International Journal of Social Science and Education Research Volume 5 Issue 5, 2022
ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

Handwritten Chinese Character Recognition Based on Deep Learning

Yuzhen Bai[1, a]

[1]School of Yunnan Normal University, Yunnan, 650500, China

[a]1480143416@qq.com

Abstract

The research of handwritten Chinese character recognition method has been the topic which the scholars in the related domain endeavor to explore. Chinese character recognition can be divided into online and offline handwritten Chinese character recognition. Deep Learning is a general term for a class of pattern analysis methods. It is a branch algorithm in the field of ML (Machine Learning) . The ultimate goal of deep learning is to make machines capable of human-like analytical learning, recognizing speech, image, text and other data. This paper analyzes a new method of handwritten Chinese character recognition based on deep learning. The method uses convolutional neural network to build a deep learning model and uses the model to recognize Chinese handwriting in images. Through this system, the user can recognize the complete handwritten Chinese character information, and has a high recognition rate for similar characters.

Keywords

Handwritten Chinese character recognition; Deep Learning; machine learning; convolutional neural network.

1. Introduction

Deep learning is about letting computers solve problems that seem intuitive but are difficult to describe in specific languages or mathematical rules, and letting them emulate human cognitive processes, learn from experience, and make computers human, the world is viewed and understood through a hierarchical and structured conceptual system, and each concept is defined in terms of its relationship to some relatively simple concepts (building simple concepts to learn complex concepts) . If we draw a graph showing how these concepts are created on top of each other, we get a “Deep”(multi-level) graph.

With the continuous progress of human life, human-computer interaction has become one of the most important technologies for processing various daily information. Moreover, in today’s society, the existing technology of deep learning for us to deal with the human-computer interaction vividly. It can automatically mine and process potential connections of data, thus avoiding the shortcomings of traditional methods. The application of handwritten Chinese character recognition technology in collecting handwritten note information, collecting examination paper content, collecting handwritten ancient books and documents can improve people’s work efficiency in these fields. Chinese characters record the growth of Chinese civilization, which brings with it the chain effects and various problems of recognition, making the subject one of the hot issues in the field of human-computer interaction, it has also been studied in depth by a large number of scholars. Therefore, in the field of handwritten Chinese character recognition, the technology of human-computer interaction deep learning has inherent advantages.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

According to Different Data Collection Methods, handwritten Chinese character recognition can be divided into two categories: on-line recognition and off-line recognition. Handwritten Chinese characters processed by on-line handwritten Chinese character recognition are text signals obtained by authors by Writing Online on physical devices such as digital pens, digital tablets or touch screens, and real-time input the writing path into the computer by sampling. Offline handwritten text recognition, however, is a two-dimensional image of handwritten text collected by an image capture device, such as a scanner or camera. Due to the different objects, the two handwritten text recognition technologies adopt different ways and strategies. The former is to identify a series of sampling point information according to the time sequence, while the latter is two-dimensional pixel information lacking stroke sequence information. Because there is no stroke sequence information, and because the photographic and scanning equipment in different lighting, resolution, writing paper and other conditions, digitization will bring some noise interference. The following figure shows the classification diagram of Chinese character recognition and the Flowchart of off-line handwritten Chinese characters:

