Proceedings of the 36th Chinese Control Conference
July 26-28, 2017, Dalian, China

Face Recognition Based on Convolution Neural Network

Kewen Yan[1], Shaohui Huang[2], Yaoxian Song[1], Wei Liu[1], Neng Fan[1]

1. School of Automation, Hangzhou Dianzi University, HangZhou 310018
E-mail: yqdawujiang@163.com

2. College of Control Science and Engineering, Zhejiang University, Hangzhou 310013
E-mail: 1148940231@qq.com

Abstract: In this paper, a face recognition method based on Convolution Neural Network (CNN) is presented. This network consists of three convolution layers, two pooling layers, two full-connected layers and one Softmax regression layer. Stochastic gradient descent algorithm is used to train the feature extractor and the classifier, which can extract the facial features and classify them automatically. The Dropout method is used to solve the over-fitting problem. The Convolution Architecture For Feature Extraction framework (Caffe) is used during the training and testing process. The face recognition rate of the ORL face database and AR face database based on this network is 99.82% and 99.78%.

Key Words: Face Recognition, CNN, Image Processing, Caffe

1 Introduction

As one of the biometric identification technology, face recognition has many advantages, such as low cost, high reliability. It has been widely used in the fields of education, e-commerce, security and so on. In recent years, as a new research direction in the field of machine learning, remarkable achievements has been made in the field of deep learning. Deep learning has made in the field of image processing. CNN is a typical deep learning method, which has unique superiority in image processing because of its special structure of local weight sharing. In particular, the multi-dimensional input vector image can be directly input into the network, which avoids the feature extraction and reduces the complexity of data reconstruction in the classification process. Convolutional neural network has been successfully applied to character recognition [1], face recognition [2], human pose estimation [3] and target detection[4]. About the aspects of the face recognition, a hybrid neural-network for human face recognition was presented which compares favourably with other methods in 1997[5]. A novel hybrid face recognition approach based on a convolutional neural architecture was presented in 2009, designed to robustly detect highly variable face patterns[6]. A face recognition method based on multilayer neural networks (MNC), the K-nearest neighbor and support vector machines SVM was proposed in 2012[7].A human face recognition method based on adaptive deep CNN was presented in 2016[8]. This paper suggested a proper method for long-distance face recognition by resolving the change in recognition rate resulting from distance change in long-distance face recognition in 2016[9]. In 2017, CNN and Support Vector Machine (SVM) are combined to recognize face images in this paper[10].

CNN is a feature based method for face recognition. It is different from the traditional feature extraction and the design of high performance classifier. Its advantage is that the feature extraction is carried out by layer-by-layer convolution, then after multi-layer nonlinear mapping, the network can automatically learn to form the feature extractor and classifier suitable for the recognition task from the training samples without special pretreatment. This method reduces the requirement of training samples, and the more the network layers are set, the more global characteristics can be learned.

A face recognition method based on CNN is proposed in this paper. And the network used here consists of nine layers. These nine layers contains three convolution layers, two pooling layers, two full-connected layers and one Softmax regression layer. The convolution layers and the pooling layers are used for feature extraction followed by two full-connected layers, and the last layer uses a Softmax classifier with strong non-linear classification capability. And activation function of the network is ReLU function . Caffe is used during the network training process and GPU is used to expedite calculation speed. As to the training algorithm, stochastic gradient descent algorithm is used to train the feature extractor and the classifier, which can extract the facial features and classify them automatically. And the Dropout method is used to solve the over-fitting problem. The testing results on ORL face database and AR face database show that this network obtained higher recognition rate than that of the traditional ICA[17], PCA[15,18] methods and other kinds of face recognition method. Moreover, the network has excellent convergence and strong robustness.

2 The structure of the nine-layer network

This network consists of three convolution layers, two pooling layers, two full-connected layers and one Softmax regression layer. Each connection layer represents a linear mapping of different types of data. The network is shown in Fig. 1. The convolution layer and pool layer are composed of multiple feature maps, and each feature map is composed of multiple neurons. The feature map of each layer is the input of the next layer, and the feature map of the

*This work is supported by National Natural Science Foundation (NNSF) of China under grant 61375072, and Natural Science Foundation of Zhejiang Province under grant LQ16F030005.

convolution layer can be related to some feature maps of the previous layer.

