Journal of Computers Vol. 31 No. 1, 2020, pp. 225-231
doi:10.3966/199115992020023101020

Multi-pose Face Recognition Based on Convolutional Neural Network

Jinyu Li, De Zhang*

School of Electrical And Information Engineering, Beijing University of Architecture & Civil Engineering, Beijing 100044, China
646738946@qq.com, zhangde@bucea.edu.cn

Received 13 January 2019; Revised 11 March 2019; Accepted 12 March 2019

Abstract. The difficulty of multi-pose face recognition is the changes of posture. Compared with the traditional face recognition, multi-pose face recognition increases the difficulty of feature extraction due to the change of posture. The feature of face recognition requires distinguishing and recognizability. In this paper, the convolutional neural network is used to first train the face pose, and then perform a single pose face recognition on the classified image. Compared with the traditional face recognition method, it can be seen from the experimental results that the classification of face poses can increase the accuracy of multi-pose face recognition.

Keywords: convolutional neural network, deep learning, face recognition, multi-pose

1 Introduction

In recent years, face recognition technology has developed fast and been applied in all aspects of social life. However, most face recognition systems are currently used for frontal face recognition [1-2]. Multi-pose face recognition is still one of the difficulties in current face recognition technology. Traditional face recognition technology is mostly applied to frontal faces [3-4]. By preprocessing the image, the target object becomes an image or data that is relatively easy to recognize. Then, the processed data is subjected to feature extraction by an algorithm. Traditional face recognition technology requires a high degree of posture on the face.

The traditional face recognition methods are as follows:

(1) Face recognition based on geometric features. Edge detection and projection functions are commonly used to extract features of major organs of the face, and various distance formulas such as the AAM algorithm are matched [5].

(2) Face recognition based on principal component analysis. The PCA algorithm converts the image two-dimensional matrix into a one-dimensional matrix and then compares the target face. However, the PCA algorithm is susceptible to external factors such as lighting and attitude [6].

(3) Face recognition based on artificial neural network. Among them, the back-propagation neural network is widely used. The algorithm has the characteristics of strong neural network learning ability and strong anti-interference ability, and has good robustness. The merits of BP algorithm recognition results depend on the rationality of manual selection features [7].

Through the above analysis, it can be seen that the traditional methods are more or less insufficient. In this paper, the original data set is processed by convolutional neural network, which enhances the self-learning of the system, and effectively avoids the problem of image distortion caused by data preprocessing.

Although there are many factors affecting the accuracy of face recognition technology, such as lighting, age, expression, etc. [8]. Posture change is the most serious factor affecting the performance of face image acquisition and face recognition system in unconstrained environment. If there are a large number of side faces in the training set, the normalization of the face will become difficult. In this paper, the pose

* Corresponding Author

feature extraction is performed on the face by convolutional neural network, and the weight model of the face pose is obtained by training through a large number of known poses. Convolutional neural network is an efficient deep learning recognition algorithm developed in recent years and has attracted extensive attention. It has become a research hotspot in the field of speech analysis and image processing. Compared with traditional neural networks, convolutional neural networks have the advantages of weight sharing and local perception. Perceived local network structure so that it is closer to biological neural network, sharing weights significantly reduce the number of model learning parameters, while reducing the complexity of the neural network structure.

The current multi-pose face recognition method is roughly divided into three recognition technologies: two-dimensional single-view face recognition, two-dimensional multi-view face recognition [9-10] and three-dimensional face recognition. Due to problems such as the degree of cooperation of the collected data of the face data, the limitation of the collected face data is caused. Two-dimensional multi-pose face recognition is still the mainstream direction of face recognition technology. Since the multi-pose face recognition training sample based on two-dimensional single view is small, it is difficult to obtain enough training samples, and the face feature information cannot be fully acquired for identification [11]. So the researchers proposed the idea of transforming single-view face recognition into multi-view face recognition through sample amplification. This paper is mainly applied to two-dimensional multi-pose face recognition.

The rest of the paper is organized as follow: The section 2 deals with the convolutional neural network. Section 3 introduces the design of multi-pose face recognition experiment. Section 4 presents the analysis of results. Finally, draw conclusions and prospects in Section 5.