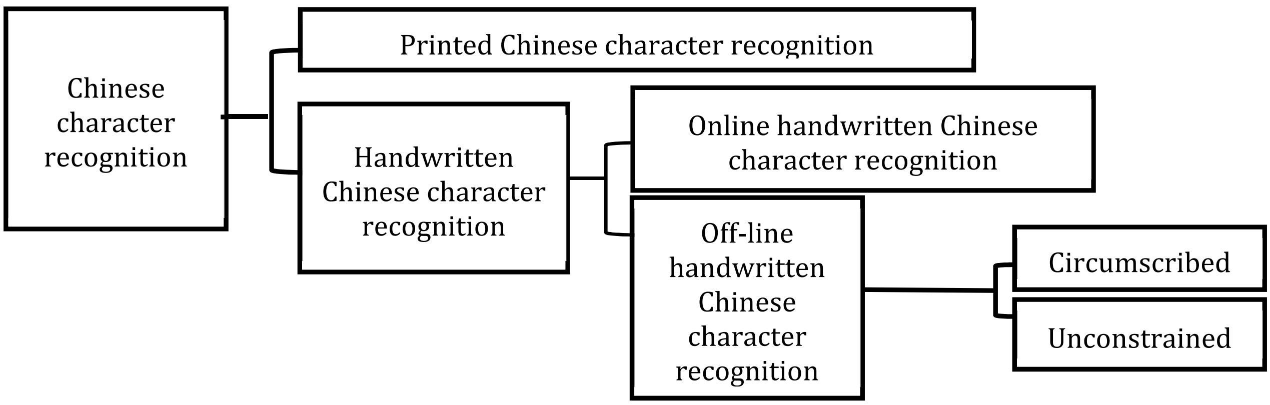

Figure 1. The classification of Chinese character recognition

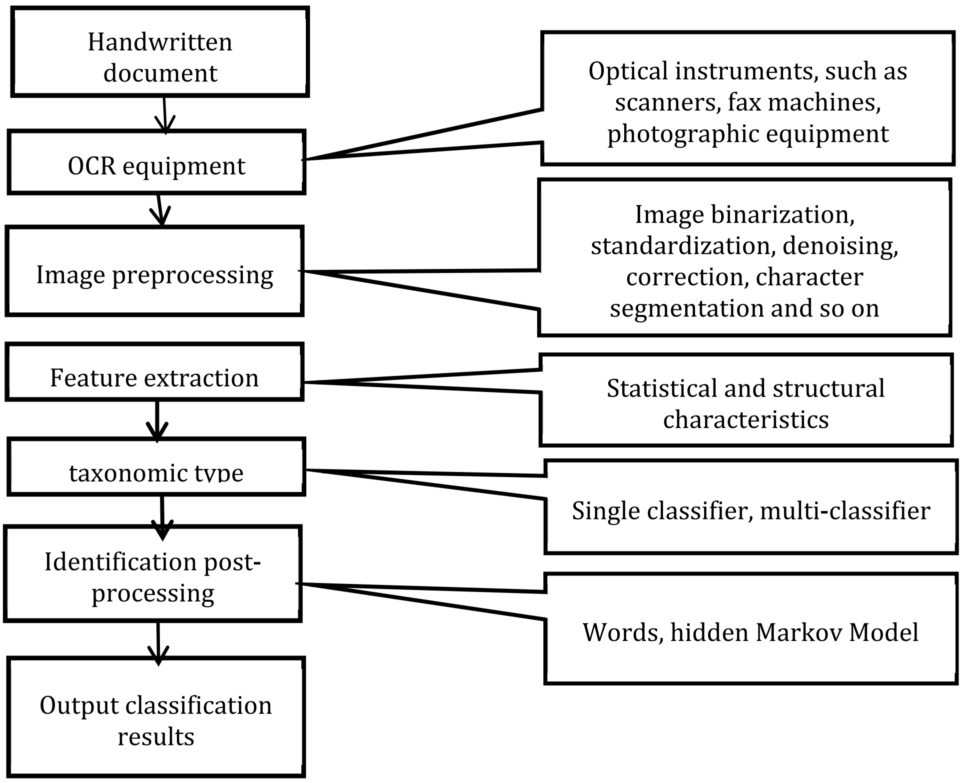

Figure 2. Flow chart of off-line handwritten Chinese character recognition

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

With the development of modern information society and the growing strength of China, the use of Chinese characters around the world and the development of artificial intelligence have a comprehensive and profound impact, in particular, the degree to which intelligent computers recognize various forms of Chinese characters. Human and computer communication is mainly through the keyboard and mouse, especially for the text input, generally through the keyboard to carry out. However, there are no Chinese characters on the keyboard. Although it is now possible to input Chinese characters by means of input codes such as phonetic codes and shape codes, the Chinese characters themselves are relatively complex and similar in shape and pronunciation, there are more or less problems in using the keyboard to input Chinese characters. Second, many documents and receipts now have to be written down on paper to be collected and stored on a computer. The recognition and input of Chinese handwriting by computer can improve the efficiency of human practical activities. In a word, the research of handwritten Chinese character recognition technology is of great significance to the development of China and even human society

2. Deep Learning

2.1. The Main Models of Deep Learning

2.1.1. Convolutional Neural Network,(CNN)

The idea is to use convolution, a special kind of linear operation, to replace at least one layer of the general matrix in a network to process data that has a similar grid structure (for example, an image can be viewed as a two-dimensional pixel grid) . The idea is to use convolution, a special kind of linear operation, to replace a General Matrix in a network. The idea is to use convolution to process data that has a similar grid structure (for example, an image can be viewed as a two-dimensional Pixel Grid) . Convolutional neural network is a key example of the successful application of deep understanding of brain research to machine learning applications, and the first neural network to address important business applications, excellent performance in a wide range of applications. Business interest in deep learning begins with Krizhevsky et Al. . Won The ImageNet contest. But before that, convolutional networks had also been used to win other machine learning and computer vision contests, though these had had little impact in previous years.