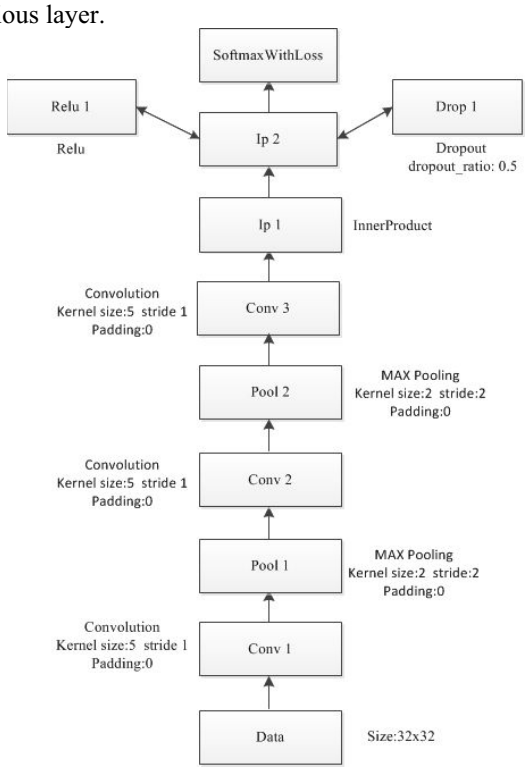

Fig. 1: The structure of the network

2.1 Convolution layer

The input of each convolution layer, like the traditional neural network, is the output of the upper layer and convoluted by several convolution kernels. The convolution kernels are used repeatedly in each sensory fields of the entire region, and the convolution result constitutes a feature map of the input image. And the convolution kernels are the contents to be learned by the convolution layer, including the weight matrix w and the bia s b . In this paper, the size of convolution kernel is 5×5 . w is initialized by the "xavier" algorithm, b is initialized with 0, they will be finally determined by network training process.

The mathematical expression of the layer [2] is:

$$x_j^l = f\left(\sum_{i\in M_j^{l-1}} x_i^{l-1}k_{ij}^l + b_j^l\right) \tag{1}$$

Where l represents the layer, k is the convolution kernel, b is the bias, and M_j represents the feature map.

2.2 Pooling layer

The output feature maps obtained after the calculation of the convolution layer are generally not much reduced in dimension. If the dimension does not change, there will be a great amount of computation need to do, and the network learning process will become very difficult, more likely to get a reasonable result.

The pooling layer is generally a layer to reduce the dimension of the feature map, and it is a non-linear down-sampling method. In the network, each feature map that has been put into the pooling layer is sampled, and the number of output feature maps is unchanged, but the size of each feature map will be smaller. Thus, the purpose of reducing the amount of calculation and resisting the change of micro displacement is achieved. In this paper, the pooling layer is sampled with the maximum value. The sampling size is 2x2 and the step size is 2.

2.3 Full-connected layer

For the network, after several continuous stack of convolution layers and pooling layers, generally, there will be a number of full-connected layers near the output layer. And these full-connected layers form a multi-layer perceptron (MLP), which plays the role of a classifier.

In this paper, we use two full-connected layers, each of which is connected to all the neurons in the previous layer. The mathematical expression of the layer [11] is:

$$y_{pj} = f\left(\sum_{i=1}^{n} x_i^{l-1}w_{ji}^l + b_j^l\right) \tag{2}$$

Where n is the number of neurons in the preceding layer $(l-1)$, w_{ji}^l is the weight for connection from neuron i in layer $(l-1)$ to neuron j in layer (l) and b_j^l is the bias of neuron j in layer (l), and f represents the activation function of layer (l).

2.4 Softmax regression layer

As the face characteristics are more complex, and the face category is more and there is no uniform template, the softmax classifier which has a strong non-linear classify ability is used at the last layer of the network. Softmax classifier is a multi-classifier, which can not only complete the dichotomy task, but also can complete the multiples (greater than 2) task.

Suppose that m samples can be divided into k classes, the training set is $\left\{\left(\left(x^{(l)},y^{(l)}\right),\cdots,\left(x^{(m)},y^{(m)}\right)\right)\right\}$, $x^{(i)}\in R^{n+1}, y^{(i)}\in\{1,2\cdots,k\}$, the Softmax regression function[12] is:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)}=1\mid x^{(i)};\theta) \\ \vdots \\ p(y^{(i)}=k\mid x^{(i)};\theta) \end{bmatrix} = \begin{bmatrix} \exp(\theta_1^T x^{(i)}) \\ \vdots \\ \exp(\theta_k^T x^{(i)}) \end{bmatrix} \Big/ \sum_j^k \exp(\theta_j^T x^{(i)}) \tag{3}$$

where $p(y^{(i)}=1\mid x^{(i)};\theta)$, and it presents the probability that $x^{(i)}$ belonging to class j, $\theta_j^T\in R^{n+1}$ is the parameters of the model.

