Handwritten Digit Recognition Based on Depth Neural Network

Yawei Hou
School of Electrical and Electronic Engineering
Shanghai Institute of Technology
Shanghai , China
houyawei2017@163.com

Huailin Zhao
School of Electrical and Electronic Engineering
Shanghai Institute of Technology
Shanghai , China
zhl@sit.edu.cn

***Abstract*—Neural network and depth learning have been widely used in the field of image processing. Good recognition results are often required for complex network models. But the complex network model makes training difficult and takes a long time. In order to obtain a higher recognition rate with a simple model, the BP neural network and the convolutional neural network are studied separately and verified on the MNIST data set. In order to improve the recognition results further, a combined depth network is proposed and validated on the MNIST dataset. The experimental results show that the recognition effect of the combined depth network is obviously better than that of a single network. A more accurate recognition result is achieved by the combined network.**

Keywords—convolution neural network; handwritten digit recognition; combinatorial network; Gabor filter

I. INTRODUCTION

Optical character recognition technology includes handwritten character recognition and printed character recognition. Handwritten digital recognition as part of handwritten character recognition is also a very important research direction. Handwritten digital recognition mainly identifies 0-9 of 10 characters, and the category of classification is much less than optical character recognition[1]. In recent years, along with the development of computer technology and pattern recognition technology, handwritten digital recognition has been widely used in postal code, financial value identification, tax form recognition, e-commerce digital processing, and even student achievement recognition[2]. Although the classifier has been further enriched, but the researchers still didn't find a way to achieve the perfect effect of the algorithm.

Artificial neural network has been widely used in character recognition because of its strong self-learning ability, adaptive ability, classification ability, fault tolerance and fast recognition[3]. In recent years, with the development of deep learning, convolutional neural network (CNN) has been widely used in image processing. CNN has some characteristics such as local connection, weight sharing and pooling. The features extracted by CNN model have good descriptive power. Artificial neural networks commonly used BP neural network and probabilistic neural network[4]. In this paper, BP neural network and convolution neural network are studied and verified on MNIST data set. In order to improve the recognition rate of the network, this paper designs a combined network of CNN and BP neural network. Through the experiment on the MNIST data set, it is verified that the classification result of the combined network is better than that of the single network.

978-1-5090-6664-3/17/$31.00 ©2017 IEEE

II. NEURAL NETWORK

A. BP Neural Network

BP neural network is a multi-layer feedforward neural network. The main feature of the network is the signal forward transmission and the error back propagation. The topology of BP neural network is shown in Fig 1.

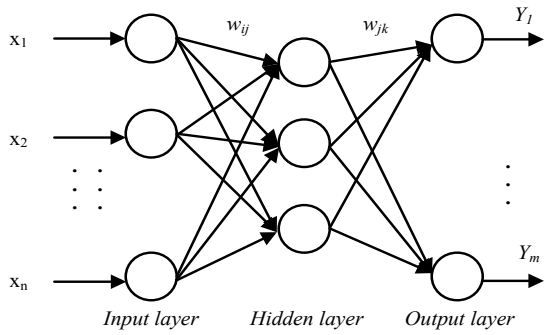

Fig. 1. The topology of BP neural network

BP neural network using the back-propagation algorithm andthe algorithm can be described as follows:

(1) Input x: set the corresponding activation value for the input layer a^1.

(2) Forward propagation: calculate the corresponding weighted input z^l and the activation value al for each l=2,3,...,L.

$$z^l = w^l a^{l-1} + b^l \tag{1}$$

$$a^l = \sigma\left(z^l\right) \tag{2}$$

In the equation (2), σ(.) represents the activation function.

(3) Calculation of output layer error δ^L

$$\delta^L = \nabla_a C \bullet \sigma'\left(z^L\right) \tag{3}$$

Where • represents the Hadamard product.

(4) Reverse error propagation: calculate δ^l for each l=L-1,L-2,...,2.

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \bullet \sigma'(z^L) \tag{4}$$

(5) Output: the gradient of the cost function is given by equation (5) and equation (6).

$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l \tag{5}$$

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l \tag{6}$$

B. Convolutional Neural Network

In 1980, the new cognitive machine was put forward in literature [5], and the concept of CNN was introduced for the first time. It became the first model of depth learning. In 2003, the literature [6] summarized the CNN. As shown in Fig 2, the convolutional neural network is a multilayer feedforward network. Each layer consists of a number of two-dimensional planes and each plane consists of multiple neurons. In Fig 2, C1 and C3 are convolutions, and S2 and S4 are subsampled. Convolution layer and the subsampling layer can have multiple layers. Generally, the first layer of CNN is the alternating layer of convolution and subsampling, and the depth of CNN is reflected [7].

