Face Recognition Based on Deep Learning

Weihong Wang(✉), Jie Yang, Jianwei Xiao, Sheng Li, and Dixin Zhou

Zhejiang University of Technology,
No.18 Chaowang Road, Hangzhou, 310023 Zhejiang, China
wwh@zjut.edu.cn, yangjie4699@163.com

Abstract. As one of the non-contact biometrics, face representation had been widely used in many circumstances. However conventional methods could no longer satisfy the demand at present, due to its low recognition accuracy and restrictions of many occasions. In this paper, we presented the deep learning method to achieve facial landmark detection and unrestricted face recognition. To solve the face landmark detection problem, this paper proposed a layer-by-layer training method of a deep convolutional neural network to help the convolutional neural network to converge and proposed a sample transformation method to avoid over-fitting. This method had reached an accuracy of 91% on ORL face database. To solve the face recognition problem, this paper proposed a SIAMESE convolutional neural network which was trained on different parts and scales of a face and concatenated the face representation. The face recognition algorithm had reached an accuracy of 91% on ORL and 81% on LFW face database.

Keywords: Facial landmark detection · Unrestricted face recognition · Deep learning · Convolutional neural network · SIAMESE network

1 Introduction

With the development of computer science and technology, face recognition have been widely applied to daily life and the environment, and the demands are also growing. Extracting and combining the semantic information of images needs effective pattern recognition algorithm. Traditional face recognition algorithm, such as the PCA [11], LDA [12], GABOR [9], LBP [10], etc., had certain deficiencies in precision and feature extraction.

In the process of human exploration, the neural network, a biologically inspired mathematical model was developed. It was an adaptive system, which could operate through a learning procedure. 2-layer BP network achieved high recognition accuracy(98%) for Mnist character database, but its convergence speed was rather slow, usually needed hundreds of times to converge for getting a satisfactory result, and easily converge to the local optimum solution [8].

In order to solve the problems mentioned above, we present convolutional neural network as the basic model to achieve the targets like fast convergence, signal noise suppression and high accuracy of feature points positioning. This model is also suitable for

Q. Zu et al. (Eds.): HCC 2014, LNCS 8944, pp. 812–820, 2015.
DOI: 10.1007/978-3-319-15554-8_73

facial landmark detection. Since the training of convolutional neural network needs massive samples [6], we propose sample transformation method in this paper to avoid over-fitting. Since Multi-input is needed, we novelly combine convolutional neural network and SIAMESE network for training on different parts and scales of a face and concatenating the face representation, to achieve the one-on-one face recognition.

2 Deep Learning

Deep Learning[5], through machine learning models with multi hidden layers and massive training data, could learn more useful features, and improve the accuracy of classification and prediction [3].

2.1 Convolutional Neural Network

The emergence of convolutional neural network solved some shortcomings of neural networks well, like the computational burden, the over-fitting of operation results and the lack of local characteristic. Through its local receptive field, sharing weights and the time domain or spatial domain samples, the displacement, scaling and distortion invariance of the results maintained [4].

The convolutional neural network could decrease the dimension dramatically by convolutional layers and pooling layers in a convolutional neural network, then outputted to a full-connected layer.

Convolutional Layer, shared weights:

$$C_{i,j,k}^{t} = g(\sum_{z=1}^{c_s} \sum_{y=1}^{w_c} \sum_{x=1}^{h_c} I_{i+x-1,\, j+y-1,\, c_k^t(z)}^{t-1} * F_{x,y,k}^{t} + B_k) \tag{1}$$

Convolutional Layer, unshared weights(UNSHARE):

$$C_{i,j,k}^{t} = g(\sum_{z=1}^{c_s,k} \sum_{y=1}^{w_c} \sum_{x=1}^{h_c} I_{i+x-1,\, j+y-1,\, c_k^t(z)}^{t-1} * F_{i,j,x,y,k}^{t} + B_k) \tag{2}$$

Pooling Layer:

$$I_{i,j,k}^{t} = f_{0<x\le d,\,0<y\le d}(C_{(i-1)*s+x,\,(j-1)*s+y,k}^{t}) \tag{2-3}$$

c_s :connections of the k-th unit in convolutional layer with the last layer. w_c :the width of a convolutional layer unit. h_c :the height of a convolutional layer unit. t :the number of a convolutional layer. c_k^t : the number of the k-th unit in convolutional layer with the last layer. I : the input of this layer. F :convolutional core. B :biasing.

The specific form of function g and f will be introduced when they were used.