CLASSIC MODEL:

(1) Lenet (LeNet-5) by Yann Lecun

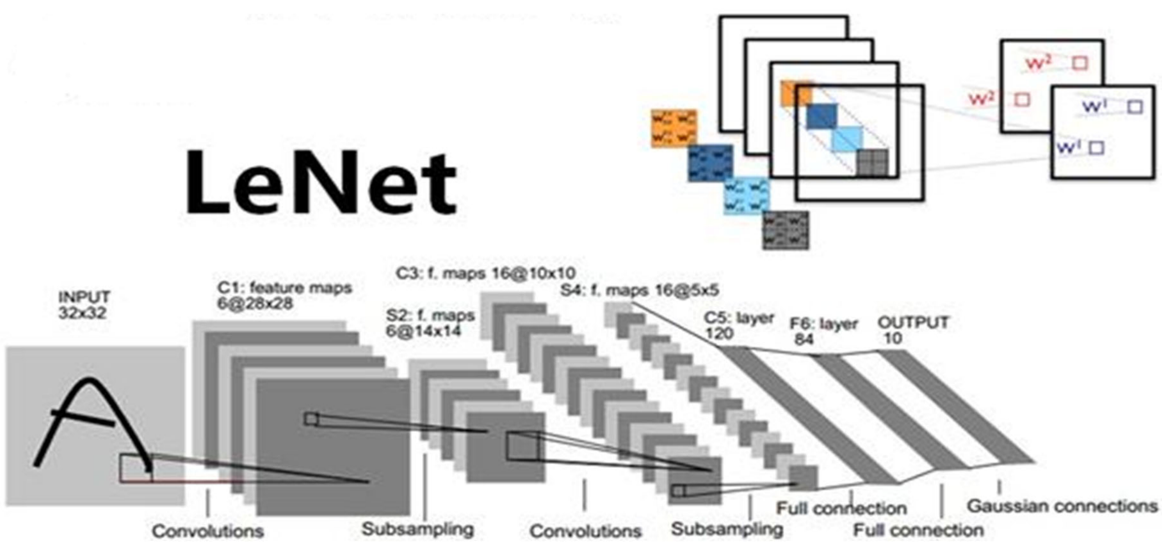

Figure 3. LeNet model

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

LeNet-5 is the basic part of another deep learning model, which does not include input. Its structure has seven layers, each layer has trainable parameters, each layer contains several feature maps, each feature map acquires a feature through a convolution filter, and each feature map has more than one neuron.

(2) Alex Krizhevsky's Alexnet

Alex Krizhevsky introduced AlexNet in 2012, winning the ILSVRC contest that year, with a 16.4% error rate in top5 predictions, largely surpassing first place. The basic structure of the network consists of eight layers of neural networks, five convolutional layers and three fully connected layers (three convolutional layers followed by a maximum cisterification layer) , including 630 million connections, 60 million parameters and 650,000 neurons.

