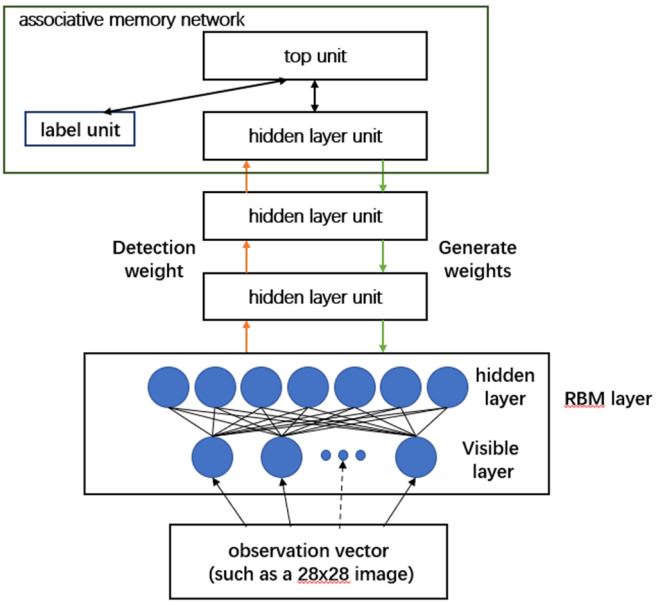

Fig 2. Structure of deep belief network

Yang[13] et al. proposed a face recognition method based on Gabor features and a deep belief network. By extracting different scale maps of Gabor face images for convolution fusion, the fused feature maps are used as the input data of the deep belief network. Multiple layers are used to obtain more abstract feature expressions. In the whole training process, cross entropy is used to fine-tune the deep belief network. The top layer of the model combines the Softmax regression classifier to classify the extracted features. The experimental results of the AR face database test show that the Gabor feature combined with the deep belief network is applied to face recognition, and its accuracy can be as high as 92.7%. Compared with other shallow learning models, the deep belief network learns the high-level data of the data. The feature also reduces the feature dimension, improves the classification accuracy of the classifier, and finally effectively improves the face recognition rate. Zhao[14]also proposed a face recognition method based on the combination of Gabor wavelet and deep belief network, which can effectively extract the abstract features of portraits, and reduce the influence of posture, light, etc. on the recognition rate, and realize the accurate identification of portraits.

4. Common Face Datasets

Face datasets play an important role in the study of face recognition methods, and the pros and cons of the datasets will affect the accuracy of the final results. This section will focus on common face datasets.

The LFW dataset[15-16] is a dataset created by the Computer Vision Lab at the University of Massachusetts at Armstrong, USA. The images of this dataset are mainly from The Yahoo website

and contain 13,233 images of 5,749 people. Since there are 4,069 people in this dataset with only one face image, this dataset is not used as a training set, but as a test set. The IJB-A dataset[17] contains 5,712 pictures and 2,085 videos of 500 people. The IJB-B dataset[18] and the IJB-C dataset[19] were subsequently developed on the basis of the IJB-A dataset. The IJB-B dataset contains 11,754 pictures and 7,011 videos of 1,845 individuals, including the IJB-A dataset. The IJB-C dataset contains the contents of the IJB-A dataset and the IJB-B dataset, including 138,000 face images and 11,000 videos. The YTF dataset[20] contains 3,425 videos of 1,595 individuals, all from YouTube video site. Video lengths ranged from 48 to 6070 frames, with an average length of 181.2 frames. A common task for this dataset is to determine whether the people in every two videos belong to the same identity. The VGGFace2 dataset[21] comes from Google and contains about 3 million face images of 9,131 people. The images cover different ages, poses, lighting, races, and occupations. In addition to identity information, the dataset also provides the face frame, 5 key points, and estimated age and pose of each face image in the dataset. The CASIA WebFace dataset[22] images come from the IMDb website, which contains 494,414 face images of 10,575 individuals. This dataset is widely used to train face recognition models based on deep convolutional neural networks[23-25]. The UMDFaces dataset[26-27] contains 367,888 images of 8,277 individuals and 22,075 videos of 3,100 individuals. The face information provided by this dataset includes face frame, pose estimation, 21 key points, and gender. At the same time, the dataset also provides three levels of face verification tests: easy, medium, and difficult. The test set for each level Contains 100,000 pairs of face images. The current common face datasets are shown in Table 1, which lists the dataset name, the number of identities it contains, and the total number of face images. The details of each dataset are as follows.

Table 1. Common face datasets

Data Set	Number of Identities	Number of Images
LFW[15-16]	5,749	13,233 pictures
IJB-A[17]	500	5,712 pictures, 2,085 videos
YTF[20]	1,595	3,425 videos
VGGFace2[21]	9,131	3.31M pictures
CASIA WebFace[22]	10,575	494,414 pictures
UMDFaces[26-27]	8,277	367,888 pictures, 22,075 videos

5. Analysis and Prospects

With the continuous development of the artificial intelligence era, the research and application of deep learning in face recognition continues to deepen. Deep learning has key advantages: it can detect complex interactions from features; it can learn low-level features from raw, unprocessed data; it can handle high-cardinality class data; and it can handle unlabeled data. With these advantages, deep learning can learn more useful data and build more accurate models. However, deep learning also has shortcomings. The training time of the model is relatively long, and continuous iteration is required to optimize the model, and the global optimal solution cannot be guaranteed. At the same time, deep learning needs to rely on a powerful data set. In the application scenario of small data sets, the deep learning algorithm cannot make unbiased estimation of the laws of the data, and the face recognition effect may not be as good as some existing traditional face recognition algorithms. Therefore, it is very important to start from the shortcomings of deep learning and improve the accuracy of the algorithm. The solution to the problem needs to rely on our continuous exploration in the future.

