Occluded Face Recognition Based on the Deep Learning

Gui Wu[1], Jun Tao[2 3], Xun Xu[4]

1. Educational Administration office, Jianghan University, Wuhan 430056
E-mail: wugui214@163.com

2. School of Mathematics and Computer Science, Jianghan University, Wuhan 430056
E-mail: martintao2006@163.com

3. Department of Electrical & Computer Engineering, Rowan University, Glassboro, New Jersey, USA 08028
E-mail: taoj0@students.rowan.edu

4. Graduate School, Jianghan University, Wuhan 430056
E-mail: 734372817@QQ.com

Abstract: With the aggravation of social public security problems, the face recognition without occluded parts can no longer meet the needs of modern society. All kinds of face recognitions in complex environment need to be realized in the real situations. The paper proposed a new method to recognize the occluded face based on the deep learning. The face recognition model is trained and learned based on convolution neural network of the deep learning, which has strong robustness to illumination difference, facial expression change and facial occlusion. Through a large number of experimental tests and result analysis, the occluded face recognition rate can reach up to 98.6%. Therefore, this method proposed in the paper realizes face recognition with occlusion in complex environment and meets the needs of practical applications.

Key Words: Deep Learning, Face Recognition, Convolutional Neural Network, Triplet Loss Function

1 INTRODUCTION

As one of the successful applications in pattern recognition and image processing, face recognition has been a research hotspot in the past twenty years. In contrast, the universality, collectability and acceptability of face recognition are relatively high, which has a series of advantages such as convenience, friendliness, easy acceptance and not easy to forge. With the development of technology, people pay more attention to public safety. People also have new requirements for face recognition technology. In actual production and life, the collected face images are not necessarily complete and clear. Robustness to illumination difference, facial expression change and facial occlusion is one of the criteria for judging a system.

At present, the research on face recognition with occlusion interference is not in-depth in our country. Most of the research is on face recognition without occlusion at close range. There are many other face recognition methods, such as linear regression classification (LRC) method. The experimental results show that LRC can recognize pure faces well, but the effect is very poor when the interference factors such as occlusion are added. Another method is based on Sparse Representation Classification (SRC). Although the sparse face has strong robustness to noise and good recognition rate for interference factors such as occlusion. However, based on SRC, there is a strong assumption that all face images must be aligned strictly in advance. Otherwise, sparsity is difficult to satisfy. In other words, for facial expression changes, the face with attitude changes does not satisfy the hypothesis of sparsity. Therefore, the classical sparse face method is difficult to be used in real application scenarios.

In this paper, the images are mapped to Euclidean space by convolutional neural network. The spatial distance is directly related to image similarity. Different images of the same person have small spatial distance, and different images of different people have large spatial distance. This method is direct and efficient, and has robustness to occlusion, expression change, attitude change and other interference factors. It can realize face recognition under certain complex conditions such as occlusion and other interference factors, and meet the requirements of practical applications.

2 THE CONVOLUTIONAL NEURAL NETWORK AND THE LOSS FUNCTION

2.1 The Convolutional Neural Network

The Convolutional Neural Network (CNNs) is a distortion of multi-layer perceptron inspired by biological vision and aimed at simplifying preprocessing operations. It is essentially a forward feedback neural network. The biggest difference between CNNs and multi-layer perceptrons is that the first layers of the network are composed of convolution layer and pooling layer alternately. It simulates simple cells and cells used for high-level feature extraction in visual cortex and complex cell alternating cascade structure.

As shown in Figure 1, the convolution layer is usually connected with the input layer in the convolution neural network, and the input image is obtained from the input

This work is supported by educational research fund of Hubei Provincial Department of Education (H2014269).

978-1-7281-0106-4/19/$31.00 ©2019 IEEE

layer. The middle layer of the network is alternately connected by the convolution layer and the down-sampling layer. After several convolutions and down sampling, the full connection structure is used to get the output of the whole network. The output layer uses the label information of the target and calculates the convergence of the network by supervised training. The convolutional neural network extracts different features by sharing the weights of neurons on the neuron plane, which effectively reduces the number of parameters in the training process of convolutional neural network and reduces the computational complexity.