2 Convolutional Neural Network

Since the application of convolutional neural networks in the field of deep learning, rapid progress has been made. Convolutional neural networks are an efficient method of identification that has developed in recent years and has attracted widespread attention. In the 1960s, when Hubel and Wiesel studied the local sensitive and directional selection of neurons in the cat's cortex, they found that their unique network structure can effectively reduce the complexity of the feedback neural network, and then propose a convolutional neural network. Convolutional neural network is a deformation of multilayer perceptron inspired by biological vision for the most simplified preprocessing operations. Its essence is a forward feedback neural network. The biggest difference between convolutional neural networks and multi-layer perceptrons is that the first few layers of the network are composed of concatenated layers and pooled layers. It simulates the alternating cascade of simple cells and complex cells used for high-level feature extraction in the visual cortex. Each neuron does not need to perceive the global image. It only needs to perceive the local part, and then combines the local information at a higher level to obtain global information. It is shown in Fig. 1.

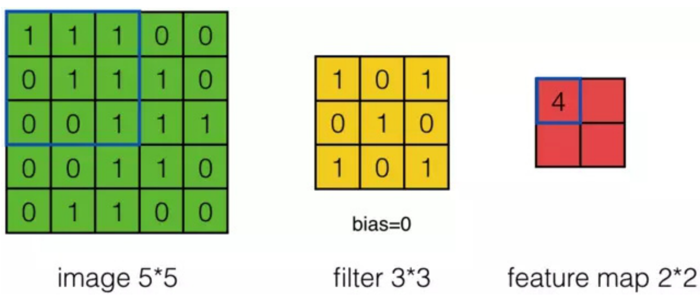

Fig. 1. Local perception field

Weight sharing is also one of the characteristics of convolutional neural networks. Regardless of the position of the image features, the statistical properties of each part of the image share weight with other parts. This also means that the features we learn in this part can also be used in another part, so we can use the same learning characteristics for all positions on this image. Since CNN's feature detection layer learns through training data, when CNN is used, explicit feature extraction is avoided, and learning is implicitly learned from training data. Furthermore, due to neuron weights on the same feature mapping surface. So the network can learn in parallel, which is also a major advantage of convolutional networks relative to the network of neurons connected to each other.

The difference between convolutional neural networks and ordinary neural networks is that the convolutional neural network contains a feature extractor composed of a convolutional layer and a sub-sampling layer. In the convolutional layer of a convolutional neural network, a neuron is only connected to a part of the adjacent layer neurons. In practical applications, multi-layer convolution is often used, and then the fully connected layer is used for training. The purpose of multi-layer convolution is that the features learned by one layer of convolution are often local, and the higher the number of layers, the learned features. The more global it is. In theory, a neural network with a shallow structure or a small number of layers may not fully express any multivariate nonlinear function when the number of nodes is large enough, but this shallow expression often needs to be too specific for implementation. Many nodes are not practical.

In this paper, a four-layer convolutional layer and a pooled layer and a three-layer fully connected layer of convolutional neural network structure are used to classify multi-pose faces. The structure, parameter results and operational performance of the convolutional neural network of this structure are described below.

3 Technical Process

The application of convolutional neural network to train face data is different from traditional face recognition, and it is not necessary to detect face regions. The face database in complex background and environment makes the face model after training more robust. In the design of the experimental process, considering the characteristics of the extracted features of the convolutional neural network, the experimental data of multiple illumination conditions were used to verify the double convolutional neural network method.

This paper uses dual convolutional neural network to train multi-pose face recognition system. The purpose of the first convolutional neural network is to classify the face poses and obtain a weight model for the facial pose features. The result of the first convolutional neural network is then used as the input layer.

(1) Input layer

In order to ensure that the convolutional neural network can obtain more image information, the face image of the experiment is not subjected to grayscale processing. Because the data set for training is too many pixels, it is not conducive to training. We process the original dataset data to 100*100 size.

(2) Convolution layer

The main purpose of the convolutional layer is to extract features from the input image, which can extract some abstract features from the input data. These abstract features are output in the form of feature maps, which are then entered into the pooled layer of the network for further processing. CNN of multi-face recognition is shown in Fig. 2.

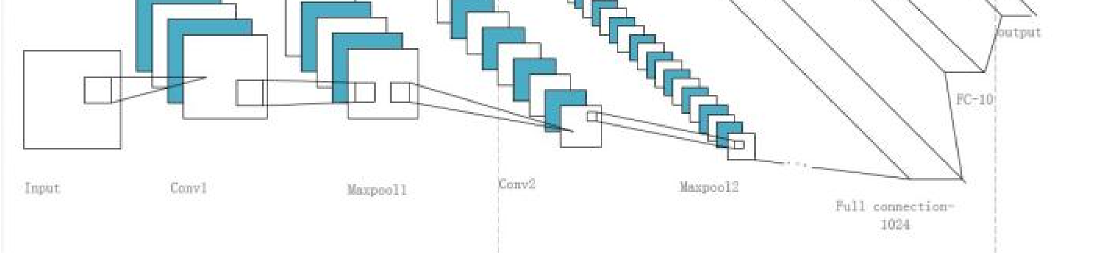

Fig. 2. CNN structure for multi-pose face recognition

The value of the convolution kernel is first set by the network initialization. After subsequent training, the value tends to the local optimal solution, and the whole convolutional neural network gradually converges. The convolution processing shows in Fig. 3.