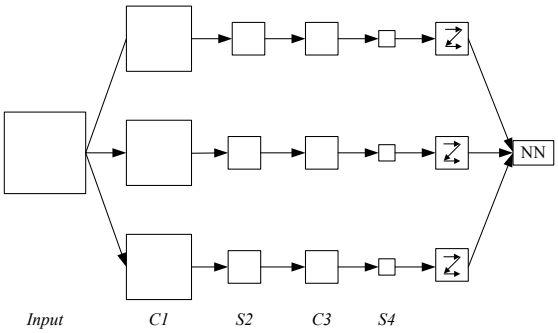

Fig. 2. The structure of convolution nerve

Convolution is an algorithm commonly used in image recognition. It refers to each pixel in the output image being weighted by the pixels of the small area corresponding to the position of the input image. This small area is called the local experience field, the regional weight is called the convolution kernel. The biased term is added after the input image is convoluted, and the feature graph is obtained by activating the function. Equation (7) gives the form of the convolution layer:

$$X_j^l = f(\sum_{i \in M_j} X_i^{l-1} * K_{ij}^l + b_j^l) \tag{7}$$

Where: l is the number of layers; X^l_j is the jth feature graph of the convolution layer l; M_j is the receptive field of the input layer; K is the convolution kernel; b is the bias; f is the activation function of the neuron.

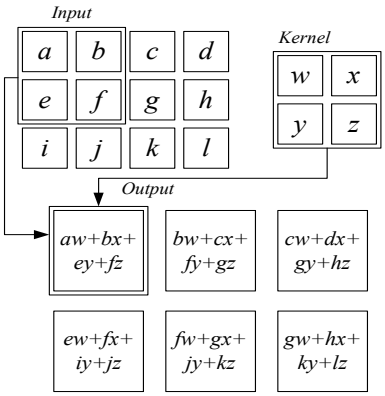

Fig. 3. Convolution process diagram

Fig 3 shows the convolution process and the resulting feature is 2 × 3.

The downsampling layer is also called a pooling layer. The image is divided into small pieces of small areas, and a value is calculated for each region. Then the calculated values are arranged in sequence to output the new image. This process is equivalent to fuzzy filter, which can improve the robustness of image feature extraction. The pooling method adopted in this paper is average pooling.

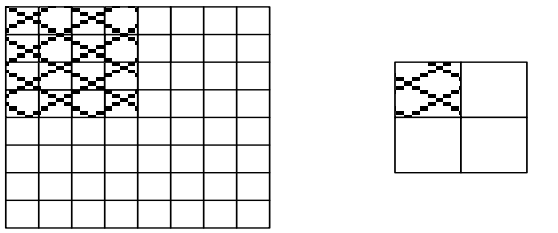

Fig. 4. Pooling process

Fig 4 (a) The original feature graph is an 8×8 matrix, where the 4×4 grid region represents the pooled domain. When the original feature pool is pooled, the moving step size is 4, and the sub sampling feature graph of Fig 4 (b) is obtained, which is 2×2 matrix. The gray region of the subsampled feature map is the pooling result corresponding to the original feature grid region.

C. Learning Rate Adaptive Adjustment Algorithm

Different learning rates can have a great impact on BP neural networks. If the learning rate is too large, the network training is not convergence; in contrast, the network training time is too long. Therefore, you need to choose a best learning rate for the BP algorithm. In order to achieve the best learning rate, the learning rate must be adjusted during the training process.

There are many ways to modify the learning rate. The method used in this paper is to automatically adjust the learning rate by checking whether the correction value of the network weights reduces the error function[8]. So that the

978-1-5090-6664-3/17/$31.00 ©2017 IEEE

network is always trained with the maximum acceptable learning rate.

$$\eta(t+1)=\begin{cases}1.05\eta(t) & E(t)<E(t-1)\\ 0.7\eta(t) & E(t)>1.04E(t-1)\\ \eta(t) & other\end{cases} \quad (8)$$

Using the learning rate adaptive adjustment algorithm which given in Equation (8), it is possible to shorten the training time of the network and make the learning algorithm more reliable.

III. Experimental Results and Analysis

The data set used in this article is a recognized MNIST dataset. The MNIST dataset is handwritten numbers and includes 60,000 training samples and 10,000 test samples.