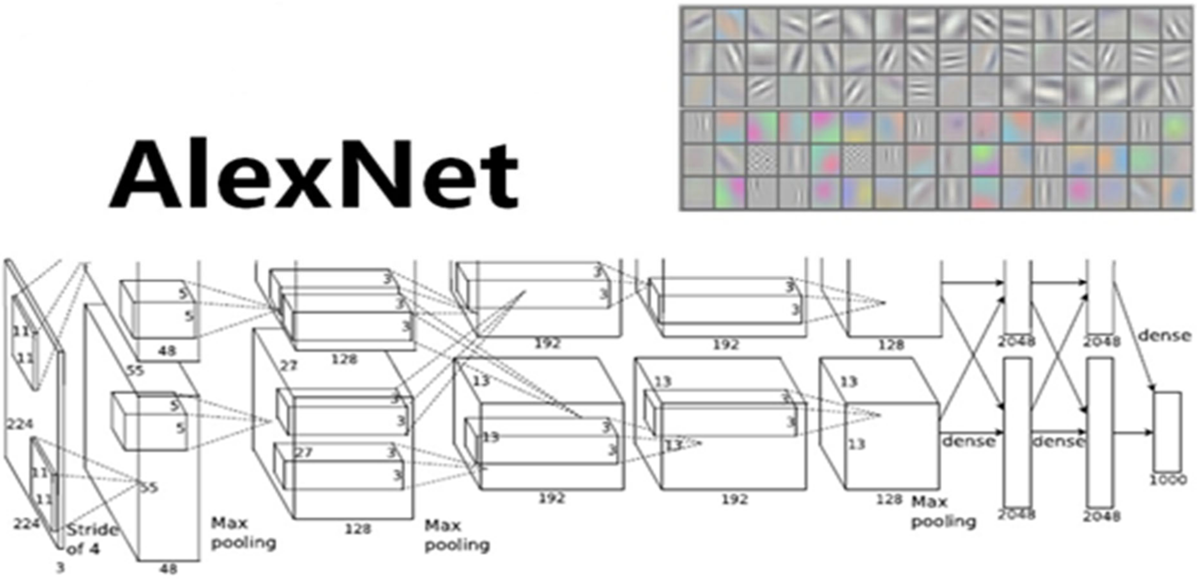

Figure 4. Alexnet structure model

(3) Developed by the Visual Geometry Group at Oxford, VGG (think of it as a souped-up version of AlexNet)

VGG Is structured as 5-layer convolution layer, 3-layer full connection layer and softmax output layer, which are separated by max-pooling.

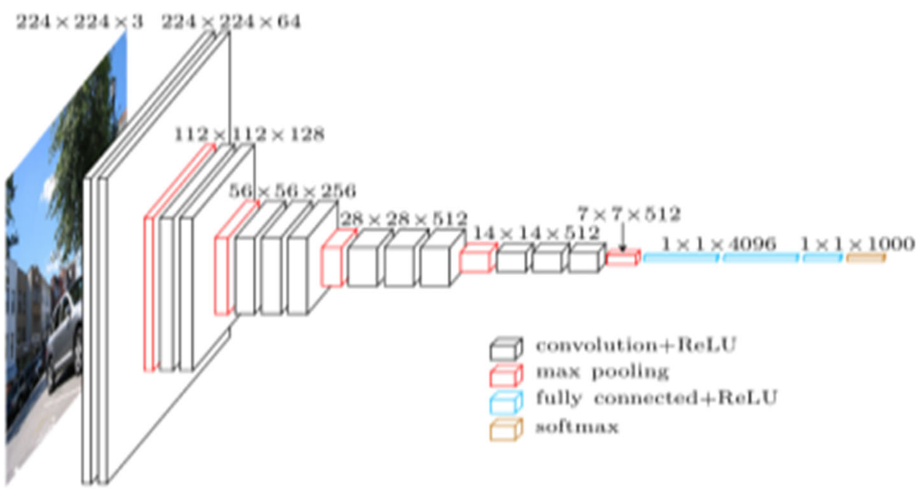

Figure 5. VGG model

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

2.1.2. Recurrent Neural Network (RNN) and Recurrent Neural Network

Recurrent Neural Network (CNN) is a kind of neural network which is used to solve sequence data in the field of neural network. It can be extended to relatively long sequences (much longer than non-sequence-based private networks), and most recurrent neural network can change the length of the sequence. The recurrent neural network can be expressed as an extension of the recurrent neural network and constructed as a deep tree structure instead of a chain structure of the recurrent neural network. This is a different type of calculation diagram. When each parent node of the recurrent neural network is connected only to a child, the structure is the same as a fully connected recurrent neural network. The potential use of recurrent neural network is in learning reasoning, which has been successfully applied to neural networks with data structures as inputs, such as natural language processing and computer vision.

2.1.3. Autoencoder

BASIC STRUCTURE OF AUTOMATIC ENCODER:

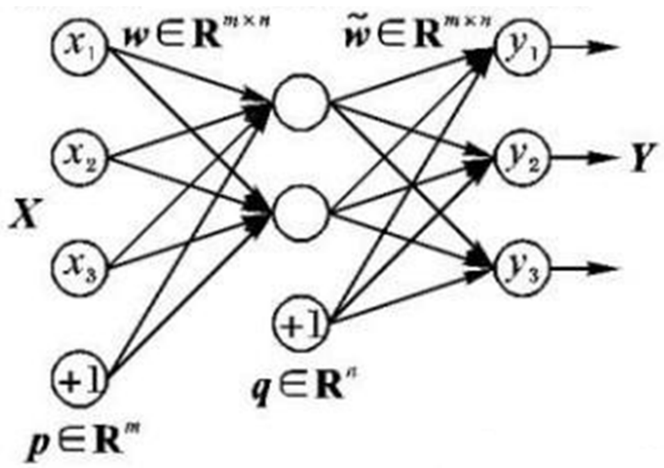

Figure 6. Automatic coding structure diagram

Autoencoder is a kind of neural network. It is trained to copy the input to the output. The output of a hidden layer h inside the encoder can be represented by the generated code. In general, we do not design the self-encoder so that the input and output are equal, but impose certain conditions to constrain the self-encoder to copy similar parts, and the input it replicates must be similar to the training data.