Acknowledgments

This article was supported in part by the 2020 Guangdong University Student Innovation and Entrepreneurship Training Program (No.202012620002) and The Nanguo Business School of Guangdong University of Foreign Studies -- practical teaching base of Guangzhou Tengke Network Technology Co., Ltd.

References

[1] LI LinglI. A Review of The Application of Face Recognition Technology Based on Deep Learning Theory.Computer and Digital Engineering.Vol.49(2021) No.9, p.1912-1914+1929.

[2] WU Jiny, ZHOU Delong. A Survey of Face Recognition Methods. Application Research of Computers. Vol.26 (2009) No.9, p.3205–3209.

[3] LIU Weikai, HAO Yaqian, ZHENG Han, et al. Overview of Face Recognition. Information Recording Material.Vol.(2018) No.7,p.13-14.

[4] Information on: https://blog.csdn.net/jacktangyao/article/details/80783242.2018.

[5] ZHUANG Xinlin. A Review of Face Recognition Methods .Science and Technology Innovation and Application.Vol.12(2022)No.2,p.130-132.

[6] YANG Jucheng, LIU Na, FANG Shanshan, et al. Review of Face Recognition Methods Based on Deep Learning. Journal of Tianjin University of Science & Technology.Vol.31(2016) No.6, p.1-10.

[7] XU Cuican, LI Huibin. A Review of Face Recognition Methods Based on Deep Learning. Journal of Engineering Mathematics.Vol.38(2021) No.4, p.451-469.

[8] SUN Y, WANG X, TANG X. Deep Learning Face Representation from Predicting 10,000 Classes. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society,2014, P.1891-1898.

[9] Information on: https://arxiv.org/abs/1506.07310.2015.

[10] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: Aunified Embedding for Face Recognition and Clustering. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015, P.815-823.

[11] Zhang Yuanyuan, Huo Jing, Yang Wanqi, Gao Yang, Shi Yinghuan. Two-Generation ID Card Heterogeneous Face Verification Algorithm Based on Deep Belief Network. Journal of Intelligent Systems. Vol. 10(2015) No.2, p.193-200.

[12] LIN Miaozhen: Research on Face Recognition Based on Deep Learning (Dalian University of Technology, China 2013).

[13] YANG Rui, ZHANG Yunwei, GOU Shuang, et al. Face Recognition Algorithm Based on Gabor Feature and DBN. Transducer and Microsystem Technologies.Vol.36(2017) No.5, p.68-70.

[14] ZHAO Yuandong: Research on Face Recognition Method Based on Deep Neural Network. (Shenyang University of Technology, China 2019).

[15] HUANG G B, RAMESH M, BERG T, et al. Labeled Faces in The Wild: A Database for Studying Face Recognition in Unconstrained Environments[R]. University of Massachusetts, Amherst, 2007: 7-49.

[16] HUANG G B, LEARNED-MILLER E. Labeled Faces in The Wild: Updates and New Reporting Procedures [R]. University of Massachusetts, Amherst, 2014.

[17] KLARE B F, KLEIN B, TABORSKY E, et al. Pushing the Frontiers of Unconstrained Face Detection and Recognition: IARPA Janus Benchmark-A .2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. IEEE, 2015, P. 1931-1939.

[18] WHITELAM C, TABORSKY E, BLANTON A, et al. IARPA Janus Benchmark-B Face Dataset. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA. IEEE, 2017, P. 90-98.

[19] MAZE B, ADAMS J, DUNCAN J A, et al. IARPA Janus Benchmark-C: Face Dataset and Protocol.2018 International Conference on Biometrics, Gold Coast, QLD, Australia. IEEE, 2018, P.158-165.

[20] WOLF L, HASSNER T, MAOZ I. Face Recognition in Unconstrained Videos with Matched Background Similarity.24th IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA. IEEE, 2011, P.529-534.

[21] CAO Q, SHEN L, XIE W D, et al. VGGFace2: A Dataset for Recognizing Faces Across Pose and Age. 13th IEEE International Conference on Automatic Face and Gesture Recognition, Xi'an, China. IEEE, 2018, P.67-74.

[22] Information on: https://arxiv.org/abs/1411.7923.2014.

[23] WANG F, XIANG X, CHENG J, et al. NormFace: L2 Hypersphere Embedding for Face Verification. Proceedings of the 2017 ACM on Multimedia Conference, Mountain View, CA, USA, 2017, P.1041-1049.

[24] LIU W Y, WEN Y D, YU Z D, et al. SphereFace: Deep Hypersphere Embedding for Face Recognition. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA. IEEE, 2017, P.6738-6746.

[25] WANG H, WANG Y T, ZHOU Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, P.5265-5274.

[26] BANSAL A, NANDURI A, CASTILLO C D, et al. UMDFaces: An Annotated Face Dataset for Training Deep Networks. 2017 IEEE International Joint Conference on Biometrics, Denver, CO, USA. IEEE, 2017, P.464-473.

[27] BANSAL A, CASTILLO C D, RANJAN R, et al. The Do's and Don'ts for CNN-based Face Verification.2017 IEEE International Conference on Computer Vision Workshops, Venice, Italy. IEEE, 2017, P.2545-2554.