2.2 SIAMESE Network

According to the present convolutional neural network, it could only support the function like $y = f(X)$, in which X was a vector to solve the actual problem, and y was the output of this module. The module didn't suit the situation which the classification or type was unknown. Therefore, we started to use Siamese Network based on normal convolutional neural network.

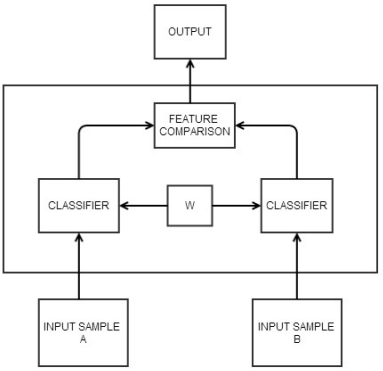


Fig. 1. Siamese Network

Fig.1 is Siamese Network of Probability, which supports the module of $y = f(X_1, X_2)$. X_1, X_2 is a vector of the actual problem, y is their similar probability. Using Siamese Network Module can solve multiple sample input and classification problem.

3 Facial Landmark Detection

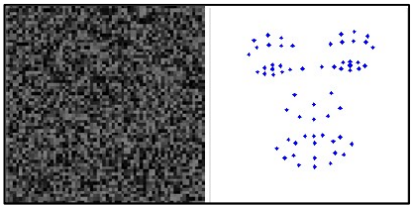
3.1 Model Analysis

This article mainly investigated on one-to-one recognition among the 2D face images in the condition of natural light which were taken by Optical camera. For the purpose of eliminating interferences like facial expression, shooting environment, the size of the picture, this paper used facial landmark detection, like eyes, eyebrows, nose, mouth and so on, to reduce the interference.

For the convenience of processing and increasing the accuracy and generalization ability, we made the facial landmark detection into two steps:(1) Positioning the face.(2) Facial landmark detection inside face.

The core problem of facial landmark detection was to consider two constraint problems:(1) Texture constraint. (2) Shape constraint. The facial texture constraint which was some parts of a face such as eyes, nose, mouth, etc. was presented by some local pixels. The facial shape constraint was the topological structure of those facial parts.

Luckily, the convolutional layer could exploit those local texture features, retrained some noise signal and the unshared layer could make the full use of the topological information from the training sample. The Fig. 2 shows the result of the inner facial landmark detection algorithm used on a random image. The result is quite face-like and stable.



(a) (b)

Fig. 2. Random Image (a) Result of Facial landmark detection (b)

3.2 Model Creation

The formula of convolutional neural network used here are implemented with 2-1,2-2,2-3, and the function g and f of face recognition are:

$$g(x) = |\tanh(x)|$$
$$f(x) = \max(x)$$

3.3 Training

The samples we used were partly from LFW [1], and also CAS-PEAL-R1 which developed by the Chinese academy of sciences, we artificially calibrated the feature points, and about 6400 samples were obtained, which contained a variety of conditions, such as the face with glasses, sunglasses, hats, and also profile face or different RACES. We mapped the 6400 samples into 12800 samples.

As Fig. 3 shows in this article, hierarchical training is divided into two convolutional neural networks. Each network has a supervised training [2], and after having adjusted the weights of the layer, we no longer update weights. Hierarchical training can avoid the situations that the underlying network is difficult to update effectively which is caused by excessive network layers, and makes the network to converge to a better level. And because of the reduction of computation, it greatly accelerates the convergence speed of the network.

In order to improve the generalization ability of training machine with limited samples and reduce the over-fitting occurred during training process, we aimed at the different characteristics of the two networks, did the sample transformation during stochastic gradient descent training process.

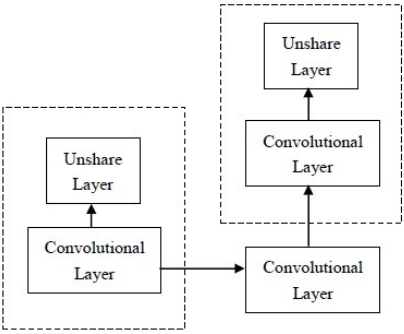


Fig. 3. Layer-by-layer training method

The first level of network — face positioning, was used to narrow the calculation scope of the second level of network and improve the generalization ability. So making some operations such as rotation, scaling, offset and other operations to the training samples in the first level could make network generalize these problem better.

The second level network—facial landmark detection, was used to reduce the noise from invalid texture constraints and invalid shape constraints caused by illumination, debris, facial expressions, etc. Therefore, to the second level of the sample, we could change the brightness, shade simulation to train shape constraints. Take shade simulation as an example, we fixed the pixel value of a region in an image as a random value, to simulate the effects of lighting or occlusions.