2.2. Deep Learning Training Process

2.2.1. Bottom-up Unsupervised Learning

This step is the most different part from the traditional neural network, can be regarded as a feature learning process, is not supervised. This layer can be regarded as a hidden layer of the three-layer neural network, which minimizes the difference between the output and the input. Because of the limited capacity and sparse constraints of the model, the model can learn the structure of the data itself, so that after learning the n-l layer, it can get more performance than the input. The output of the n-l layer is taken as the input of the n layer, train n levels, then get the parameters for each level.

2.2.2. Top-down Supervised Learning

It trains data with tags, transmits errors from top to bottom, and makes minor adjustments to the network. On the basis of the parameters obtained at the beginning, the parameters of the whole multi-layer model are further optimized. This step is a supervised training process. The first step is similar to the initial value process of the neural network. Because the first step is

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

not random initialization, but based on learning the structure of the input data, the original value is closer to the global best, then you can achieve a more complete effect. Therefore, to achieve a better effect of deep learning, to a large extent depends on the first step of the feature learning process.

2.3. Deep Learning Algorithms

If the L system has n layers, the first layer is L1, the next layer is L2, push down in this order until the n layer is LN. I and O represent input and output, respectively. Under the condition that I and I are equal, the information of system L is complete during the whole process of input l, which means that there is another representation of input I in any layer of system L, then we will automatically get the various levels of features corresponding to input I. In short, deep learning is the use and reference of machine learning algorithms to achieve hierarchical representation of input information.

3. Data Sets and Models

In this section, we will focus on the preprocessing of the data set used in this experiment, as well as the details of the convolutional neural network.

3.1. Preprocessing of Data Sets

Since the HWDB data is collected by hand-written electronic version, the image generated after the collection is a grayscale image, the value range of each pixel is between 0 ~ 255, and the strokes in Chinese characters can only reflect the writer's own writing style, for the recognition of Chinese characters itself is not of special significance. Considering these objective factors, I used the basic method of image processing to binary the original data set. After the binary, we can clearly see that the character of the font is not lost because of the binary, instead, it becomes clearer where things are less clear. And since the pixel value of each pixel in the binary image is composed of 0 or 1, using the binary image to store the Dataset also reduces our storage overhead on the Dataset, after preprocessing, the physical storage size of the image is reduced about 8 times. The raw data versus the binary data is illustrated in figure 7 below.

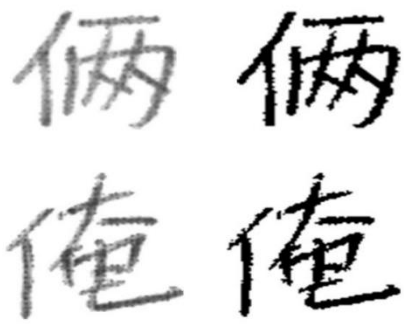

Figure 7. After binarization compared with the original data

As can be seen from figure 7 above, the Binary Image is obviously higher in definition than the original image, but it is difficult to choose the threshold of binary image, a single threshold can not be applied to images in all data sets. Based on the above problems, after the image binarization, the edges of the image appear more obvious saw teeth. To solve this problem, I used gauss filtering to process the image before binary processing, the main principle of GAUSS filtering is to give different weights to the neighborhood of the Pixel when smoothing the image, the main advantage of this method is that it can smooth the image and retain the whole gray

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

distribution of the original image. The GAUSS filtered and binarized image is shown in figure 8 below.

