Face Recognition Based on improved Convolutional Neural Network

Ru-Yang Zhang*, Sang-Yun Park**, Eung-Joo Lee*
*Dept of Information and Communication Engineering, Tongmyong University
**Dept of Bachelor's Support Team, Silla University
e-mail: ejlee@tu.ac.kr

Abstract

In the past few years, with the fast development of face recognition technology, people lead a much better life than before. So as a result, research about developing face recognition technology, improving the recognition efficiency and time has become a hot field in face recognition studying. Base on that we proposed a face recognition system based on improved Convolutional Neural Network which can automatically extracts facial features and classifies them. The network trains feature extractors and classifiers by batch gradient descent. With the experiment using ORL and AR face image database, the results show a higher recognition rate, speed and robustness for different environment changes of face images.

1. Introduction

As we all know, nowadays face recognition is one of the hottest spots in recognition fields. It has been widely used in safety supervision, society security, traffic management and so on. It has advantages like fast speed, high reliability and high efficiency. But it also has disadvantages such as long consumption time and large work demand and can easily be effected by different environment. So in order to improve those disadvantages, this paper proposed a 6-layer convolutional neural network for face recognition. The first 4-layer using alternate concatenation of convolutional and pooled layers for feature extraction, then access a fully connected layer and the last layer uses the Softmax classifier with strong nonlinear classification ability. The whole Convolutional Neural Network system has a stable studying rate and using sigmoid as activation function.

2. 6-layer convolutional neural network model

The 6-layer convolutional neural network model proposed in this paper consists of two convolutional layers, two pooling layers, one fully connected layer and one Softmax regression layer. The specific composition is shown in Figure 1:

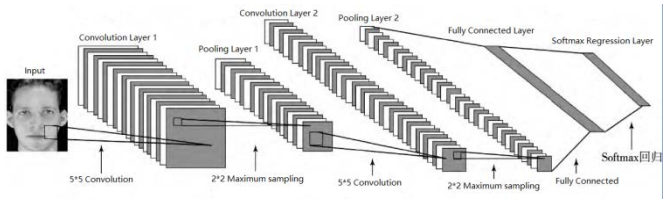

Figure1. 6-layer convolutional neural network

(1) Convolution Layer

The convolutional layer simulates the process of extracting some primary visual features from simple cells with local receptive fields through local connection and weight sharing. The above set of the same connection strength is a feature extractor, which appears as a convolution kernel during the operation process, and the convolution kernel value is randomly initialized first, and finally determined by network training. In order to extract face features more fully, we improved convolution layer with two methods, which are: add 2 convolutional feature extractors, which is the number of convolution kernels and the feature map of the convolutional layer 2 is related to all the feature maps of the pooled layer 1. The operating of convolution layer is shown in Figure 2:

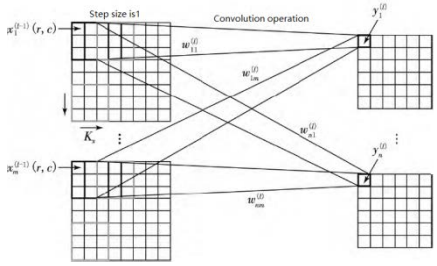

Figure2. Convolution layer operation

The mathematical expression for this layer is:

$$y_m^{(k)} = f\Big(\sum_{i \in M_{m(k-1)}} \sum_{(p,q) \in K^k} W_{mi(p,q)}^{(k)} \times x_i^{(k-1)}(c+p, r+q) + b_m^k\Big)$$

(2) Pooling Layer

The pooling layer simulates the process by which complex cells screen and combine primary visual features into more advanced, abstract visual features, which are implemented by sampling in the network. The pooling layer proposed in this paper uses the maximum sampling, and the sampling size is 2×2. Mathematical expression is:

$$y_m^{(k)} = \mathrm{down}(y_m^{(k-1)})$$

(3) Fully Connected Layer

In order to enhance the nonlinear mapping capability of the network and limit the size of the network, the network accesses a fully connected layer after four feature extraction layers. Each neuron in this layer is interconnected with all neurons in the previous layer, and the neurons in the same layer are not connected. Mathematical expression is:

$$S_m^k = f(\sum_{i=1}^{n} x_i^{(k-1)} \cdot w_{mi}^{(k)} + b_m^k) \cdot$$

(4) Softmax Regression Layer

Face features are more complex than digital features, and there are no uniform templates for face types. Therefore, the last layer of the network uses Softmax regression with strong nonlinear classification ability as the classifier.

