Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230946

Studies advanced in face recognition

Haoyu Wang

Northeast Forest University, Haerbin, Heilongjiang Province, 150040, China

2020213504@nefu.edu.cn

Abstract. Face recognition has always been a hot topic in the field of computer vision research. Its purpose is to enable computers to have the ability to recognize people through face images like humans. The early face recognition technology was mainly based on artificial features, mainly including five parts: face image acquisition, face image detection, face image preprocessing, face image feature extraction and matching recognition. Thanks to the rapid development of convolutional neural networks, face recognition technology based on deep learning has gradually become the mainstream method. Focusing on the above two categories of frameworks, in this paper, we introduce representative algorithms for face recognition, including their design ideas, basic processes and key steps. We also discuss existing problems in the field of face recognition and predict future directions for this topic.

Keywords: Face Recognition, Deep Learning, DeepID.

1. Introduction

Face recognition has been a hot research issue in the fields of artificial intelligence, computer vision, and psychology in recent years. It aims to use computers to extract effective recognition information from face videos or images, and finally identify the identity of face objects. Similar to other biological characteristics (such as iris, fingerprints, etc.) of the human body that have been used for identification.The application of face recognition is becoming more and more widespread, such as for criminal case detection, intelligent transportation, entrance and exit control, Internet services, etc.

Research on face recognition can be traced back to the late 1960s, and the main idea is to design feature extractors and use machine learning algorithms for classification. Face recognition in this period is mainly by detecting and measuring some facial features unique to the face, such as ear size, eyebrow shape, eye size, eye distance, nose length, etc. Beginning in the 1970s, some scholars began to use geometric parameters to describe human faces, whose representative work was the launch of the face recognition system Kenade. After the 1980s, principal component analysis (PCA) algorithms were used to identify human faces, which reduces the dimensionality of human face data to generate eigenfaces so that can reduce computational complexity to a certain extent. At the same time, related face recognition technologies such as independent component analysis (ICA) and linear discriminant analysis (LDA) are derived. After introducing deep learning into face recognition in 2012, feature extraction was transferred to neural network, and deep learning achieved great success in face recognition. Different from the early manual features, face recognition based on deep learning has

DOI: 10.54254/2755-2721/6/20230946

made breakthroughs in both accuracy and speed, and has gradually become the mainstream framework for face recognition in recent years.

Focusing on the above two categories of methods, in this paper, we introduce the research progress of face recognition technology in detail. Specifically, in section 2, we will introduce representative face recognition methods, including their design ideas, basic processes, and key principles. In section 3, we will introduce common face recognition datasets and analyze the performance of different methods to compare the application boundaries of these methods on different datasets. Finally, we discuss the existing problems in the field of face recognition, such as occluded face recognition, and predict the future development direction of this topic in section 4.

2. Traditional face recognition algorithm

2.1. Basic framework of traditional face recognition

Early face recognition usually input took pictures or videos and output the name of one or more people appearing in the input, which mainly have five parts (1) Face image acquisition. Different face images (different positions or expressions) can be collected by the camera, such as static images and dynamic images. (2) Face image detection. Face detection is responsible for checking whether there are faces in the input picture or video, and the location area of each face. The essence of face detection is a binary classification problem. (3) Face image preprocessing. Face image preprocessing is based on the image processing of face detection results to serve the feature extraction process. (4) Face image feature extraction. The feature extraction module is responsible for extracting various feature information required by the face recognition module from the detected face area, and the feature information required by different recognition algorithms is different. (5) Face recognition. For the obtained face features, face recognition constructs a classification model to compare it with the obtained face feature templates, and make judgments based on the degree of similarity with the face identity information [1].

2.2. Representative methods based on geometric features

Representative traditional face recognition methods are mainly based on geometric features. A set of feature measures (such as distance and angle) are derived from the salient points for recognition, which are determined according to the profile curve. The most representative traditional face recognition methods are the PCA and LDA, which we will introduce in the following subsections in detail.