The loss function of the model is:

$$J(\theta)=\frac{-1}{m}\left[\sum_{i=1}^{m}\sum_{j-1}^{k} l\left\{y^{(i)}=j\right\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] \quad (4)$$

$l\left\{y^{(i)}=j\right\}$ is indicative function, if $y^{(i)}=j$ is true, then $l\left\{y^{(i)}=j\right\}=1$, otherwise $l\left\{y^{(i)}=j\right\}=0$

3 Training the network

In order to study the effect of the network model proposed in this paper, we use the Caffe framework to train and test the ORL face database and the AR face database in this paper. ORL face database consists of 40 people, 10 photos for per person, a total of 400 pictures, including facial changes, small posture changes and the scale changes less than 20%. AR face database contains 100 people face images, including 50 men and 50 women, 26 photos for per person, a total of 2600 pictures, And the pictures have face expression change, block change, light change and other changes. Two people's pictures are selected randomly from the ORL face database, as is shown in Fig. 2. And one people's pictures are selected randomly from the AR face database, as is shown in Fig. 3.

Fig. 2: ORL dataebase sample

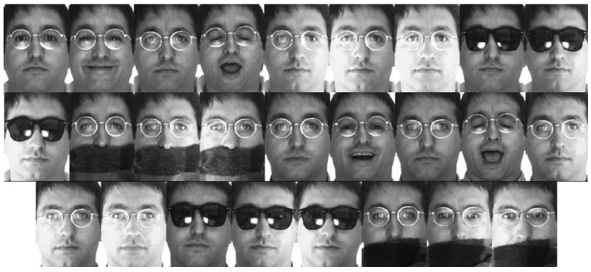

Fig. 3: AR database sample

3.1 Data preprocessing

Firstly, process the pictures from the two databases as follows: Use the Caffe tool to make all the images into 32x32 pixel size, then do the mirror symmetry. After that, normalize the input image data from [0-255] to [0-1]. Finally, 90% of the human faces in the database were selected as the training set, and the remaining 10% were used as the test set.

3.2 Network training algorithm

In this paper, the stochastic gradient descent method which has the fast convergence rate is used. The learning rate (*base_lr*) is initialized to 0.01. During the training process, the learning rate is updated as:

$$base\_lr*\left(1+gamma*iter\right)\wedge(-power) \quad (5)$$

Where *iter* is the number of iterations, $iter=100$, $gamma=0.0001$, $power=0.75$ in this paper. At the same time, each iteration traverse through all the batch blocks of the training set, the network parameters are updated after traversing one batch block, and the network parameters are updated once a batch is completed. The updating formula is:

$$v^{i+1}=\varepsilon v^{i}-\alpha\cdot\nabla J(\theta)^{(i)} \quad (6)$$

$$\theta^{(i+1)}=\theta^{(i)}+v^{i+1} \quad (7)$$

Where i is the iteration number, $\nabla J(\theta)^{(i)}$ presents partial derivative of the loss function $J(\theta)^{(i)}$ in the i group, v is the impulse parameter, α is the learning rate which equals to $base\_lr$. Attenuation coefficient λ in $\nabla_{\theta_i} J(\theta)^{(i)}$ is recommend as 0.0005[13]. Although λ is small, experimental results show that it can effectively improve the classification accuracy.

4 Experimental results

In this paper, the Ubuntu16.04 operation system is used as the experimental environment and the Caffe framework is used as training and testing tool. The computer CPU is Inter Xeon (R) E3-1231, clocked at 3.40GHz, and the memory size is 4G, use GPU to accelerate computation speed, and graphics card is NVIDIA GeForce GTX 1060.

4.1 Analysis of experimental results

Fig. 4 and Fig. 5 show the loss value and accuracy of when training the AR face database. Fig. 6 and Fig. 7 show the loss value and accuracy of when training the ORL face database. After training, these figures will be conveniently obtained using the Caffe tools to analyze and process the training data.