4 Face Recognition

4.1 Model Analysis

This paper used SIAMESE convolutional neural network cluster constructed on different parts of a face image such as eyes, nose, mouth, etc. to extract feature vectors and concatenate them as one feature vector to represent the face image [7].

Here the classifier in the SIAMESE network was the set of deep convolutional neural networks, the characteristics of this method was that it extracted the feature using machine learning model directly from different parts and scales of images, and then directly used for recognition.

4.2 Model Creation

The formula of convolutional neural network used here are implemented with 2-1,2-2,2-3, and the function g and f of face recognition are:

$$g(x) = \log(1 + e^x)$$

$$f(x) = \max(x)$$

4.3 Training

Samples mainly came from the result of facial landmark detection and obtained almost 8000 samples, which were selected from LFW and CASPEAL-R1. Adopting the layer-by-layer training method, the face region of each training layer was as followed: (1) Training the critical face region around eyes, nose and mouth by the first layer of convolutional neural network.(2) Training the critical face region around the left and right cheek by the second layer of convolutional neural network.(3) Training the critical face region based on the whole face by the third layer of convolutional neural network.

The training of SIAMESE network used the gradient based feedback algorithm and stochastic gradient descent. Because the deep network, we could use the same method as the third chapter (the layer-by-layer training method). The Fig.4 shows the structure of layer-wise training method as below.

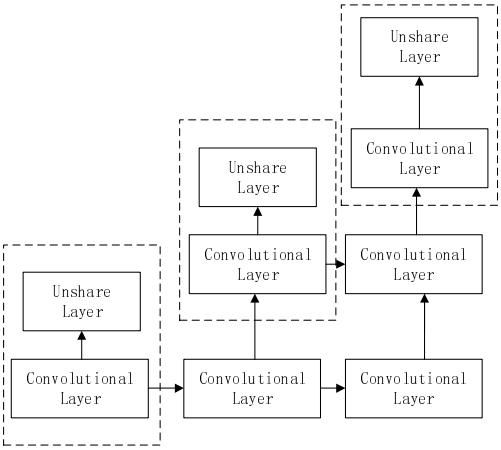


Fig. 4. Layer-wise training method

Training was achieved by two steps: feedforward and feedback.
Feedforward:

$$L(s^1, s^2, s^3, ..., s^M \mid h_1^1, h_2^1, h_1^2, h_2^2, ..., h_1^M, h_2^M; \alpha)$$
$$= -\sum_{m=1}^{M} (s^m \log(\mathrm{g}(h_1^m, h_2^m, \alpha)) + (1 - s^m) \log(1 - \mathrm{g}(h_1^m, h_2^m, \alpha)))$$

The training set consists of M pairs of input examples and corresponding similarity labels s^m, h_1^m is the first sample of the m-th input example. h_2^m is the second sample of the m-th input example. α is the positive coefficient that is optimized during learning. $\mathrm{g}(x)$ is the cost function of SIAMESE network.

Feedbak:

$$\frac{\partial L}{\partial \alpha} = \sum_{m=1}^{M} \frac{s^m - g^m}{g^m(1-g^m)} \left(\frac{\partial g(h_1^m, h_2^m, \alpha)}{\partial \alpha} \right)$$

$$\frac{\partial L}{\partial h_1^m} = \sum_{m=1}^{M} \frac{s^m - g^m}{g^m(1-g^m)} \left(\frac{\partial g(h_1^m, h_2^m, \alpha)}{\partial h_1^m} \right)$$

$$\frac{\partial L}{\partial h_2^m} = \sum_{m=1}^{M} \frac{s^m - g^m}{g^m(1-g^m)} \left(\frac{\partial g(h_1^m, h_2^m, \alpha)}{\partial h_2^m} \right)$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial h_1^m} \frac{\partial h_1^m}{\partial w} + \frac{\partial L}{\partial h_2^m} \frac{\partial h_2^m}{\partial w}$$

w is the parameters of training machine before the SIAMESE network, g^m is the output of m-th sample.

The advantage of using this function was to make the same sample have similar output as much as possible, and to make the output in different samples had larger gap, and the results would be normalized between 0 and 1 to show the similar probability.

5 Experiments

5.1 Results on Facial Landmark Detection

AT&T ORL: In AT&T ORL face database, the change of rotation and illumination among the face samples was rather small. And the image was clearer than others. There were no significant differences between the average error and sample error, so the over-fitting did not happen. The average error of facial landmark detection on AT&T ORL is shown in Fig.5 and the partial result of facial landmark detection on AT&T ORL is shown in Fig.6.