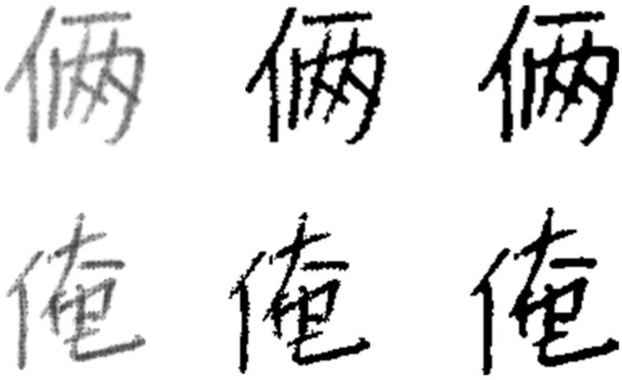

Figure 8. Through gaussian filtering, not through a gaussian filter, the original image contrast

In figure 8, the image on the far left is the original image, the image in the middle is only binary without gauss filtering, and the image on the left is both gauss filtering and binarization. It is clear from figure 8 that the overall features of the gauss filtered image have not changed much, but the edges have been significantly fused due to the binarization, this processing can avoid the sawtooth shape of the image to bring additional feature information to the image. After the above processing, our data set has met the expected requirements of this test.

3.2. Convolutional Neural Network

3.2.1. Foundation Model

The main technical principle of the experimental model is to use deep convolutional neural network to extract image features for recognition. Zhang Y [10] proposed eight neural networks with different depths and channel numbers, each consisting of 5 to 11 layers. After comprehensive consideration, my basic convolutional neural network is a six layer convolutional neural network, which lies in the middle of all eight networks, regardless of its depth or the number of characteristic channels, good balance between performance and system overhead during training. The 6-layer network includes 4 layers of convolution layer and 2 layers of full connection layer. A schematic diagram of its network structure is shown in figure 9.

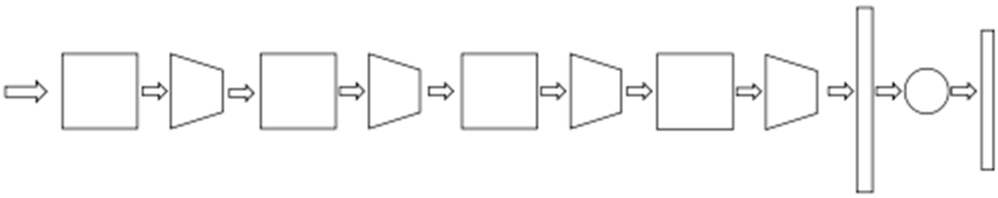

Figure 9. Original model structure diagram

I then reproduced the network structure according to the original literature, and trained it using the pre-processed images in the previous section, but the recognition rate of the network could not reach the effect of the literature, based on this, I've made improvements to the M6 network.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

3.2.2. Model Improvements

Based on the previous section I reproduced the selected model and found that the model has no ability to learn the features of the image, then I change the loss function of this model from crossentropy loss to Negative Log Likelihood loss and change the last classifier of neural network from softmax to softlog-softmax. After the above changes in the original model has been able to start learning the features of the image, but its classification accuracy is not too good based on this I made a major improvement to the neural network. Because the original model is only a simple linear model, so its fitting degree to the features is not enough, so the performance of the test set is not satisfactory. So my main idea in the design of the improved model is to use two parallel linear convolutional neural network with different convolution kernels in convolutional neural network to ensure that the extracted features are not the same, then the feature map of the two convolutional neural network is added together, and the final layer is identified by the fully connected layer, which is used to increase the model fitting ability, compared with the original model, the proposed method has a great improvement in recognition ability. The new channel reduces the size of the receptive field compared with the original model to prevent the loss of smaller details when the receptive field is too large. The improved model structure Diagram is shown in figure 10 below.

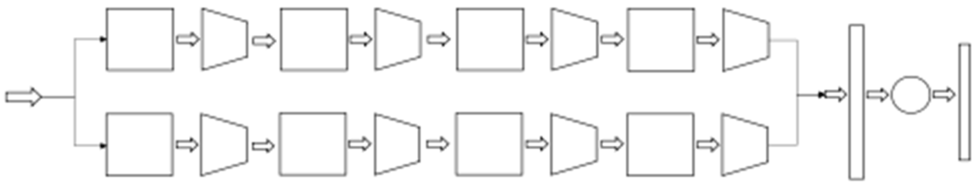

Figure 10. The improved model structure diagram

From the above diagram, we can see that the improved model structure has two-tier structure. The structural parameters of the two-tier structure will be shown in Table 1 below.