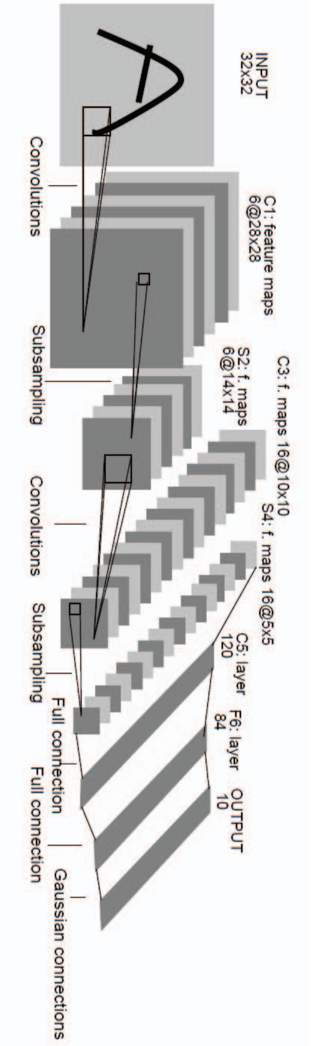


Fig 1. The schematic diagram of the Convolutional Neural Network

Face images are disturbed by occlusion factors, which results in the loss or alienation of the recognized image pixels, thus resulting in the recognition errors of traditional models. However, the deep learning model based on convolutional neural network has the advantages of weight sharing, low complexity and small number of weights, which makes the model have strong anti-jamming ability and can still be recognized under the interference of occlusion factors.

2.2 The Triplet Loss Function

The Loss function is used to estimate the degree of inconsistency between the predicted value and the real value of your model. The loss function used in this model is Triplet Loss.

The Triplet Loss is a loss calculated from a triple of three pictures. The triple is composed of Anchor (A), Negative (N), and Positive (P). Any picture can be used as a base point (A), and then the picture that belongs to the same person is its (P), and the picture that does not belong to the same person is its (N).

The learning objectives of Triplet Loss can be visualized as follows:

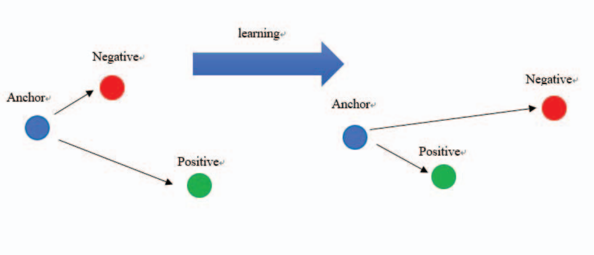


Fig 2. The schematic diagram of Triplet Loss

As shown in Figure 2, the optimizing function is formula (1).

$$\|f(x_i^a)-f(x_i^p)\|_2^2+\alpha<\|f(x_i^a)-f(x_i^n)\|_2^2, \forall(x_i^a,x_i^p,x_i^n)\in T$$

(1)

The formula is that the distance in the left class (plus the margin) is less than the distance between the right classes. This constraint needs to be valid on all Triplet image pairs. Conversely, its loss function becomes as follows: minimization (intra-class distance-class distance + margin). Among them, the two norms on the left represent the distance within the class, the two norms on the right represent the distance between the classes, and alpha is a constant. The optimization process is to use gradient descent method to make the loss function decrease continuously, that is, the distance between classes decreases continuously and the distance between classes increases continuously.

But there are problems in this way: if the best triple is chosen, the local extremism will be created, and the network may not converge to the optimal value. Therefore, to select all Positive in mini-batch can make the training process more stable. For the selection of Negative, the semi-hard Negative is adopted, so that the distance from a to n is greater than the distance from a to P.

In industry, Cross Entropy (Soft Max) is often used as loss function. Traditional classification is the recognition of large categories of birds, birds and dogs, but there are some requirements to be precise to the individual level, such as precise to which person's face recognition applications, Triplet loss function is more effective than Soft Max function. And when the sample size is very small, using Soft Max training will be easier to converge. However, when the training set contains a large number of different individuals (more than 100,000), the output of Soft Max on the last layer becomes very large, but the training using Triplet loss can still work properly.