2.2.1. Face recognition based on PCA. The face recognition method based on principal component analysis. From a statistical point of view, the face recognition method based on principal component analysis finds the basic elements of the face image distribution to approximate the face image, The feature subsurfaces generate a subspace in the corresponding image space, called the subsurface space. Compute the projected distance of the test image window in subface space. If the window image satisfies the threshold comparison condition, it is judged as a human face.

2.2.2. Face recognition based on LDA. Considering a situation where we have two pictures of the same person, the only difference between them is that one of the pictures is with the eyes closed. At this time, the global features of the two extracted faces are basically similar (each feature vector can see the entire picture), which adds a lot of difficulty to the recognition of the two images. In theory, in order to improve feature quality and recognition results, we should pay more attention to the local area around the eyes. Local Binary Pattern (LBP) is a representative face recognition algorithm based on local information, which is used for image local texture feature extraction. As the Figure 1 shown, the original LBP operator is defined as thresholding the central pixel of the window and comparing the gray values of 8 adjacent pixels in a 3*3 window [2].

DOI: 10.54254/2755-2721/6/20230946

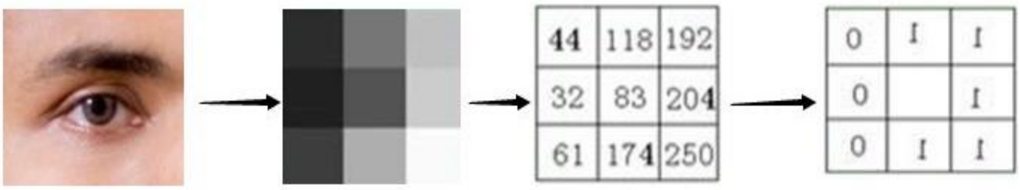
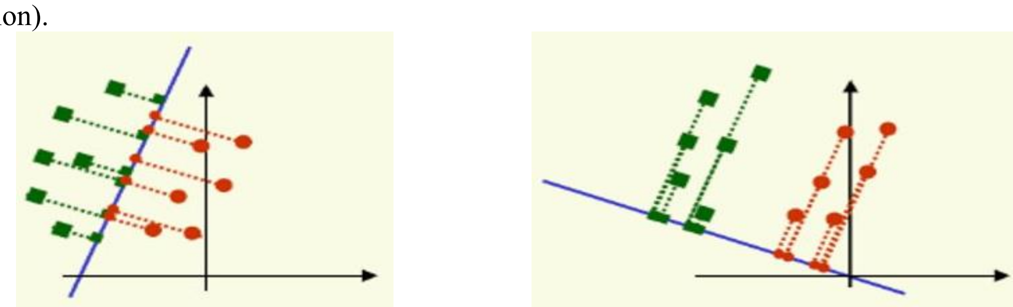


Figure 1. Visualization of the LBP calculation process.

The LBP operator quantizes points using the relationship between surrounding points.The original LBP was proposed, and researchers are increasingly proposing improvements and optimizations to it, so that it has properties such as rotation invariance.

2.2.3. Comparison of face recognition based on PCA and LDA. The left picture is the PCA, and all it does is map the data to the axis that is most convenient for representing that set of data. The classification effect may be modest. Right is the LDA, clear, after increasing the classification information, two groups of input mapping to another on the axis, with such a mapping, becomes easier to distinguish between two sets of data (in the lower dimensions can distinguish, reduce a lot of computation).



(a) Example of PCA classification (b) Example of LBP classification

Figure 2. Comparison of the LBP and PCA.

3. Face recognition algorithm based on Deep Learning

Neural network technology has begun to recover and has shown great power in image processing and recognition. The use of neural networks in face recognition has also become popular and achieved good results [3]. Neural network technology not only trains and learns labeled face data in supervised learning, but also uses unsupervised learning technology to extract face features, and uses supervised learning technology to select face features to reduce classification errors [4]. Representative face recognition methods based on deep learning mainly include DeepFace and DeepID.