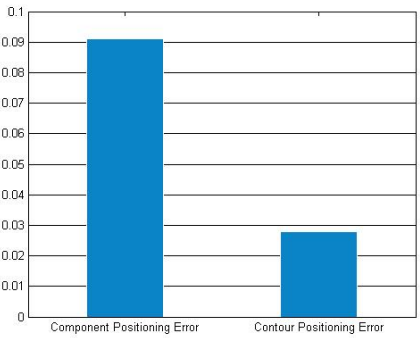


Fig. 5. The average error of facial landmark detection

Fig. 6. The partial result of facial landmark detection

5.2 Results on Face Recognition

AT&T ORL: In this face database, the change of samples is rather small, but the resolution of the image is relatively low, only about 92 * 92. In the process of testing, the image will be size up to 146 * 146. However, the actual image resolution in the training set is generally more than 146 * 146, so that the result remains to be improved. The result of face recognition on AT&T ORL is shown in Fig.7.

LFW: The images in LFW had the most expressions and the largest angle of face rotation. Beacause this article mainly researched positive face recognition, so it was not accidental that LFW performed the worst among face databases. The result of face recognition on LFW is shown in Fig.8.

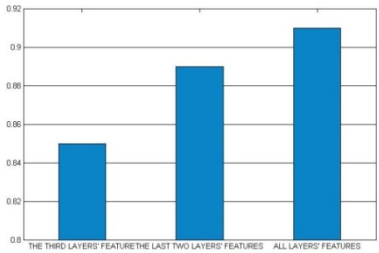


Fig. 7. The result of face recognition on AT&T ORL

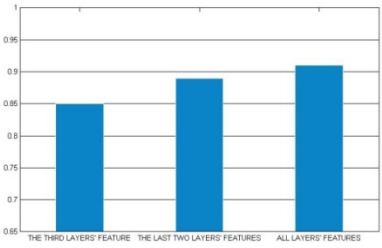


Fig. 8. The result of face recognition on LFW

6 Discussion and Conclusion

This paper studied the deep learning of the convolutional neural networks, via a layer-by-layer training method and a sample transformation method, made it converge

faster and avoiding over-fitting. SIAMESE convolutional neural network model effectively solved the problems of multi-input and unknown type of classification. By repeated tests on ORL and LFW, we achieved high accuracy in the facial landmark detection and face recognition. As a result of the limitation of time and samples, the training process of the method still needed further improvement. At last, this method can reliably bring certain economic benefits for manufacturing enterprises.

Acknowledgment. The work is supported by National Natural Science Foundation of China (Grant No.61340058) and Zhejiang Provincial Natural Science Foundation of China (Grant No. Z14F020006). The authors also would like to express appreciation to the experts and reviewers, who helped improve the paper with their valuable advice.

References

1. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, October 2007
2. Zhou, E., Fan, H., Cao, Z., Jiang, Y., Yin, Q.: Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In: 2013 IEEE International Conference on Computer Vision Workshops (ICCVW), pp. 386–391. IEEE (2013)
3. Abu-Mostafa, Y.S., Magdon-Ismail, M., Lin, H-T.: Learning from Data. AMLbook.com, March 2012
4. Bouvrie, J.: Notes on convolutional neural networks. (2006)
5. Huang, G.B., Lee, H., Learned-Miller, E.: Learning hierarchical representations for face verification with convolutional deep belief networks. In: 2012 IEEE Conference on, Computer Vision and Pattern Recognition (CVPR), pp. 2518–2525. IEEE (2012)
6. Wolf, L.: DeepFace: Closing the Gap to Human-Level Performance in Face Verification (2014)
7. Fan, H., et al.: Learning Deep Face Representation (2014). arXiv preprint arXiv:1403.2802
8. Lu, C., Tang, X.: Surpassing Human-Level Face Verification Performance on LFW with GaussianFace (2014). arXiv preprint arXiv:1404.3840
9. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local Gabor binary pattern histogram sequence (LGBPHS): a novel nonstatistical model for face representation and recognition. In: 2005 Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 1, pp. 786–791. IEEE (2005)
10. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(12), 2037–2041 (2006)
11. Zhao, W., et al.: Discriminant analysis of principal components for face recognition. In: Face Recognition, pp. 73–85. Springer, Heidelberg (1998)
12. Chen, L.-F., et al.: A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognition **33**(10), 1713–1726 (2000)