3. Network training algorithm

Due to the large number of face database training samples, this paper uses a batch random gradient descent method which has faster convergence in practice. For the training of the convolutional neural network of the ORL face database, the batch block size is 40, the momentum is 0.9, and the learning rate is 0.12 constant. For the training of the convolutional neural network of the AR face database, the batch block size is 65, the momentum is 0.9, and the learning rate is 0.15 constant. Each iteration traverses all the batch blocks of the training set, and updating the network parameters once as soon as traverses a batch block each time.

For the over-fitting problem that the parameter is much larger than the number of training samples may cause, that is, the output of the input layer or the hidden layer is set as 0 with a certain probability p($0 \leqslant p \leqslant 1$) during network training, the neuron set to 0 no longer participates in the forward propagation and error back propagation of the network, and in the test, the input or hidden layer neuron output is multiplied by p as the output of the layer. Since each time we input a image the neurons in the network are randomly set to 0, it is required that the hidden layer neurons must extract the most essential features in the picture, so the "dropout" can be used to solve the over-fitting problem.

4. Experiment

The experimental data comes from the ORL and AR face databases. There are 600 different lighting conditions in the ORL database, different poses, and different facial expressions. Each person has 12 images totaling 50 people. Compared to ORL face database, AR face database is a large-scale color face image database. The database has a total of 4,000 images with 768 × 576 pixels, including 126 people. Each person in the library has 8 images under normal lighting conditions, 6 images under varying light conditions, and 4 images with different expression changes. The samples of those two face databases are as follows, first is the ORL face database:

Figure3.ORL face database

Second is the AR face database:

Figure4.AR face database

Do the following for the two databases: randomly extract 90% of the faces in the library as the training set, and the remaining 10% of the faces as the test set, and then normalize the faces in the 2 sets.

Compared the proposed convolutional neural network model with other algorithm based on both ORL and AR face image database, the results are shown in the Table1 and Table2 below:

Table1: Proposed network model compared with other algorithms based on ORL database

Recognition Methods	Recognition Rate (%)

Eigenface	97.02
ICA	94.15
2DPCA	98.27
Proposed Method	99.57

Table2: Proposed network model compared with other algorithms based on AR database

Recognition Methods	Recognition Rate (%)
PCA	86.12
2DPCA	95.98
PCA+GSRC	97.33
Proposed Method	99.69

Through the results we can see that the network system proposed in this paper has a better performance based on both face database, and the network has stronger ability to resist the interference of illumination, facial expression changes, and whether there is any obstruction or not.

5. Conclusion

This paper proposed a 6-layer convolutional neural network and uses it to do experiment on both ORL and AR face image database. The recognition rate for all the samples in those two databases are 99.57% and 99.69%. When it was used in AR face database the result showed that it robust to interference such as illumination differences, facial expression changes, and the presence or absence of obstructions. And also when we did the experiment based on the environment of Microsoft system and Matlab, the face recognition time of ORL and AR face database is very fast, so the real-time recognition is good.

참고문헌

[1] Nam H H, Kang B J, and Park K H. “Comparison of Computer and Human Face Recognition According to Facial Components,” Journal of Korea Multimedia Society, Vol. 37(21), pp.40-50, 2012.

[2] SYAFEEZA A R, KHALIL-HANI M, LIEW S S, et al. “Convolutional Neural Network for Face Recognition with Pose and Illumination Variation,” International Journal of Engineering & Technology, Vol. 6(1), pp. 44-57, 2014.

[3] Toshev A, Szegedy C. “Deeppose: Human Pose Estimation via Deep Neural Networks,” 2014 IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, pp. 1653-1660, 2014.

[4] Srivastava S, Hintion G, Krizhevsky A, et al. “Dropout: a simple way to prevent neural networks from over-fitting,” the Journal of Machine Learning Research, Vol. 15(1), pp. 1929-1958, 2014.

[5] Li Y X, Wang R. “Face Recognition based on Gabor Wavelet Feature Extraction and PCA,” Journal of Computer Knowledge & Technology, Vol. 11(32), pp.138-141, 2015.