Handwritten Digit Recognition System Based on Convolutional Neural Network

Jinze Li
College of Mechanical and Electrical Engineering
Shenyang Aerospace University
Shenyang, China

Gongbo Sun
College of Mechanical and Electrical Engineering
Shenyang Aerospace University
Shenyang, China

Leiye Yi
College of Mechanical and Electrical Engineering
Shenyang Aerospace University
Shenyang, China

Qian Cao
College of Computer Science
Nanjing University of Science and Technology Zijin College
Nanjing, China

Fusen Liang
College of Mechanical and Electrical Engineering
Shenyang Aerospace University
Shenyang, China

Yu Sun
College of Mechanical and Electrical Engineering
Shenyang Aerospace University
Shenyang, China

***Abstract*—Image recognition is widely used in the field of computer vision today. As a kind of image recognition, digit recognition is widely used. Today, the online recognition technology in digit recognition is relatively mature while the offline recognition technology is not. This paper mainly introduces an offline recognition system for handwritten digits based on convolutional neural networks. The system uses the MINST dataset as a training sample and pre-processes the picture with the Opencv toolkit. Then it uses LeNet-5 in the convolutional neural network to extract the handwritten digit image features, repeatedly convolution pooling, and pull the result into a one-dimensional vector. And finally find the highest probability point to determine the result to achieve handwritten digit recognition with the Softmax regression model. The application of this system can greatly reduce labor costs and improve work efficiency, which is of great significance in many fields.**

Keywords—handwritten digit recognition system, deep learning, convolutional neural network

I. INTRODUCTION

With the rapid development of electronic information, computer input has become more and more common, but handwriting is still an irreplaceable way for people to transfer information. As a link combining handwritten characters and computer input, handwriting recognition has received more and more attention for its practicability. Handwriting recognition technology is the basis of handwriting interpolation and handwriting identification. In the past decade, machine learning and pattern recognition have extended many highly intelligent handwriting recognition classifications, including artificial neural networks (ANN) [1], support vector machine (SVM) [2], modified quadratic discriminant function (MQDF) [3] and hidden Markov model [4], etc.

Arabic numerals are the only universal symbols in the world, and they have an indelible contribution to the development of world science and culture. Convolutional neural networks have advantages in image processing in all neural networks, so this paper designs a handwritten digit recognition system based on convolutional neural networks. Handwritten digit recognition is divided into online recognition and offline recognition. online recognition refers to letting the computer recognize the characters written on the handwriting device, and recognize it on the basis of handwriting strokes and stroke order. The theoretical research of this technology is very mature. Offline recognition is the recognition of characters written on paper by the computer. There is no stroke and stroke order as the basis. Only the picture recognition provides less information, so it is still facing greater challenges. The system designed in this paper is a handwritten digit recognition system that can realize offline recognition. The flow chart of the handwritten digit recognition system based on convolutional neural network designed in this paper is shown in Figure 1.

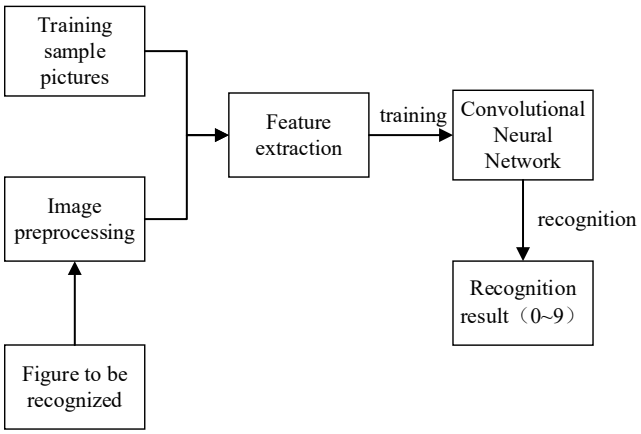

Fig. 1. The flow chart of the handwritten digit recognition system

The identification system in this paper can be divided into two modules, a data source module and a digital identification module. The data source module includes the provision of original handwritten digits and the feature extraction of handwritten digit pictures. Digit recognition module includes convolutional neural network (CNN) and recognition.