Table 1. Model structure parameters

Layers	Conv1	Conv2	Conv3	Conv4
Top	64@11@5	128@7@3	256@5@2	512@3@1
Bottom	64@3@1	128@3@1	256@3@1	512@3@1

As shown in the table above, Top is the Top layer of the convolutional neural network, Bottom is the Bottom layer, which is my new layer, and Conv is the convolution layer, the second number represents the convolution kernel of the current convolution layer, and the last number represents the number of lines zero-adding to the feature graph before convolution is performed on the convolution layer. The size of the pooled cores in the whole upgrade network is 2. And the nonlinear activation layer (ReLU) is linked between each convolution layer and pool layer. The main function of Relu is to express the feature in non-linear way, which can express the more complex data feature.

3.2.3. Hyperchoice

The authors set the learning rate as 0.01 in the original text, adjusted the learning rate to half of the previous round after 3 rounds of training and set the L2 regularization and dropout.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

However, according to the results of replication, this learning rate is relatively small, so that the update results of the model is too slow. I then adjusted my learning rate and updated my strategy by scaling the learning rate up 10 to 0.1 and discarding the L2 regularization, instead of dropping 50% of the neurons randomly in the dropout between the first and the next full-connected layer, 80% of the neurons were dropped to prevent overfitting later in the model, the idea is that when data is passed to the second fully connected layer, the number of fully connected neurons in the second layer is randomly hidden away by 80% and the remaining 20% of the neurons are used until the end of the training cycle, because the hidden neurons are random, each can be trained after several rounds of training. Finally, the learning rate was reduced by 0.8 after every 5 rounds of training. I used Stochastic gradient descent in the optimizer selection.

4. Test Setup and Result

4.1. Test Setup

The data set I used in this experiment is the online Chinese handwritten Chinese character database (HWDB) [9] provided by the Institute of Automation, Chinese Academy of Sciences. All handwritten Chinese characters are written using an electronic tablet, which contains 7,185 characters and 171 symbols, but is limited by my own computing platform, so in this experiment, I took a random sample of 200 Chinese characters, each of which was written by about 300 different writers, for a total of 59,703 pictures of Chinese characters. Using 80% of the training set and 20% of the test set, I got a total of 47,768 images in the training set and 11,935 images in the test set.

300 training sessions per trial. In the whole experiment design, three experiments were carried out, first on the original model with the data set I selected, then on the improved model with the same data set, in order to prevent the accidental deviation caused by the single test, the improved model is used to conduct the second test under the same conditions.

4.2. Test Results

First, I replicated the model in the original paper and ran the experiment, which involved 300 rounds of training. The results of the training set and the test set are shown in figure 11 below.

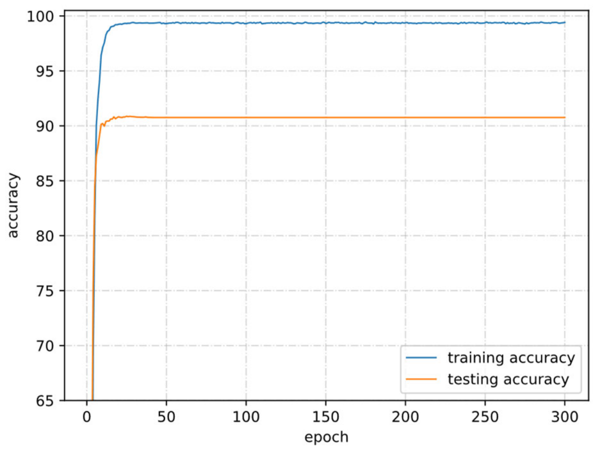

Figure 11. Original model

According to the diagram above, we can see that after 20 rounds of training in the original model, the accuracy rate of the model has converged to less than 91% , at the same time, it is worth noting that in this experiment, except for the structure of the model itself, the parameters

ISSN: 2637-6067
DOI: 10.6918/IJOSSER.202205_5(5).0034

of the model have been readjusted, and if the model is set according to the original parameters, the recognition ability of the model to the image features is not very good. The first experiment was carried out on the same data set using the improved model, and the parameters of this experiment were all described above. The results of the same 300 rounds of training are shown in figure 12 below.