A. BP Neural Network Test

By comparing the different feature extraction methods of BP neural networks, the network correctness is shown in Table 1.

TABLE I. BP Neural Network Recognition Rate With Different Feature Extraction

Feature extraction method	Correct rate
No feature extraction	98.37%
PCA feature extraction	98.30%
Combination of Sobel and PCA Feature Extraction	98.56%
Gabor feature extraction	99.15%

BP neural network structure is: input layer - hidden layer 1 - hidden layer 2 - output layer. The neurons of the hidden layer 1 are 500, the neurons 2 of the hidden layer 2 are 300, and the neurons in the output layer are 10. The learning rate of the network is 0.25, and using the learning rate adaptive adjustment algorithm.

In the PCA feature extraction, 99% of the variance is retained. In the combination of Sobel and PCA feature extraction, the input image is passed through the Sobel operator to obtain four horizontal images in the horizontal direction, the vertical direction and the two diagonal directions [9]. In the Gabor feature extraction, 5 scales and 8 directions are selected, and 40 Gabor filters are used. In this paper, the feature image is sampled four times under reference [10], and the resulting eigenvector has the size of 1960.

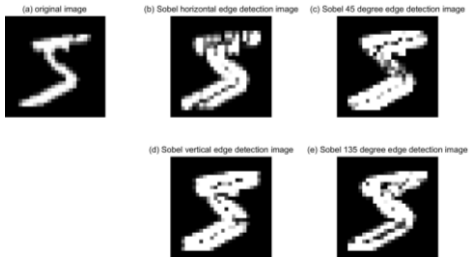

Fig. 5. The original image and the 4 gradient images obtained by the Sobel operator

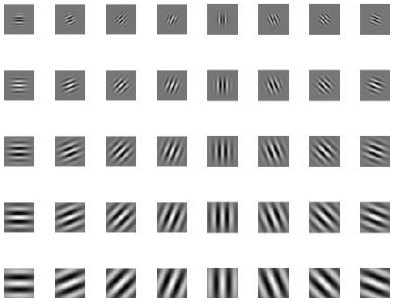

Fig. 6. Gabor filter with 5 scales and 8 directions

B. Convolutional Neural Network Test

In the study of handwritten character recognition, the structure of the convolutional neural network is established as shown in Table 2.

TABLE II. CNN Structural Parameters

Number of layers	type	Output feature dimension	Filter size
0	input layer	28×28	---
1	Convolution layer	10×24×24	5×5
2	Sub-sampling layer	10×12×12	2×2
3	Convolution layer	20×8×8	5×5
4	Sub-sampling layer	20×4×4	2×2

The activation function of the network is sigmoid function, the sub-sampling uses the mean sampling, and the network output is 10 neurons.

Because the learning rate is very important to the final recognition result of the network, this paper adopts the learning rate adaptive adjustment algorithm. By using the MNIST data set to test, a recognition rate of 99.23% was obtained.

In order to obtain better network recognition rate, this paper improves the structural parameters of Table 2, and obtains the network structure shown in Table 3.

TABLE III. CNN Structural Parameters

Number of layers	type	Output feature dimension	Filter size
0	input layer	28×28	---
1	Convolution layer	10×24×24	5×5
2	Sub-sampling layer	10×12×12	2×2
3	Convolution layer	20×8×8	5×5
4	Sub-sampling layer	20×4×4	2×2
5	Convolution layer	40×2×2	3×3
6	Sub-sampling layer	40×1×1	2×2

978-1-5090-6664-3/17/$31.00 ©2017 IEEE

The network learning rate changed to 1.5, the other parameters have not changed. The correct rate is 99.43% by testing on the MNIST dataset.

C. Combination Neural Network Experiment

In order to get a better recognition rate, this paper analyzes the network identification results under different feature extraction. The convolutional neural network learns the features of the original image through the cooperation of the multilayer and the lower sampling layer, while the Gabor feature extraction is performed by the Gabor filter. According to the difference of feature extraction, this paper presents a combined neural network. The structure of the network shown in Fig 7.

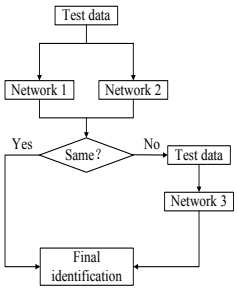

Fig. 7. Block diagram of the combined network structure

The network 1 and the network 2 recognize the test data at the same time. If the recognition result of the network 1 and the network 2 is the same, the result is the correct recognition result of the test data. If the recognition result of the network 1 and the network 2 is different, The test data is identified by the network 3, and the final recognition result of the test data is the result of the recognition of the network 3. There are three ways to combine three networks, so Table 4 shows the recognition results for different combinations.