II. DATA SOURCE MODULE

A. MNIST Dataset

The MNIST dataset comes from the National Institute of Standards and Technology (NIST). The training set consists of numbers handwritten by 250 different people. 50% of them are high school students and 50% are from the Census Bureau staff. The test set is also the same proportion of handwritten digits. The MINST library contains two parts, SD1 and SD3, from which 30000 words are taken as the training set. The test set has a total of 10,000 test samples, and SD1 and SD3 each have 5000.

All digit pictures in the MINST dataset are standardized by the central composite size, and the size of each picture is 28 *28 pixels. This system completes the process of training sample pictures based on the MNIST data set. This system completes the process of training sample pictures based on the MNIST data set. The handwritten digit recognition system extracts the features of the handwritten digit pictures in the MINST data set, performs recognition training, and then inputs the handwritten digits to be recognized for identification.

B. Opencv Toolkit

Opencv (Open Source Computer Vision Library) is an open source library that is allowed in the case of BSD and covers many computer vision algorithms. The system in this paper uses the Opencv toolkit to perform image preprocessing and feature extraction on the digits to be recognized. Due to the uneven size of the handwritten pictures, we need to use the Opencv toolkit to compress the pictures into 28 * 28 pixels of the same size as the MNIST dataset. In addition, in terms of image processing, Opencv can achieve linear and non-linear image filtering, graphics changes (including setting size, radiation and perspective, using basic table settings mapping), color space changes, histograms, etc. [5].

This system uses Opencv toolkit to filter grayscale pictures of handwritten digits, do grayscale histogram, threshold extraction and binarization, and finally extract the outline features of handwritten digital pictures.

III. DIGITAL IDENTIFICATION MODULE

A. Deep Learning

In 2006, Hinton et al. [6] introduced the concept of deep learning for the first time in the paper. Deep learning can have good adaptability in the local minimum of the training process. Deep learning networks are developed on the basis of artificial neural networks (ANN). Artificial neural networks which are composed of many perceptrons are also called Muti-layer perceptron neural networks (MLP). The perceptron model is shown in Figure 2.

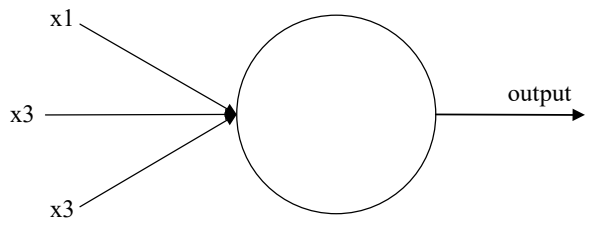

Fig. 2. The perceptron model

The output in Figure 2 is the formula (1):

$$output = \begin{cases} 0 & if \sum_j \omega_j x_j \le threshold \\ 1 & if \sum_j \omega_j x_j \ge threshold \end{cases} \quad (1)$$

The perceptron was invented by Frank Posenblatt [7] in the 1960s. Its working mechanism is to receive several binary inputs x_1, x_2, ... x_n and produce a binary output. ω_j is called the weight, which represents the real number of the importance of the input to the output. Formula (1) indicates whether the output result is 0 or 1 depending on whether the sum $\sum_j \omega_j x_j$ after weight distribution is less than the threshold or greater than the threshold.

The structure diagram of the artificial neural network composed of multiple perceptrons is shown in Figure 3:

