Journal of Robotics, Networking and Artificial Life
Vol. 6(1); June (2019), pp. 52–55
DOI: https://doi.org/10.2991/jrnal.k.190601.002; ISSN 2405-9021; eISSN 2352-6386
https://www.atlantis-press.com/journals/jrnal

Research Article

The Recognition and Implementation of Handwritten Character based on Deep Learning

Fengzhi Dai[1,*], Zhongyong Ye[1], Xia Jin[1,2]

[1]Department of Automation, College of Electronic Information and Automation, Tianjin University of Science and Technology, China
[2]Tianjin Tianke Intelligent and Manufacture Technology Co., Ltd., China

ARTICLE INFO

Article History
Received 23 October 2018
Accepted 20 November 2018

Keywords
Deep learning
machine learning
pattern recognition
handwritten character recognition
convolution neural network

ABSTRACT

This paper mainly focuses on the recognition of Chinese handwritten characters. By using the deep learning technology, a deep convolution neural network (CNN) is constructed to identify the handwritten character set, and finally gets the network structure that can be used for recognition. By comparing with other handwritten characters, the engineering application value of the network structure used in this paper is proved, and finally the handwritten character recognition system based on this model also embodies the feasibility of the network structure.

1. INTRODUCTION

In recent years, deep learning and deep network structure based on multiple hidden layers have been proposed, which are used to learn and deal with some popular problems in the field of machine learning, such as image retrieval and image recognition [1].

At present, although there are some applications of deep learning technology to solve the handwritten digital character recognition, for the handwritten Chinese character recognition, it is still the traditional method that is based on the artificial feature extraction [2]. Chinese characters are more complex than other common characters such as English letters or Arabic numerals, and even the same Chinese character has many written styles because of the different habits of personal writing. Therefore, the recognition of handwritten Chinese characters has always been a hot research problem in the field of machine learning.

2. APPLICATION OF DEEP CONVOLUTION NEURAL NETWORK IN HANDWRITTEN CHINESE CHARACTER RECOGNITION

2.1. Data Preparation and Preprocessing

The sampling data of the handwritten Chinese character set are selected from the HWDB1.1 database [3]. In the experiment, the sampling data of HWDB1.1 are divided into four sub-datasets, that is, Ten classification, Hundred classification, Thousand classification, 3755 classification of the dataset will be named as Set1, Set2, Set3 and Set4 respectively. The composition of each dataset is shown as follows in Table 1.

In Table 1, Set name represents the data name of four sub-datasets, Writer1 is the number of people who wrote the training samples. That is to say the sampling source of the training sample came from a different 240 persons. Writer2 represents the number of people who wrote the test sample. The Training column represents the total number of training samples in the dataset, while the Testing column indicates the total number of samples used for testing in the dataset. The last Total column represents the total number of training and test data samples.

2.2. Model Structure and its Improvement Method

The structure of the network proposed in this paper is based on the Alexnet deep Convolution Neural Network (CNN) and we have improved on its structure. It consists of five convolution layers, among them three of which are followed by a pooling layer with maximum sampling respectively, after that followed two full connection layers, and the last layer of classifier with 1000 output nodes. If the input layer is also included, the total number of layers of the network reaches nine layers, which is shown in Figure 1.

The innovation of this paper is the improvement of the Alexnet network [4]: (1) the number of convolution kernels in the convolutional layer is adjusted to reduce the number of convolutional layers. (2) Adjust the number of the whole connection layer and readjust

*Corresponding author. Email: daifz@tust.edu.cn

the network training parameters according to the performance of the computer. The explanation is listed as follows:

- Reduce the dimensions of the input layer from 3 to 1. For handwritten Chinese characters, the image can be changed to grayscale so as to get a single-channel picture. It need not retain the high-resolution Red, Green, Blue (RGB) three channels.
- Change the number of convolution kernels in the first convolution layer from the original 96 to 60, and the corresponding pooling layer is changed accordingly.
- Remove the first fully connected layer and connect the final output layer with only one full-connection layer.
- Change the number of output nodes of the final classifier to 10, 100, 1000 and 3755, respectively. And they are named as Set1, Set2, Set3 and Set4, respectively.