3 THE MODEL STRUCTURE

The basic model structure of this model is the convolutional neural network model. As shown in Figure 3, it is the model structure of the program. Each module has the following functions:

1. Batchs: it refers to the input face image sample, where the sample has been found through face detection and cut to a fixed size of the image sample.

2. DEEP ARCHITECTURE: In-depth learning framework, the experiment adopts Inception-ResNet-v1 network structure.

3. L2: The feature is normalized to $||f(x)||2 = 1$, so that all image features are mapped to a hypersphere.

4. EMBEEDING: Embedded layer.

5. Triplet Loss: Triple loss function.

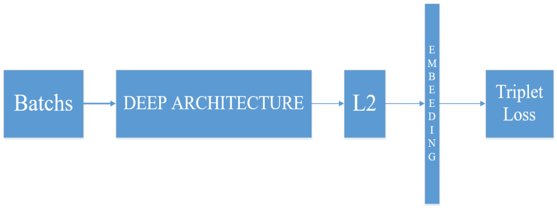


Fig 3. The system constructure

The purpose of the model is to embedding face images into D-dimensional Euclidean space. However, unlike general deep learning frameworks, traditional deep learning frameworks are usually Double Loss or Single Loss. In this paper, we use the loss function Triplet Loss of three images to directly learn the separability between features: the feature distance between identical identities should be as small as possible, while the feature distance between different identities should be as large as possible.

4 EXPERIMENTAL RESULTS AND DATA TEST

The experiment is based on the face image shown in figure 4. The image pixels are 160*160.

Fig 4. The benchmark of experimental human face

The experiment selects a group of different faces of the same person and adds occlusion interference. Then, the Euclidean distances between them and the reference photographs are calculated respectively. In theory, the Euclidean distances should be less than 1.

Based on figure 4, the following experimental results are obtained by comparing the face images of the same person:

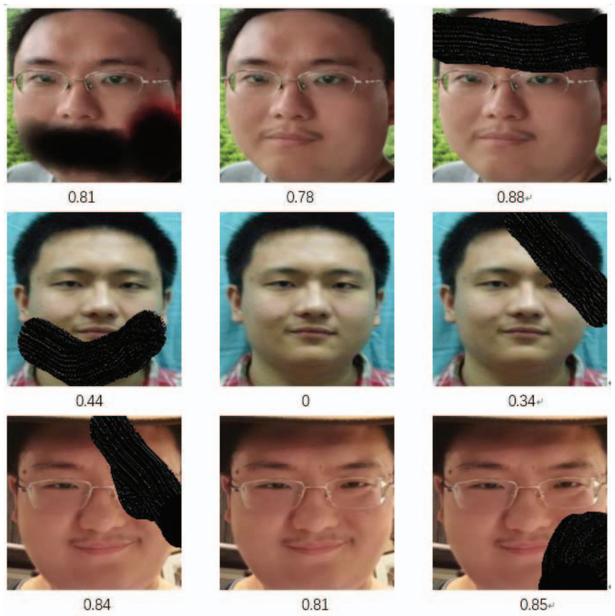


Fig 5. The face test results of the same person

As shown in figure 5, the photos of the same person in different periods were selected, and the occlusion interference factors were added to interfere with key feature information such as eyes and mouth. The Euclidean distance calculated by the model is less than 1.

It can be proved that the model is identical to the one with good recognition, although the time and age are added to the photos and the key features of the face are occluded.