The architecture of DeepFace consists of 6 convolutional layers and 2 fully connected layers.Through three layers of convolution with shared weights, the features of each layer are extracted on the entire image (within the same layer, features of the same dimension are extracted because the filters are the same). In the last three layers, Local-Conv is used to introduce various features at different positions of the face. Due to the huge number of Local-Conv parameters, it is necessary to use a large database for training. DeepFace is based on Facebook's own SFC dataset, which contains data on 4.4 million faces from 4,030 people. The feature maps generated by over-convolution are sent to the 7th layer, the fully connected layer, and can perceive all image information transferred from the previous layer. Face recognition of various architectures, most of which include this layer, because the connection layer can capture the correlation between distance and face features. In all features of layer 7 L2 normalization, the range between [0, 1] softens the light sensitivity image.

The architecture of DeepFace consists of four convolutional layers and one fully connected layer. According to the distribution of facial feature points, 10 regions are selected from the input image and

DOI: 10.54254/2755-2721/6/20230946

each region is resized into three photos with different scales. After the data enhancement, an image has generated 120 regions at this time. Each region and its horizontally flipped regions are sent into the network for feature extraction. In this step, a total of 60 neural networks are trained [5]. Take the input of a region as an example. If the area is a rectangle, it will be adjusted to 39*31; if the area is a square, it will be adjusted to 31*31. After four layers of convolution, the feature maps of the third and fourth layers are 3*2*60 and 2*1*80 dimensions, respectively.

4. Experiments and performance analysis

4.1. Common data

4.2. LFW

The full name is Labeled Faces in the Wild. This data set is bound to be used in face assessment, containing 13,000 face maps from 1680, which were searched online. They're mostly positive faces. This
data set is also the simplest, with a maximum accuracy of around 99.9%.

4.3. CFP

This data set consists of about 7,000 images for 500 identities. What makes this data set special is that it has 10 positive images and 4 side images for each person.

4.4. VGG-Face

2 million images from 2,622 people. Each person has a summary of 2000+ pictures, and there are many overlaps with MS-Celeb-1M. This data set is often used as the data of the training model, and the noise is relatively small, so relatively good training results can be obtained.

4.5. CASIA-WebFace

The dataset was collected from the IMBb website and contained 500K images of 10K people. At the same time, similarity clustering is done to remove some noise. The data set source of CAISA-WebFace is the same as that of IMDb-Face, but there are fewer images than IMDb-Face due
to data cleansing. Noise is not very much, suitable for training data [6].

Datasets	Publish Time	#photos	#subjects	# of photos per subject [1]	Key Features
MS-Celeb-1M (Challenge 1)[45]	2016	10M 3.8M(clean)	100,000 85K(clean)	100	breadth; central part of long tail; celebrity; knowledge base
MS-Celeb-1M (Challenge 2)[45]	2016	1.5M(base set) 1K(novel set)	20K(base set) 1K(novel set)	1/-/100	low-shot learning; tailed data; celebrity
MS-Celeb-1M (Challenge 3) [163]	2018	4M(MSv1c) 2.8M(Asian-Celeb)	80K(MSv1c) 100K(Asian-Celeb)	-	breadth;central part of long tail; celebrity
MegaFace [44], [164]	2016	4.7M	672,057	3/7/2469	breadth; the whole long tail;commonalty
VGGFace2 [39]	2017	3.31M	9,131	87/362.6/843	depth; head part of long tail; cross pose, age and ethnicity; celebrity
CASIA WebFace [120]	2014	494,414	10,575	2/46.8/804	celebrity
MillionCelebs [165]	2020	18.8M	636K	29.5	celebrity
IMDB-Face [124]	2018	1.7M	59K	28.8	celebrity
UMDFaces-Videos [166]	2017	22,075	3,107	–	video
VGGFace [37]	2015	2.6M	2,622	1,000	depth; celebrity; annotation with bounding boxesand coarse pose
CelebFaces+ [21]	2014	202,599	10,177	19.9	private
Google [38]	2015	>500M	>10M	50	private
Facebook [20]	2014	4.4M	4K	800/1100/1200	private

[1] The min/average/max numbers of photos or frames per subject

Figure 3. The commonly used FR datasets for training.