The structure through the above first three steps is named CNNet, and because its final output node is slightly different (Set1, Set2, Set3 and Set4), it is named CNNet1, CNNet2, CNNet3 and CNNet4 respectively.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. The Influence of Calculating Mean Image on the Result in Preprocessing

In this experiment, we preprocessed the image [5] mean value and verified the influence of preprocessing on network training. Under the condition that relevant network parameters remain unchanged, the network CNNet3 is obtained by improving Alexnet network and is trained from Set3, and verify the difference between training preprocessed by the mean value calculation and without mean value calculation, respectively. The result of the two training models are shown in Figure 2.

The horizontal coordinate in Figure 2 represents the number of iterations of the network, while the longitudinal coordinate represents the loss value of the network model.

Table 1 | Composition of the data sets

Set name	Writer1	Writer2	Training	Testing	Total
Set1	240	60	2400	600	3000
Set2	240	60	24000	6000	30000
Set3	240	60	240000	60000	300000
Set4	240	60	901200	225300	1126500

3.2. Setting the Training Parameter

Using the strategy of the relevant parameters in Le Net5, the learning rate is set from 0.01 to 0.02, which is equivalent to increasing the step length. The gradient descent method is used, and the network oscillation is avoided, thus it can cross the local minimum point and approach the extremum. Finally, it is proved by experiments that the value of Batch_size is reduced to 64, the network can converge at a faster speed, and the classification accuracy is higher, which proves that the network also achieves the optimal solution, so the paper sets the Batch_size to 64 and the learning rate to 0.01.

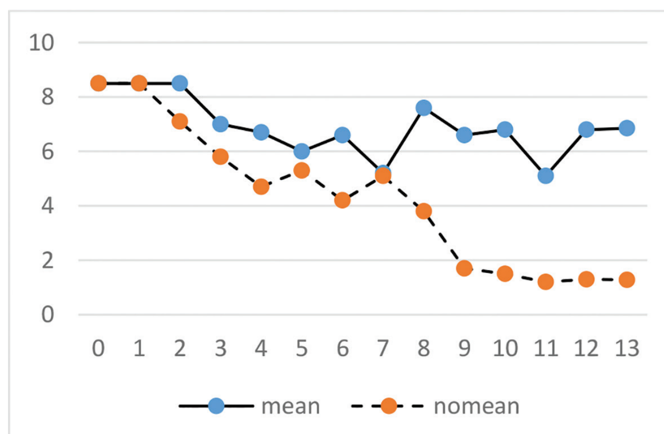

Figure 2 | Relationship between calculating mean value and network convergence in preprocessing.