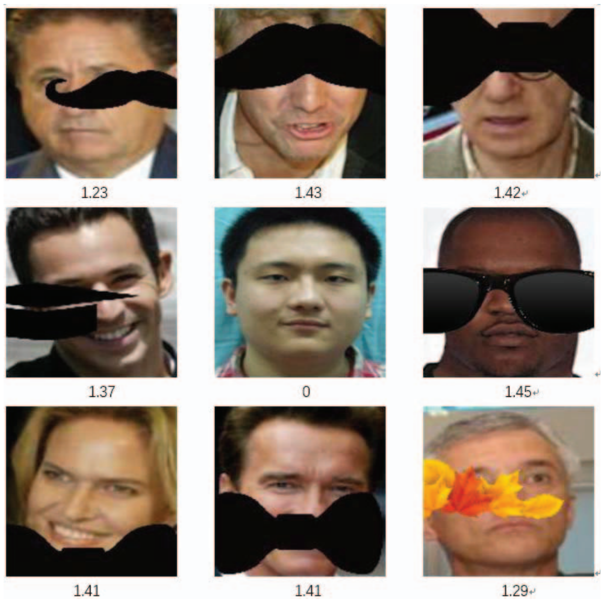


Fig 6. The face test results of difficult person

Based on the facial photos of figure 4, different people's pictures are selected, and occlusion interference factors are added, and the Euclidean distance between them is calculated respectively.

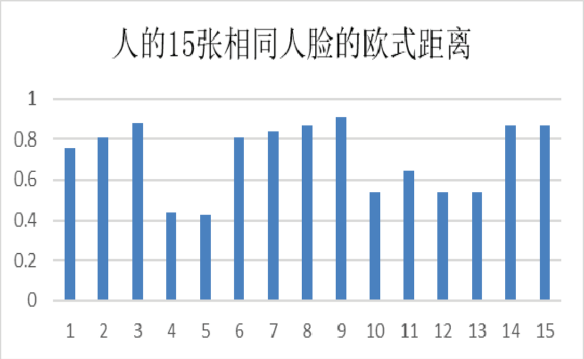
In theory, the Euclidean distance should be greater than 1.

As shown in figure 6, the experiment uses the middle face as the benchmark and adds contrast pictures. The added contrast pictures contain interference factors such as ethnic differences, age differences and occlusions, and occlude key feature factors such as eyes and mouth. It can be seen from the calculated Euclidean distance that the Euclidean distance is greater than 1.

From the above experiments, it can be concluded that the model can distinguish the faces of different people well under the interference of the key feature information such as eye and mouth, and the differences of race, sex and age.

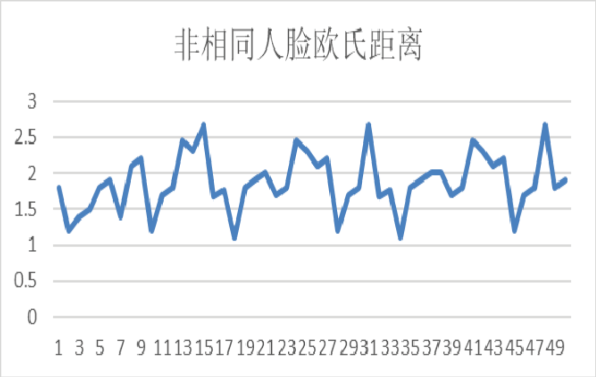
According to the above interference factors, the number of test pictures was added and the charts were listed as following.

Table 1 Euclidean Distances Between The Same Person



As shown in Table 1, 15 face images of the same person in different periods based on figure 4 are added with occlusion interference factor of about 30%. The Euclidean distance is calculated by comparing with the base face images. It can be seen that the distance values are less than 1.

Table 2 Euclidean Distance Of Different People's Faces



As shown in Table 2, 50 face photos of different races, sexes and ages can be selected from LFW data on the basis of figure 4. The key features of face information can be occluded by adding interference factors, and the occlusion rate is about 30%. Calculate the Euclidean distance. The values are all greater than 1.

Therefore, it can be concluded that under the interference factor of 30% occlusion rate, the model can still have a good recognition rate for different races, different ages and different genders. In summary, the same person pictures selected in the experiment have time and age differences, and different people have gender, race, age differences, and add interference factors for the occlusion of key feature information of face. However, the model can still recognize face well.

Table 3 The Recognition Rate Of Different Algorithms When The Occlusion Rate Is 30%

Algorithm Name	Recognition Rate	Image Number of Recognition Failure
SRC	86.8%	556
LRC	65.2%	1475
Algorithm in the paper	98.6%	59

As shown in Table 3, the experiment intercepts the LFW database and produces 4234 images from the face image data collected by the camera. The recognition rate of SRC and LRC algorithm and the algorithm in this paper is tested under the condition of adding 30% occlusion interference factor. The recognition rate of the algorithm can reach 98.6%, which has certain practical application value, but the recognition rate of SRC and LRC algorithm is not satisfactory.