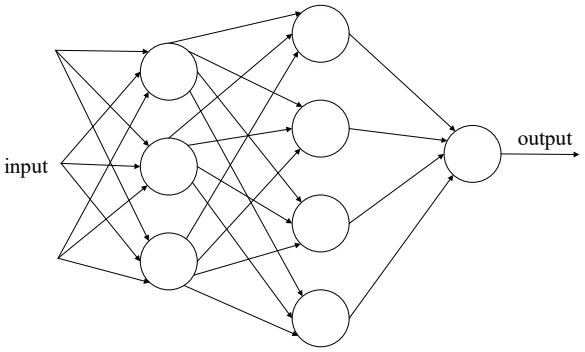

Fig. 3. Structure diagram of artificial neural network

The artificial neural network was inspired by the biological neural network, and it simplifies the biological neural network. Artificial neural network is optimized through a learning method based on mathematical statistics, so artificial neural network is also a practical application of mathematical statistics. Artificial neural networks can have simple deciding and judging abilities like human thinking, and have many characteristics of biological systems, such as robustness, high parallelism, nonlinearity, fault tolerance, and good learning ability, etc. This method is more reliable than logical reasoning in many scenarios.

Deep learning networks are composed of multiple artificial neural networks. The deep learning network consists of three layers, namely the input layer, the hidden layer and the output layer. The structure diagram of deep learning network is shown in Figure 4.

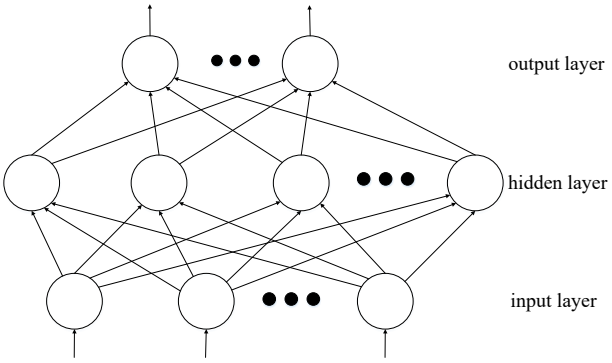

Fig. 4. Deep learning network structure diagram

B. Principle of Convolutional Neural Network

1) Convolutional neural network structure

The structure of the convolutional neural network is shown in Figure 5:

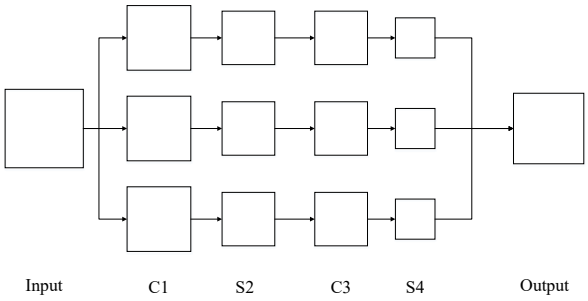

Fig. 5. The structure of the convolutional neural network

It can be seen from Figure 5 that the convolutional neural network is mainly composed of the convolutional layer C and the pooling layer S. The image is input from the input layer, and the convolutional layer C1 is obtained through 3 trainable digital filters. C1 also has 3 feature maps. Each feature map represents a set of learned features. The convolutional layer C1 is pooled to obtain the pooled layer S2. S2 is re-convolution filtered to obtain C3, and C3 is then pooled to obtain S4. After S4 is obtained, the features can be drawn into a straight line to act on the neural network.

2) Convolutional layer

The flow chart of the convolution layer operation is shown in Figure 6:

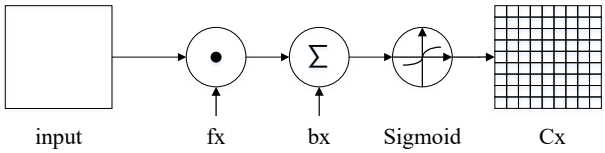

Fig. 6. The flow chart of the convolution layer

In Fig. 6, fx is a digital filter and bx is a bias voltage. Since we need to process two-dimensional images written with numbers, we need to use the convolution formula in two-dimensional:

$$s(i,j)=(X*W)(i,j)=\sum_{m}\sum_{n}x(i+m,j+n)\omega(m,n) \tag{2}$$

It can be seen from equation (2) that the elements in the final matrix are obtained by multiplying the matrices of different parts of the input image and the elements of the convolution kernel matrix, and then adding them. The examples is shown in Figure 7:

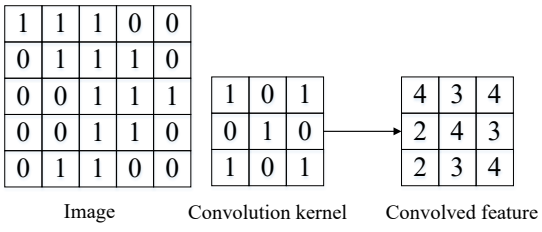

Fig. 7. Image convolution example

As shown in Figure 7, the 5 * 5 matrix becomes a new 3 * 3 matrix after the image convolution operation. It can be seen from Figure 6 that the image needs to pass through the Sigmoid function after passing through the filter and the bias voltage. The Sigmoid function image is shown in Figure 8.