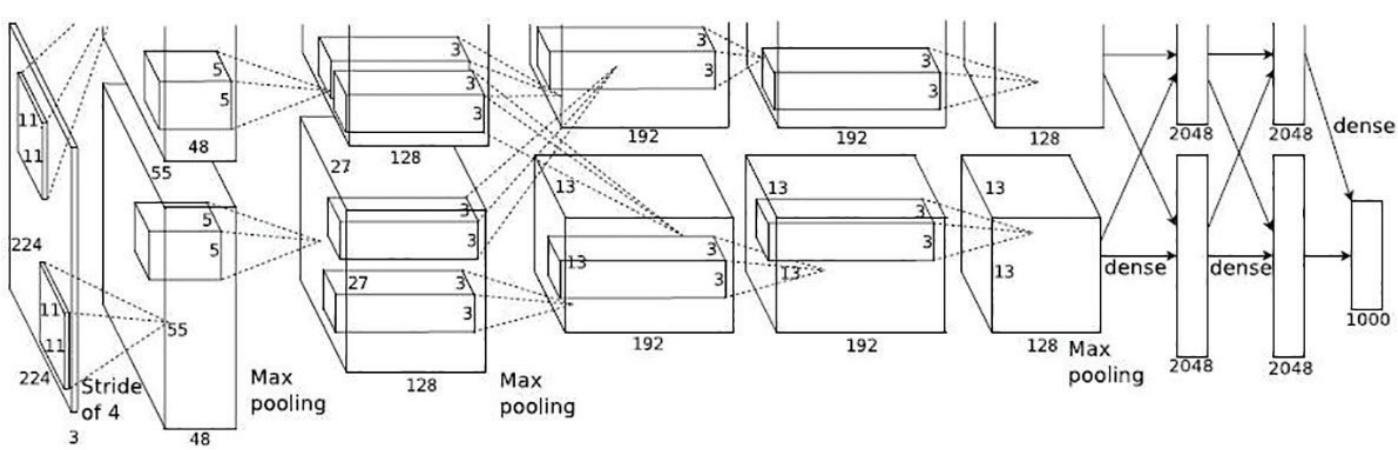

Figure 1 | Alexnet network structure diagram.

3.3. Experimental Results and Analysis of Four Data Sets and MNIST in CNNet

After adjusting the relevant parameters of the network so that the network can converge, the four sub-datasets Set1, Set2, Set3 and Set4 of HWDB1.1 handwritten characters are trained with four sub-networks of CNNet respectively. Sine CNN networks with higher complexity, the recognition rate of handwritten Chinese characters is higher. Also their ability to express the sampling features is much higher than that of the simple CNNet2, that is, more sample categories can be classified.

Although the convolution network with shallow accuracy has increased, the complexity of CNNet is higher, and the following Table 2 is the comparison between Net2 and CNNet networks in the number of convergence.

Under the same experimental conditions [6,7], CNNet is much higher than the Net2 network in the number of iterations required to make the network converge. Taking the training of DataSet Set4 as an example, the number of samples used for training has reached more than 900,000, and the size of each image is large, the Batch_size for each training weight needs to be used is also large. Coupled with the need for all the test data after each 10,000 iterations to verify the classification accuracy of the network, it takes nearly 4 days for the network CNNet to be fully trained on the training Set4, the time spent on each parameter is enormous. This is also a major disadvantage for deep CNNs compared with simple CNN.

4. SYSTEM DESIGN AND IMPLEMENTATION

4.1. System Structure and Design

4.1.1. System framework and related modules

The system adopts the BS structure as shown in Figure 3, that is, the user only needs to have simple interaction with the browser, and most of the core business of the system is on the server side. The system uses Python-based Tornado web framework.

According to the different division of functions, the system can be divided into the following modules: the input module, model information loading module, model classification module, result mapping module and the final output module, which is shown in Figure 4.

Table 2 | Comparison of convergence times between Net2 and CNNet

Set name	Net2	CNNet
Set1	6100	48000
Set2	11600	84000
Set3	14300	123000
Set4	—	179000

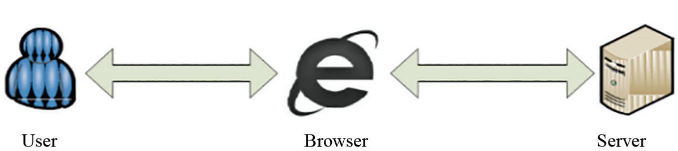

Figure 3 | BS architecture diagram.

4.1.2. Core class design

The main modules of this system are model loading, model classification and result mapping module. Each module corresponds to a core class, and all of them complete the main functions of the system [8].

4.2. Data-related Preprocessing Operation

The model used is CNNet3, the input is single channel grayscale images. If the color image of RGB is entered, it needs to be preprocessed to grayscale.