It can be concluded that, compared with other algorithms, in the case of occlusion factors interfering with the model, the model does not lose stability, and can still distinguish the faces of different people very well, and the recognition rate is higher. In conclusion, compared with traditional algorithms, the algorithm proposed in the paper is more practical in face recognition with occlusion.

5 CONCLUSION

The algorithm model in this paper is based on convolutional neural network and the loss function is Triplet Loss. The separability between direct learning features is that the feature distance between identical identities should be as small as possible while the feature distance between different identities should be as large as possible.

The algorithm model is trained and tested by a large number of pictures, and the recognition rate of masked faces is 98.6%. It is verified that the model can still recognize face accurately under the interference factors such as occlusion. Compared with other traditional face recognition models, the model is robust to occlusion and other interference factors, and has the advantage of high face recognition rate. It has practical application value and achieves face recognition under occlusion interference factors.

References

[1] Yen Shi-Jim, Yang Jung-kuei, Kao Kuo-Yuan, et al. Bitboard knowledge base system and elegant search architectures of connect6, Knowledge-based systems, Vol.34 Special Issue, 43-54, 2012.

[2] Wu I-Chen, Lin Hung-Hsuan, Lin Ping-Hung, et al. Job-Level Proof-Number Search for Connect6, Computers and games, Vol.6515, 11-22, 2011.

[3] Jun Tao, 3D modeling of small object based on the projector-camera system, Kybernetes, Vol.41, No.9, 1269-1276, 2012.

[4] Yen Shi-Jim, Yang Jung-Kuei, Two-Stage Monte Carlo Tree Search for Connect6, IEEE Transactions On Computational Intelligence And Ai In Games, Vol.3, No.2, 100-118, 2011.

[5] Wu I-Chen, Lin Ping-Hung, Relevance-Zone-Oriented Proof Search for Connect6, IEEE Transactions On Computational Intelligence And Ai In Games, Vol.2, No.3,191-207, 2010.

[6] Jun Tao, Development and application of functionally gradient materials, International conference on industrial control and electronics engineering, 1022-1025, 2012.

[7] Qiao Zhihua, Yang Ming, Wang Zijuan, Technologies Analysis of Connect6 Computer Game, achievements in engineering materials, energy, management and control based on information technology, 679-682, 2011.

[8] Xu Chang-ming, Ma Z. M., Tao Jun-jie, Xu Xin-he, Enhancements of Proof Number Search in Connect6, 21st Chinese control and decision conference, vols 1-6, proceedings, 4525-4529, 2009.

[9] Jun Tao, Design and visualization of optical feedback laser based on computer vision, International conference on industrial control and electronics engineering, 1030-1032, 2012.

[10] Lin Yi-Shan, Wu I-Chen, Yen Shi-Jim, TAAI 2011 Computer-Game Tournaments, ICGA JOURNAL, Vol.34, No.4, 248-250, 2011.

[11] Yoshizoe Kazuki, Kishimoto Akihiro, Mueller Martin, Lambda Depth-first Proof Number Search and its Application to Go, 20th International Joint Conference on Artificial Intelligence, 2404-2409, 2007.

[12] Jun Tao, Face reconstruction based on camera-projector system, International conference on industrial control and electronics engineering, 1026-1029, 2012.

[13] Tao Jun-jie, Xu Chang-ming, Han, Kang, Construction of Opening Book in Connect6 with Its Application, 21st Chinese control and decision conference, vols 1-6, proceedings, 4530-4534, 2009.

[14] Wu I. –Chen, Lin Hung-Hsuan, Sun Der-Johng, Job-Level Proof Number Search, IEEE transactions on computational intelligence and ai in games, Vol.5, No.1, 44-56, 2013.

[15] Saito Jahn-Takeshi, Winands Mark H. M., van den Herik H. Jaap, Randomized Parallel Proof-Number Search, advances in computer games, Vol.6048, 75-87, 2010.