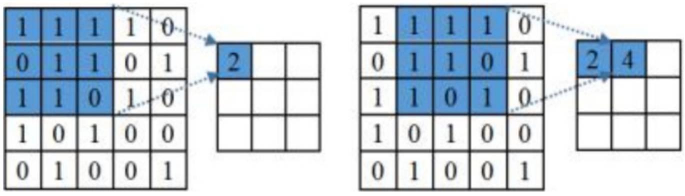

Fig. 3. Convolution process

The activation function of this article uses the Relu function. Generally, the convolutional neural network used in the field of image recognition has the characteristics of deep network layers, many neurons, and complex network structure. Therefore, the Relu function is often better than other activation functions.

$$x_j^l = \max(0, \sum\nolimits_{i \in M} x_i^{l-1} w_n^l + b_n) \quad \textbf{(1)}$$

In above formula, x is the convolutional layer element, l is the current network layer number, w is the convolution kernel element, b is a random independent item, and n is the location information of the convolution kernel element. M is the selection of the input image, i and j are a group of image selections corresponding to each other.

(3) Pooling layer

The pooling layer down samples the input. The commonly used pooling method is to find the maximum, average, and median of the output of each filter. One of the most important features of pooling is that it outputs a fixed-size matrix. After the pooling layer is processed, it always gets the output of the same dimension and passes it to the classifier of the next layer. Pooling also reduces the dimensions of the output, but retains significant features.

In the convolutional neural network established of this paper, the pooling layer reduces the feature image dimension by calculating the mean of the block of the region. The feature image data outputted by the upper convolution layer is segmented into 2*2 size blocks, and each 2*2 size block is replaced by a mean value.

The pooling processing shows in Fig. 4.

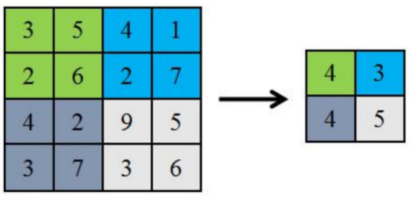

Fig. 4. Pooling processing

(4) Fully connected layer

Each node of the fully connected layer is connected to all nodes of the previous layer to combine the features extracted from the front. Due to its fully connected nature, the parameters of the fully connected layer are also the most common. In the network structure described in this paper, the first layer of the fully connected layer sets 1024 neurons, the activation function is the Relu function. The second layer sets 512 neurons, and the activation function is the Relu function. The third layer sets 13 neurons, no activation function. This paper uses two consecutive convolutional neural networks. The first convolutional neural network aims to classify the pose. The result is input to the second convolutional neural network training, and the output is identification.

4 Experiment

4.1 Data Preprocessing

The database used in this experiment is the Multi-pie face database. In the Multi-pie database, each training individual is divided into 15 angles and 20 light intensities. The data set consists of more than

300 men and women of different ages. The database contains people from multiple countries and skin colors. In order to make the multi-pose image facilitate the convolutional neural network training, this paper firstly preprocesses the database and divides the whole data set into 13 categories according to the posture. Organize the data of the same posture in the database into the corresponding folder to facilitate the training of the convolutional neural network. It is shown in Fig. 5.