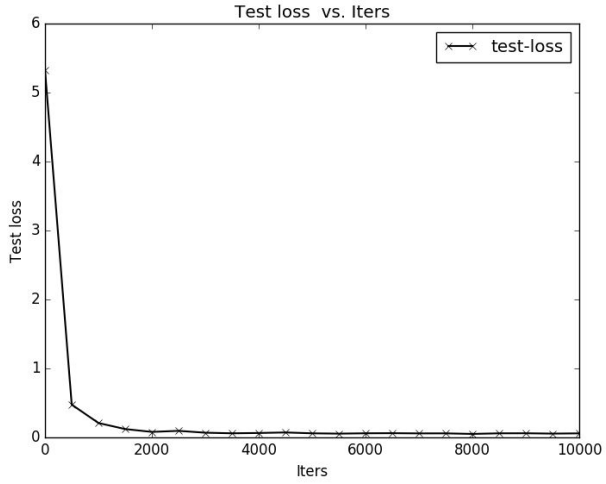

Fig. 4: The change of loss when training the AR database

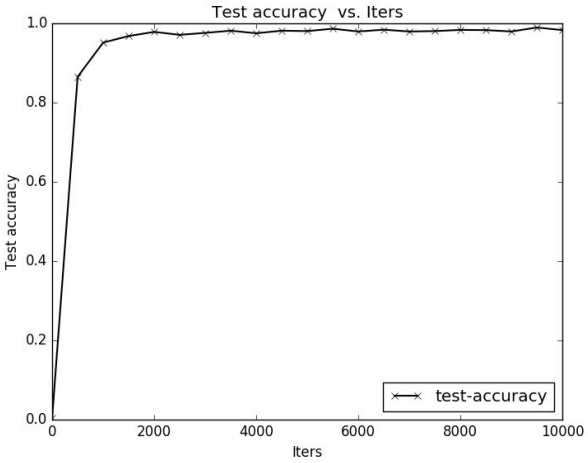

Fig. 5: The change of accuracy when training the AR database

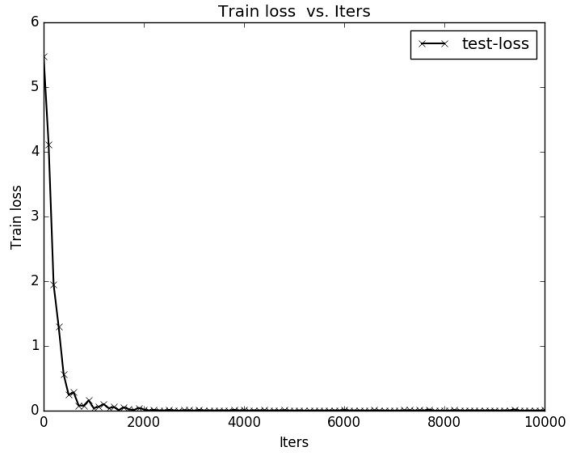

Fig. 6: The change of loss when training the ORL database

It is seen from Fig. 4 and Fig. 6 that the network has a good convergence, and the loss value can quickly reduce. It becomes stable after 2000 iterations, and which is close to 0. It is seen from Fig. 5 and Fig. 7 that the network model can quickly reach the accuracy rate of 90%, and is close to 97% after 2000 iterations. With the increase of the number of training iteration, it increases slowly and smoothly, and finally stabilized at around 98.8%.

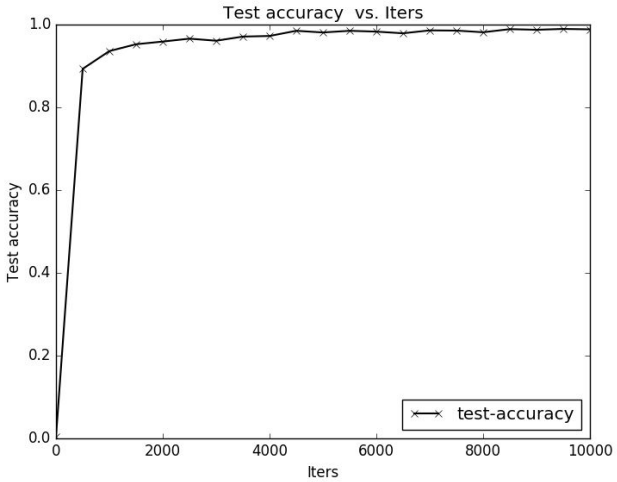

Fig. 7: The change of accuracy when training the ORL database

So, the network model proposed in this paper not only has good convergence, but also has high accuracy. The correct rate of face recognition for AR face database is 99.78%, and the correct rate of face recognition for ORL face database is 99.82%. Moreover, the network has a strong resistance to the change of facial expression, and whether there is occlusion or not.