TABLE IV. IDENTIFY THE RESULTS OF DIFFERENT COMBINATIONS

Combination method	Error rate
Gabor feature extraction and Table 2 CNN	0.45%
Gabor feature extraction and Table 3 CNN	0.47%
Table 2 and Table 3 CNN	0.46%

By analyzing Table 4, we can get the best result of a combined network of 99.55%. Compared with the Reference [11], the recognition rate increased 0.14%. Compared with the Reference [12], the recognition rate increased 0.08%. Compared with the Reference [13], the recognition rate increased 0.35%.

IV. CONCLUSION

In order to study the network model with high recognition rate, this paper firstly studies the double hidden layer BP neural network with different feature extraction. By analyzing Table 1, it is found that the Gabor feature extraction method can obtain a relatively high recognition rate compared with other feature extraction methods in Table 1. Secondly, the CNN is studied. In CNN, the time of the network training is longer with the increase of the convolution layer. The network will be difficult training because of many layers. Therefore, the CNN structure is relatively simple in this paper, and the network structure is given in Table 2 and Table 3. The recognition results will be limited because of the simple structure of CNN. The best result of the CNN used in this paper is 99.43%. In order to obtain a network model with a higher recognition rate, this paper finally studies the combined depth neural network. By experimenting with the MNIST data set, the optimal result of the combined depth network is 99.55%. Compared with the three single network models, the recognition results are improved.

REFERENCES

[1] Yang Shuying, Image recognition and project practice. Beijing: Publishing House of Electronics Industry, 2014.

[2] Basu S., Das N., Sarkar R., Kundu M., Nasipuri M., Basu D.K. (2009) Recognition of Numeric Postal Codes from Multi-script Postal Address Blocks. In: Chaudhury S., Mitra S., Murthy C.A., Sastry P.S., Pal S.K. (eds) Pattern Recognition and Machine Intelligence. PReMI 2009. Lecture Notes in Computer Science, vol 5909. Springer, Berlin, Heidelberg.

[3] Impedovo S, Pirlo G, Modugno R, Ferante A, "Zoning Methods for Hand-Written Character Recognition： An Overview," International Conference on Frontiers in Handwriting Recognition, IEEE Computer Society, 2010, pp. 329-334.

[4] Huailin Zhao, Yawei Hou, Shifang Xu, Congdao Han, Masanori Sugisaka, "Research on a Method of Character Recognition for Self-learning Errors," The 2017 International Conference on Artifical Life and Robotics, Japan, 2017, pp.190-193.

[5] Deng L, Hinton G, Kingsbury B, "New types of deep neural network learning for speech recognition and related applications : an overview," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver BC: IEEE, 2013, pp. 8599-8603.

[6] Xu Peng, Bo Hua, "Facial expression recognition based on convolutional neural networks," Microcomputer & Its Applications, no. 12, pp. 45-47, 2015.

[7] Zhang C, Zhang Z Y, "Improving multiview face detection with multi-task deep convolutional neural networks," Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO:IEEE, 2014. pp. 1036-1041.

[8] Wu Meixian, Zhang Xueliang, Wen Shuhua, Guo Qin, "BP neural network dual learning rate adaptive learning algorithm," Modern Manufacturing Engineering, no. 10, pp. 29-32, 2005.

[9] Yi Chaoren, Deng Yanni, "Multichannel convolution neural network image recognition method," Journal of Henan University of Science & Technology(Natural Science), no. 3, pp. 41-44, 2017.

[10] M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," Expert Systems with Applications, vol. 42, no. 21, pp. 7905-7916, 2015.

[11] Wan L, Zeiler M, Zhang S, LeCun Y, Fergus R, "Regularization of neural networks using dropconnect," Proc of the 30th International Conference on Machine Learning, 2013, pp. 1058-1066.

[12] Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y, "What is the best multi-stage architecture for object recognition?" Proc of IEEE 12th International Conference on Computer Vision, 2009, pp. 2146-2153.

[13] Chen Haoxiang, Cai Jianming, Liu Kengran, Lin Qiushuang, Zhang Wenling, Zhou Tao, "Handwritten digital depth feature learning and recognition," Computer Technology and Development, vol. 26, no. 7, pp. 19-23, July 2016.

978-1-5090-6664-3/17/$31.00 ©2017 IEEE