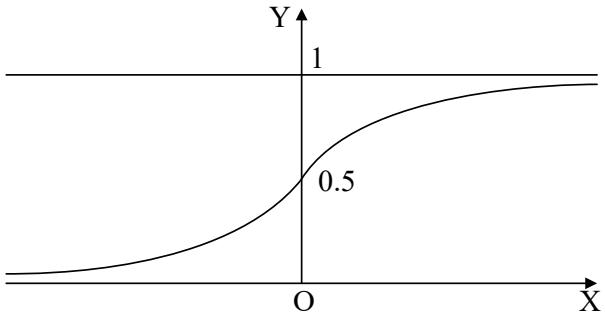

Fig. 8. Sigmoid function image

It can be seen from Figure 8 that the Sigmoid function is a continuous function at (0, 1). Its function is to make the image have displacement invariance in the process of filtering and applying bias voltage. As long as the size of the input layer and the size of the motion sensing area are determined, the size of the sensing layer C can be determined.

3) Pooling layer

Pooling operations are relatively simple compared to convolution operations. The essence of the pooling operation is to compress each sub-matrix in the input tensor, turning every n * n elements of the input sub-matrix into an element. There are usually two compression standards, taking the maximum value of the sub-matrix or the average value of the sub-matrix. In order to obtain higher accuracy in image processing, this paper uses the average compression method. Examples are shown in Figure 9.

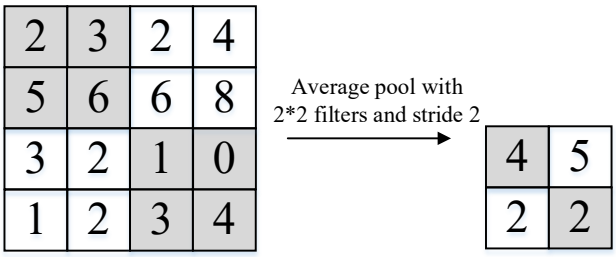

Fig. 9. Examples of pooling operations

The role of the pooling operation in the convolutional neural network is to map the weighted average of 4 consecutive pixels (2 * 2 area) in the convolutional layer C to the point of S, which reduces the dimension of the matrix.

C. Convolutional Neural Network Based on LeNet-5

The convolutional neural network used in this system is based on the LeNet-5 [8-9] model. The model is a convolutional neural network designed by Yann LeCun in 1998 for handwritten digit recognition. At that time, most US banks used it to recognize handwritten digits on checks. It is one of the most representative experimental systems in early convolutional neural networks.

The model has 5 layers, as shown in Figure 10.