DOI: 10.54254/2755-2721/6/20230946

4.6. Performance comparison

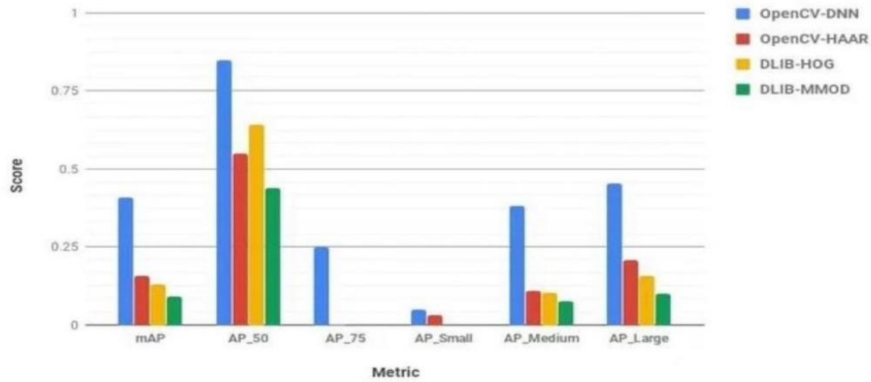
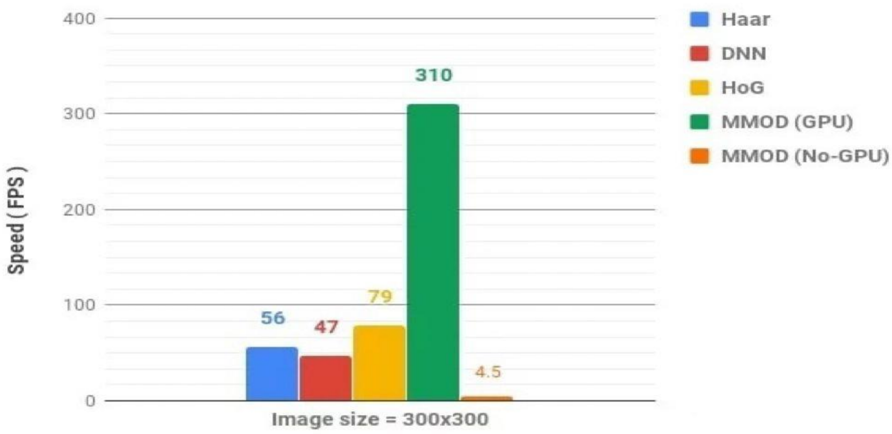


Figure 4. Wrong way of comparing dlib face detection with others.

I tried to evaluate the four models using the CFP dataset, using the scripts used to evaluate the OpenCV-DNN model. What I found, however, was surprising. Dlib results are worse than Haar, although the output from Dlib looks much better visually. Here are the accuracy scores for each of the four methods.

Let's use a 300 by 300 graph to compare the two methods. MMOD detectors can run on Gpus, but support for NVIDIA Gpus in OpenCV is still not there. Therefore, we evaluate the method only for the CPU and report the results of the MMOD on the GPU as well as the CPU. As you can see, for an image of this size, all methods except MMOD are executed in real time. The MMOD detector is very fast on the GPU, but very slow on the CPU. It should also be noted that these numbers may be different on different systems [7].



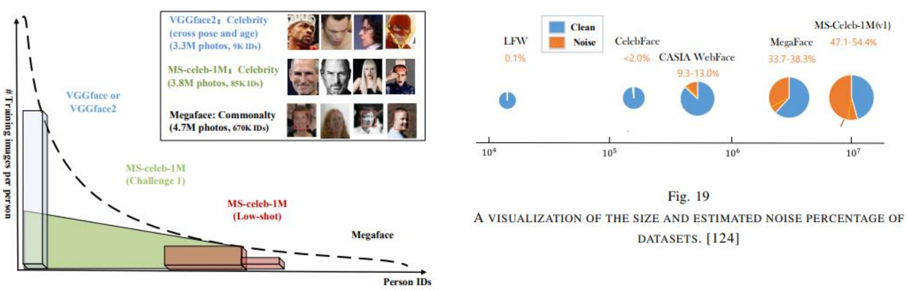


Figure 5. A visualization od the size and estimated noise percentage of datasets.