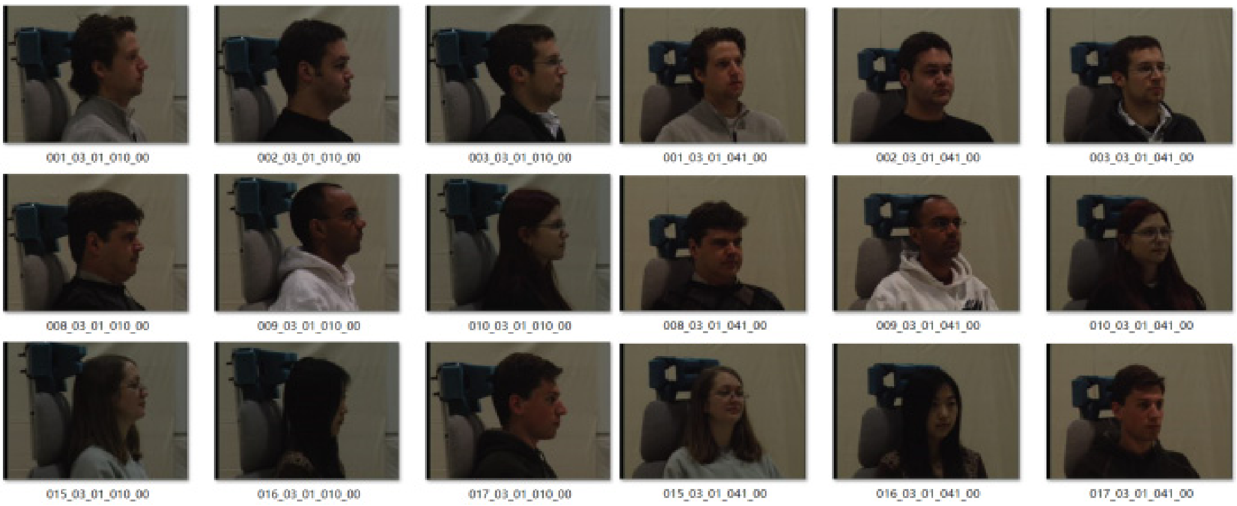

Fig. 5. Multi-pose database classification preprocessing

4.2 Robustness of Illumination

Light effects have always been an important problem in multi-pose face recognition feature extraction. Traditional methods deal with face recognition in multi-light conditions, mostly preprocessing the training data set. The basic idea of the traditional method is to extract features that do not change with the illumination or are at least insensitive to changes in illumination. Or pre-processing to eliminate the effects of illumination changes before the face image is recognized. The convolutional neural network used in this paper is robust to illumination conditions, and the robustness is verified under the convolutional neural network of pose classification. Fig. 6 shows the multi-light comparison data in the multi-pie database.

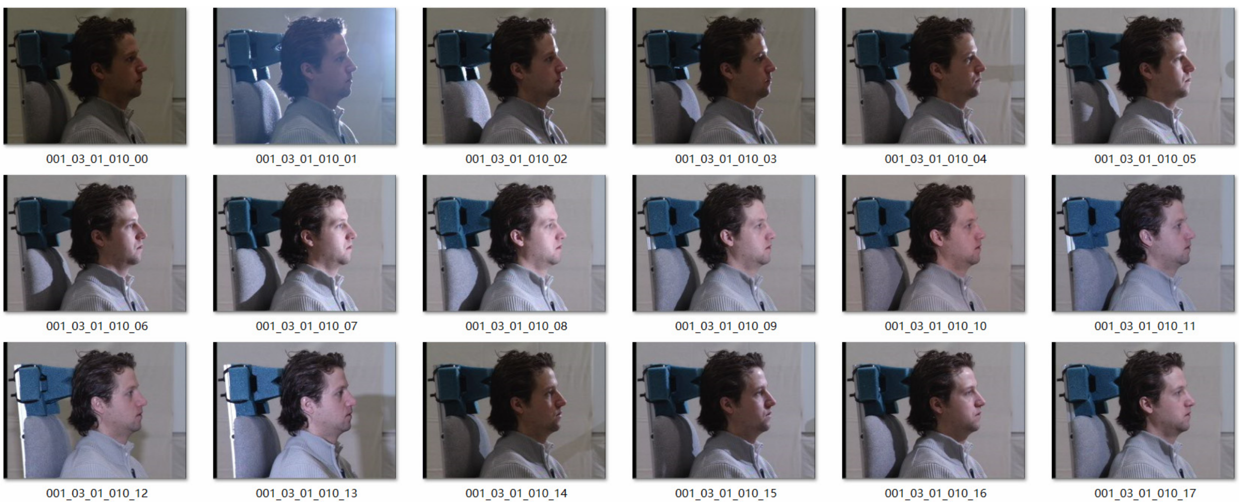

Fig. 6. Samples of multi-lighting from multi-pie dataset

In this paper, the training data of the first convolutional neural network is trained using different illumination training samples. To confirm the robustness of the convolutional neural network to the effects of illumination, each time a single illumination is used for training. Table 1 shows the results of training comparisons for different lighting conditions.

Table 1. Results from different light conditions

Light condition	Pose classification rate
1	98.9%
2	98.4%
3	99.6%
4	97.9%
5	99.1%

4.3 Experiment results

In this paper, the multi-pie database is used to verify the accuracy of the double convolutional neural network. The training set is divided into 13 categories to train multi-pose convolutional neural networks. The results are as follows. Table 2 shows the relationship between the detection accuracy rate and the number of network iterations for the face pose recognition convolutional neural network.