There are many ways to process the RGB image to grayscale, such as the component method, which uses the pixel value of one of the RGB channels as the grayscale value; Mean method, the average of three channels is used as the grayscale value; Weighted average method, that is, according to a certain weight to the RGB three component values to calculate the grayscale value. This paper uses the last method, and the formula is as follows:

$$f(x,y) = 0.30\mathrm{R}(x,y) + 0.59\mathrm{G}(x,y) + 0.11\mathrm{B}(x,y) \qquad (1)$$

The weight of G component is set to the highest 0.59, and the lowest value 0.11 is set to B component. The input of the system is a multi-formatted character picture, and the correct characters in the picture are also proposed. In the network, the initial result of the output is the corresponding character category in the training data, and the corresponding transformation is carried out in the mapping module. Therefore, the mapping between characters and categories must be established first, as shown in Figure 5.

The CNNet3 network is used and the output node is 1000. We need to establish the first 1000 classes of the DataSet. In Figure 5, each folder contains samples of the same character dataset, corresponding to different handwritten characters.

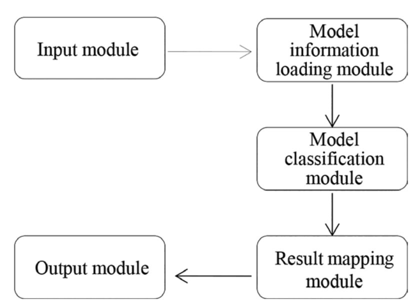

Figure 4 | Overall module flow chart of the system.

Figure 5 | Mapping relationships for datasets.

CONFLICTS OF INTEREST

There is no conflicts of interest.

ACKNOWLEDGMENTS

The research is partly supported by Project of Tianjin Enterprise Science and Technology Commissioner to Tianjin Tianke Intelligent and Manufacture Technology Co., Ltd. It is also supported by Industry-University Cooperation and Education Project (201802286009).

REFERENCES

[1] R. Zhang, W. Li, T. Mo, Review of deep learning, Inform. Control 5 (2018), 385–397.

[2] B. Sun, D. Kong, W. Zhang, W. Jia, Survey on human action recognition from depth maps, J. Beijing Univ. Technol. 10 (2018), 1353–1366.

[3] K. Yang, Y. Xu, X. An, Y. Li, K. Liu, Army Military Transportation University, Vehicle detection method based on deep learning, Comput. Netw. 7 (2018), 58–61.

[4] S. Zhou, Techniques review of deep learning, Inform. Technol. Exploration 10 (2018), 116–117.

[5] A. Liu, S. Zhang, X. Chang, B. Chen, Real time human action recognition and monitoring method based on deep learning, Heilongjiang Sci. 10 (2018), 14–16.

[6] Y. Huang, C. Wan, Human behavior recognition algorithm based on deep learning, Appl. Electron. Tech. 10 (2018), 1–10.

[7] Q-C. Mei, W. Lyu, Research on commodity image recognition based on deep learning, Mech. Electric. Eng. Technol. 9 (2018), 28–151.

[8] H. Zhao, Z. Wu, X. Wang, A multi-robot rescuing system. The 2018 International Conference on Artificial Life and Robotics (ICAROB 2018), 1–4 February 2018, Beppu, Japan, 23, pp. 582–586.

Authors Introduction

Dr. Fengzhi Dai

He received an M.E. and Doctor of Engineering (PhD) from the Beijing Institute of Technology, China in 1998 and Oita University, Japan in 2004 respectively. His main research interests are artificial intelligence, pattern recognition and robotics.

He worked in National Institute of Technology, Matsue College, Japan from 2003 to 2009. Since October 2009, he has been the staff in Tianjin University of Science and Technology, China, where he is currently an associate Professor of the College of Electronic Information and Automation.

Mr. Zhongyong Ye

He is a second-year master candidate in Tianjin University of Science and Technology, majoring in numerical analysis, matrix theory, Internet of things and other important control disciplines. During his study, he has published many conference papers and patents.

Ms. Xia Jin

She is a second-year master candidate in Tianjin University of Science and Technology. Her research area is about machine learning, image processing and Intelligent manufacturing.