The advantage of the algorithm based on set features is that the algorithm has a fast recognition speed and requires a small memory.

The advantage of PCA algorithm is that it makes the data set easier to use. Save the cost of the algorithm; Noise reduction; Make the results easy to understand; And there are no parameters.

The advantage of LDA algorithm is that category prior knowledge experience can be used in the process of dimension reduction.The dimensionality reduction of LDA can be reduced up to the dimensionality of class k-1. Of course, there are some upgrades to LDA that can avoid this problem.

The advantage of LBP algorithm is to alleviate the problems caused by light changes to a certain extent. It has rotational invariance. The characteristic dimension is low and the calculation speed is fast. The disadvantage is that if the lighting is uneven, then the LBP value does not reflect the true texture characteristics. Be sensitive to direction [8].

DeepID was trained using CelebFaces+ with 10,177 people and 202,599 images; 8,700 trained the DeepID and 1,477 trained the Joint Bayesian classifier. The number of shred patches was 100, and 5 different scales were used. The resulting vector length of each image was 32000, and the dimensionality was reduced to 150 using PCA. When training ConvNet, the cross-entropy loss function of class 10000 was used as the target function. Finally, the validation accuracy of the paper reached 97.45% on LFW. The disadvantage is that the algorithm is not easy to explain, because the number of neurons is large, the operation time is long, and the need for multiple face images for training, in the training process often need to adjust some parameters, so the suitable range is limited to small face library.

In today's society, face recognition applications are everywhere. Due to the impact of the epidemic, many mobile phones and cameras are equipped with face recognition devices that can recognize people wearing masks. This technology has increased the requirements for feature extraction [9]. Face recognition is mainly used for identity recognition.Because of the rapid popularity of video surveillance, many video surveillance applications urgently need a rapid identification technology in the remote and non-cooperative state of users to quickly confirm the identity of distant personnel and realize the function of early warning. This enables fast identification. Face recognition products have been widely used in various fields. With the further improvement of technology and the enhancement of social communication, face recognition technology will be applied in more fields [10].

5. Conclusion

In this paper, we introduce the research progress of face recognition from two aspects of traditional face recognition and face recognition based on deep learning. Traditional face technology is mainly used for the recognition and processing of still pictures, and now it has basically reached the applicable level in a controlled environment, and there are still some technical obstacles waiting to be overcome for face recognition in an uncontrolled environment. In recent years, with the emergence and rapid development of the Internet of Things and big data technology, how to retrieve specific faces from massive video data and track specific faces poses new challenges to face recognition technology. The spatio-temporal information of human face detection, recognition and tracking is the main direction of current research workers[11].

References

[1] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of cognitive neuroscience, 1991, 3(1): 71-86.

[2] Turk M A, Pentland A P. Face recognition using eigenfaces[C]//Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on. IEEE, 1991: 586-591.

[3] Belhumeur P N, Hespanha J P, Kriegman D. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997, 19(7): 711-720.

[4] M. D. Kelly, “Visual identification of people by computer.,” tech. rep., STANFORD UNIV

DOI: 10.54254/2755-2721/6/20230946

CALIF DEPT OF COMPUTER SCIENCE, 1970.
[5] T. KANADE, “Picture processing by computer complex and recognition of human faces,” PhD Thesis, Kyoto University, 1973.
[6] K. Delac and M. Grgic, “A survey of biometric recognition methods,” in 46th International Symposium Electronics in Marine, vol. 46, pp. 16– 18, 2004.
[7] U. Park, Y. Tong, and A. K. Jain, “Age-invariant face recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, pp. 947–954, 2010.
[8] Z. Li, U. Park, and A. K. Jain, “A discriminative model for age invariant face recognition,” IEEE transactions on information forensics and security, vol. 6, no. 3, pp. 1028–1037, 2011.
[9] Brunelli R, PoggioT. Face Recognition: Features Versus Templates [J]. IEEE T rans on Pattern Analysis and M achine Inteligence , 1993, 15 (10):10421052
[10] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of cognitive neuroscience, vol. 3, no. 1, pp. 71–86, 1991.
[11] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711–720, 1997.