Table 2. Comparison of accuracy and loss under different iterations

Epoch	Training loss	Training accuracy
1	610.45	78.4%
5	6.53	94.8%
10	0.24	98.9%

It can be seen from the results that the first step classification in multi-pose face recognition has reached a high accuracy. This is the basis for the identification of the second continuous convolutional neural network.

The first convolutional neural network result is input into the double convolutional neural network. Fig. 7 shows the loss rate of double convolutional neural network training.

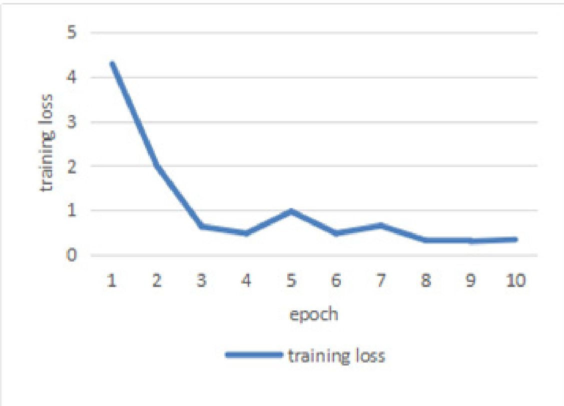

Fig. 7. training loss of double CNN

The first convolutional neural network result is input into the double convolutional neural network. That is to achieve the effect of multi-pose face recognition through the training classification of two convolutional neural networks. Table 3 compares the classification of double convolutional neural networks with the classification of ordinary convolutional neural networks.

Table 3. Comparison between double CNN and ordinary CNN

Method	Correct recognition rate
Ordinary CNN	78.40%
Double CNN	98.5%

It can be seen from the comparison results that in the case of multi-pose face recognition, the double convolutional neural network classification improves the recognition accuracy compared with the ordinary convolutional neural network.

5 Conclusion

Based on the existing 2D multi-pose face recognition technology, this paper proposes a new face recognition algorithm. When dealing with two-dimensional images, the complicated pre-processing and the cumbersome artificial feature extraction process are avoided, which shows good feature extraction performance and can solve the over-fitting problem well. Experimental results on the multi-pie database indicate that it can be seen from the results that the robustness of multi-pose face recognition can be greatly improved by double convolutional neural network classifications. The disadvantage is that the double convolutional neural networks are not robust to 3D faces.

Multi-pose recognition is always a challenging problem in the field of facial biometrics. In this paper, the convolutional neural network is used to solve the problem of pose classification and demonstrates its effectiveness on a popular large database. However, there is still time cost problem to be further reduced in the future work. The most efficient network should be constructed and applied in this CNN-based pose classification system. It is believed that an improved CNN would work well according to our current experiments.

Acknowledgements

This paper is supported by the fundamental research funds for Beijing University of Civil Engineering and Architecture (No. X18068).

References

[1] G.-F. Zou, G.-X. Fu, H. Li, M.-L. Gao, K.-J. Wang, A survey of multi-pose face recognition, Pattern Recognition and Artificial Intelligence 28(7)(2015) 613-625.

[2] Y. Wu, W. Qiu, Face recognition based on improved deep convolutional neural network, Computer Engineering and Design 38(8)(2017) 2246-2250.

[3] F. Wang, Y. Zhang, D. Zhang, H. Shao, C. Cheng, The application of shortcut neural network based on shortcut in face recognition, Journal of Electronic Measurement and Instrument 32(4)(2018) 80-86.

[4] Z. Chen, D. Zhou, J. Huang, Emotional invariant 3D face recognition based on convolutional neural network, Electronic Measurement Technology 40(4)(2017) 161-171.

[5] J. Zheng, Y. You, Identification of standard frontal face images, Computer Engineering 1(1992) 1-6.

[6] C. Jin, X. Qiu, H. Zhang, Design and implementation of face recognition algorithm based on PCA algorithm, Fujian Computer 11(2018) 108-117.

[7] X. Chen, C. Bai, Y. Huang, Research on face recognition system based on BP neural network, Intelligent Computers and Applications 8(3)(2018) 57-60.

[8] X. Cui, W. Cui, Partial occlusion of facial expression recognition based on convolutional neural network, Journal of Changzhou Vocational College of Information Technology 16(1)(2017) 12-14.

[9] X. Xu, H. Liu, Face expression recognition based on convolutional neural network, Foreign Electronic Measurement Technology 37(1)(2018) 106-110.

[10] H. Li, B. Shi, Face recognition algorithm based on Convolutional Neural Network, Software Guide 16(3)(2017) 26-29.

[11] Z. Wu, Multi-pose face recognition, [dissertation] Beijing: Beijing University of Posts and Telecommunications, 2017.