4.2 Comparison with other methods

Table 1 and Table 2 show of the proposed network model in this paper with other models in ORL and AR face database respectively.

Table 1: the comparison of the accuracy for AR database

Method	accuracy（%）
DWT[14]	90.80
PCA+GSRC[15]	97.14
LC-KSVD[16]	97.80
CNN in this paper	99.78

Table 2: the comparison of the accuracy for ORL database

Method	accuracy（%）
Eigenface[17]]	97.50
ICA[17]	93.75
2DPCA[18]	98.30
CNN in this paper	99.82

5 Conclusions

A face recognition method based on convolution neural network (CNN) is presented in this paper. And the network has nine layers. The Caffe is used during the training and testing process. In experiments on the testing set of ORL face database and AR face database, the recognition rate are 98.95% and 98.30% respectively, In experiments on all the data of ORL face database and AR face database, the recognition rate are 99.82%and 99.78% respectively. Moreover, the network has excellent convergence and strong robustness.

References

[1] Ahranjany S S, Razzazi F, Ghassemian M H. A very high accuracy handwritten character recognition system for Farsi Arabic digits using convolutional Neural Networks[C]. *Theories and Applications(BIC-TA),2010 IEEE Fifth International Conference on Bio-Inspired Computing. Bejing: IEEE,* 20l0: 1585—1592.

[2] Syaffeza A R, Khalil-Hani M, Liew S S, et al. Convolutional neural network for face recognition with pose and Illumination Variation[J]. *International Journal of Engineering & Technology*, 2014, 6(1): 44—57.

[3] D. Cheng, Controllability of Switched Bilinear Systems. *IEEE Trans. on Automatic Control*, 2005, 50(4): 511-515.

[4] Toshev A, Szegedy C. Deeppose: Human pose estimation via deep neural networks[C]. *2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Los AIamitos: IEEE*, 2014: 1653—1660.

[5] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: a convolutional neural-network approach [J]. *IEEE Transactions on Neural Networks*, 1997, 8(1):98.

[6] R?zvanDaniel Albu. Human Face Recognition Using Convolutional Neural Networks[J]. *Journal of Electrical & Electronics Engineering,* 2009, 2(2):110.

[7] M. Fakir, B. Bouikhalene. Face recognition based on statistical approaches, neural networks and support vector machines. *International Research Journal of Computer Science and Information Systems.* 23 November, 2012.

[8] Chen L, Guo X, Geng C. Human face recognition based on adaptive deep Convolution Neural Network[C]. *Chinese Control Conference.* 2016:6967-6970.

[9] Moon H M, Chang H S, Pan S B. A face recognition system based on convolution neural network using multiple distance face[J]. *Soft Computing,* 2016:1-8.

[10] Guo S, Chen S, Li Y. Face recognition based on convolutional neural network and support vector machine[C] *IEEE International Conference on Information and Automation.* IEEE, 2017.

[11] Ziming Z, Song F. Profiling and analysis of power consumption for virtualized systems and applications[C]. *Proceedings of IEEE 29th International Performance Computing and Communications Conference IPCCC),* 2010: 329-330.

[12] Huang V S, Shadmehr R, Diedrichsen J. Active learning: learning a motor skill without a coach[J]. *Journal of neurophysiology,* 2008, 100(2): 879—887.

[13] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. *Advanees in neural information processing systems*. 2012: 1097-1105.

[14] Yaji G S, Sarkar S, Manikantan K, et a1. DWT feature extraction based face recognition using intensity mapped unsharp masking and laplacian of gaassian filtering with scalar multiplier[J]. *Procedia Technology*, 2012, 6: 475-484.

[15] Yang M, Zhang L. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary[M]. *Computer Vision—ECCV 20lO. Berlin Heidelberg: Springer,* 2010: 448—461.

[16] JiangG Z, Lin z, Davis L S. Learning a discriminative dictionary for sparse coding via label consistent K—SVD[C]. *2011 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Washington：IEEE,* 2011: 1697—1704.

[17] Yang M H, Kernel eigenfaces vs. kernel fisherfaces：Face recognition using kernel methods[C]. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, *Washington: IEEE*, 2002: 0215—0215.

[18] Yang J, Zhang D, Frangi A F, et a1. Two-dimensional PCA: a new approach to appearance-based face representation and recognition[J]. *Pattern Analysis and Machine Intelligence. IEEE Transactions on,* 2004, 26(1): 131—137.