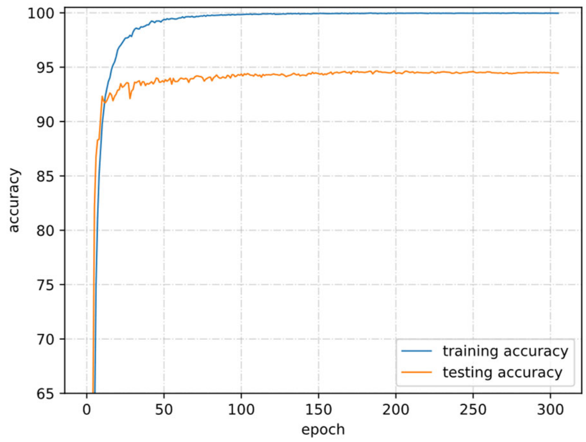

Figure 12. The improved model of training for the first time

As can be seen from the diagram above, the accuracy of the test set of my improved model after 150 rounds of training has almost converged to 95% , which is about 4% higher than that of the original model, although the convergence speed of the improved model is slower than that of the original model, the accuracy of the improved model is improved greatly, and the performance of the improved model is better than that of the original model. In order to eliminate the chance in the experiment, we use the same strategy to do the experiment again. The result of the experiment is shown in figure 13.

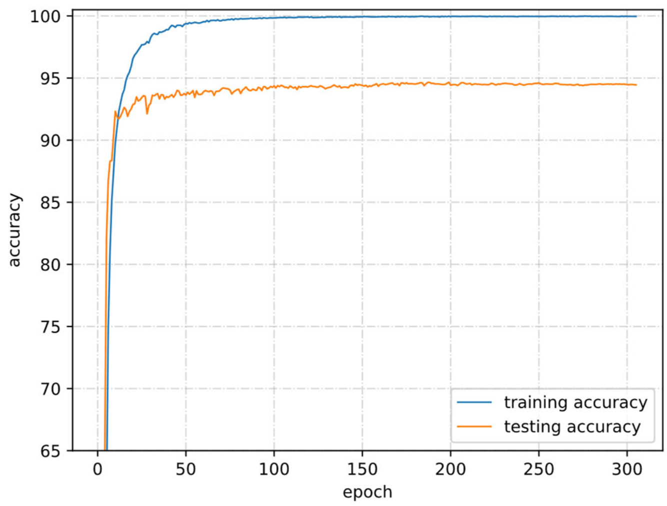

Figure 13. Improved model test for the second time

From figure 4-3, we can see that the trend of the accuracy of the training set and the test set is almost identical, which shows that the improved model has good performance in accuracy and stability.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0034

5. Conclusion

In this paper, I make the improvement and adjustment of the model structure based on the original model, and adjust the updating strategy of hyper-parameter. The experiment results show that the accuracy of the improved model is better than that of the original model. At the same time, there are some corresponding problems in the improved model, for example, the convergence speed decreases compared with the original model, which is mainly due to the more complex structure of the model and more parameters. The accuracy is better than the previous model, but I think it should be better.

Acknowledgments

This thesis was completed under the Gracious Care and guidance of my tutor. Her serious scientific attitude, rigorous academic spirit, excelsior style of work, deeply impressed and inspired me. From the selection of the project to the final completion of the project, she has always given me careful guidance and unremitting support. I would like to extend my sincere thanks and high esteem to her. At the same time, I would like to thank all the authors cited in the paper, their hard work and research results have given me great help and inspiration. Thanks to the experts who reviewed this article, their suggestions and insights added a great deal to my paper. Just a few words can not carry my incomparable gratitude, I will be these love and harvest permanent collection, bless you have a beautiful future and life!

References

[1] Zhao Jiyin, Zheng Ruirui, Wu Pao-chun, Li Min. Summary of off-line handwritten Chinese character recognition [j] . Journal of Electronics, 2010,38(02) : 405-415.

[2] Chen si-che. Deep Learning Algorithms and applications based on convolutional neural network. Technology Communication, 2017,9(18) : 59-60.

[3] Chen hsien-chang. Research on convolutional neural network based deep learning algorithms and applications. Zhejiang Gongshang University, 2014.

[4] Surya. A study of handwritten Chinese character recognition based on deep learning. Inner Mongolia University, 2019.

[5] Xu Yamei. Research on some key techniques of handwritten Uighur character recognition [d] . Xidian University, 2014.

[6] Gu Zhenye. Convolutional neural network for small-scale image classification. Huazhong University of Science and Technology, 2017.

[7] Wong kwok-kwan. Research on single target tracking algorithm based on deep learning in complex background. University of Science and Technology of China, 2018.

[8] Si Peng Ju, Hu Wei. Application of deep learning in biomimetic eye monitoring system. Measurement and control technology, 2016,35(12) : 139-143.

[9] Institute of Automation, Chinese Academy of Sciences Http://www.nlpr.ia.ac.cn/databases/handwriting/home.html,

[10] Zhang Y.Deep convolutional network for handwritten Chinese character recognition[J]. Computer Science Department,Stanford University,2015.

[11] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.