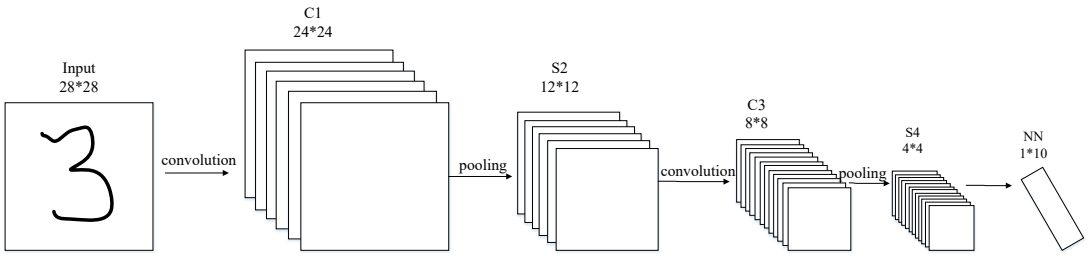

Fig. 10. Convolutional neural network model based on LeNet-5

It can be seen from Figure 10 that LeNet-5 consists of 2 convolutional layers, 2 pooling layers, and 1 fully connected layer. The processes of C1, S2, C3, and S4 are almost the same as those in Figure 5. The number of feature maps in the pooling operation is constant, and the number of feature maps in the convolution operation is independent of the previous layer. The C1 layer obtains 6 feature maps through 6 trainable digital filters, and the 12 feature maps of the C3 layer are all superimposed after being mapped from the S2 feature map. There are 12 feature maps in the C3 layer in this system. The process from S2 to C3 is a multi-to-one mapping. The role of the fully connected layer is to pull S4 into a 1 * 10 one-dimensional vector.

D. Classification of Recognition Results

The results to be identified are the ten discrete values 0-9, the probability sum of their results is 1 and it belongs to the multi-classification problem. Therefore, this system adopts Softmax regression model. Therefore, the Softmax regression model used in this system is a generalization of the logistic regression model on multi-classification problems. The regression result is to add a layer of function mapping on the continuous value, which has the characteristics of linear superposition, and then use the G (z) mapping, the value of the continuous discrete value mapping is 0-1.

Softmax regression model can better solve the multi-classification problem. In the Softmax regression model, the class label y can take different K values. The probability of category i in Softmax regression is:

$$s_i = \frac{e^{v_i}}{\sum_j e^{v_j}} \qquad (3)$$

The function of Softmax in this system is to map the probability of 0-9 numbers to be output to the value of (0, 1) interval, and the sum of these values is 1. When the output node is selected in the output layer, the point with the highest probability is selected as the output node, which is regarded as the prediction target.

IV. CONCLUSION

This paper designs a handwritten digit recognition system based on convolutional neural network. The system adopts the method of deep learning and uses the MNIST data set as a training sample. After the training is completed, the handwritten digits in the picture can be recognized through the Softmax regression model. The training and recognition process of this system is completed by LeNet-5 based convolutional neural network repeated convolution operation and pooling operation. Handwriting recognition is of great significance. Handwriting recognition is a bridge between handwriting and machines, and can play a huge role in finance, accounting, education and other fields, greatly reducing labor costs. The recognition results given by the existing recognition technology often still need manual review, and the accuracy needs to be improved. The design in this article also has the problem of too long recognition time. How to improve the recognition accuracy and reduce the recognition delay is still an important issue today.

REFERENCES

[1] O. Agaton, S. Kustrin, R. Beresford, “Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research”, Journal of biomedical analysis, vol. 22, no. 5, pp. 717-727, 2000.

[2] S. Y. Zhang, Y. M. Zhao, J. L. Li, “SVM-based image classification algorithm and implementation. Computer engineering and application”, vol. 25, pp. 40-42+86, 2007.

[3] K. Ding, “Research on adaptive technology of Chinese handwriting writers based on incremental learning”, South China University of Technology, 2011

[4] P. Zhou, “Research on Acoustic Modeling of Speech Recognition Based on Deep Neural Network”, University of Science and Technology of China, 2014

[5] Q. Y. Bu, “Design and Implementation of Face Recognition System Based on Open CV”, Jilin University, 2015.

[6] G. E. Hiton, S. Osindero, Y. Teh, “A fast learning algorithm for deep belief nets”, Neural computation, vol. 18, no. 7, pp. 1527-1554, 2006.

[7] S. Y. Zhang, “Depth model and its application in visual text analysis”, South China university of Technology, 2016

[8] L. Bottou, “Notes on convolutional neural networks”, 2006.

[9] W. Yuan, “Research on key words extraction and retrieval technology in online video subtitles”, North China University of Technology, 2017.

[10] G. H. Rosa, J. P. Papa (2019). Soft-Tempering Deep Belief Networks Parameters Through Genetic Programming. Journal of Artificial Intelligence and Systems, 1, 43–59.