The recognition and implementation of handwritten character based on deep learning

Zhongyong Ye[1], Fengzhi Dai[1*], Xia Jin[1], Yasheng Yuan[1], Lingran An[1], Yujie Yan[1], Yiqiao Qin[12], Hao Li[3]

[1] *College of Electronic Information and Automation,*
Tianjin University of Science and Technology, China;

[2] *Tianjin Tianke Intelligent and Manufacture Technology CO.,LTD, China;*

[3] *CETC Ocean Information Co., Ltd, Beijing 100041, China*
*E-mail: * daifz@tust.edu.cn*
www.tust.edu.cn

Abstract

This paper mainly focuses on the recognition of handwritten characters, especially handwritten Chinese characters. Using deep learning technology constructs a deep convolution neural network to identify the character set of handwritten Chinese characters, compares the performance differences of the same depth network, and finally gets the network structure which can be used for recognition, and realizes the recognition system of handwritten characters based on the network. By comparing with other hand writing characters, the engineering application value of the network structure used in this paper is proved, and finally the handwriting character recognition system based on this model also embodies the feasibility of the network structure in this paper.

Keywords: Deep learning, Machine learning, Pattern recognition, Handwriting character recognition, Convolution neural network

1. Introduction

Deep Learning is a deep-seated network structure based on multiple hidden layers proposed in recent years, which is used to study 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 more application of deep learning technology to solve the application of handwritten digital character recognition, but in the field of handwritten Chinese character recognition, it is more traditional method based on artificial feature extraction. Because Chinese character strokes are more complex than other common characters English letters or Arabic numerals, and handwritten Chinese characters are more varied because of the different styles and habits of personal writing, Therefore, the recognition of handwritten Chinese character 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

Because this paper is aimed at the recognition of handwritten Chinese characters, the sample data using handwritten Chinese characters set is also used from the HWDB1.1 database to obtain some of the data samples. In this experiment, the sample data HWDB1.1 divided into 4 sub-datasets, that is, Ten classification, Hundred classification, Thousand classification, 3755 classification of the dataset will be named Set1, Set2, Set3 and Set4 respectively. The composition of each dataset is as follows in table 1.

© The 2019 International Conference on Artificial Life and Robotics (ICAROB2019), Jan. 10-13, B-Con Plaza, Beppu, Oita, Japan

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

In table 1, set name represents the data name of four child datasets, Writer1 that the number of people writing training samples, that is, 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, Testing column indicates the total number of samples used for testing in the dataset, and the last column represents the total number of samples used to experiment with the dataset.

2.2. Model structure and its improvement method

The network used in the design of this paper is improved on the basis of Alexnet deep convolution neural network. It contains five convolution layers, 3 of which are followed by a pool layer with a maximum sampling, followed by 3 full-attached layers, and the last Oftma classifier with 1000 output nodes, and if the input layer is not included, the total number of layers of the network reaches 8 layers, the network structure is shown in Fig.1.

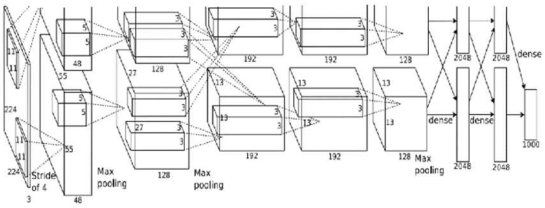

Fig.1. Alexnet Network structure diagram

In this paper, the main improvement of the Alexnet network is to adjust the number of convolution nuclei in the convolution layer, reduce the number of convolution layer, and adjust the number of the full connection layer, according to the performance of their own computer to readjust the network training parameters, the specific improvement is as follows:

1. Reduce the dimension of the input layer from 3 to 1, because for handwritten Chinese characters, the image can be grayscale to get a single-channel picture, it does not have the same as high-resolution natural images, in the RGB three channels have a wealth of information.

2. Change the number of convolution cores in the first convolution layer from the original 96 to 60, and the corresponding pooling layer changes accordingly.

3. Remove the first fully connected layer and connect the final output layer with only one full-connection layer.

4. Change the number of output nodes of the final classifier to 10,100,1000 and 3755, respectively, that is, the datasets used are Set1, Set2, Set3, and Set4, respectively. The structure obtained by the network through the above 1,2,3 step is named CNNet, and because its final output node is slightly different, 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, in order to verify whether the mean processing of images in preprocessing will have an impact on the training of the network, under the condition that the relevant parameters of the network are unchanged, the network CNNet3 obtained by the improvement of the Alexnet network are preprocessed by mean calculation Set3 Training and Training sample SeT3 without mean calculation preprocessing are trained, and the relevant results of the two training models are shown in Fig.2.

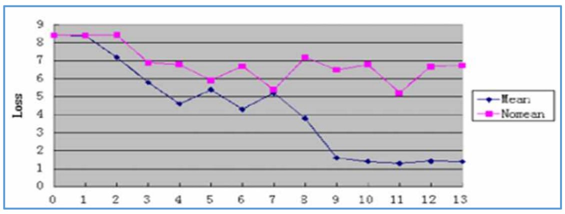

Fig .2 Relationship between calculating mean value and network convergence in preprocessing

The horizontal coordinates in the figure represent the number of iterations of the network, that is, the number of training times, the longitudinal coordinates represent the loss value of the network model.

3.2. network related training parameter settings

Using the strategy of the relevant parameters in Le Net5, the learning rate is increased from 0.01 to 0.02, which is equivalent to increasing the step length, The decrease of the gradient descent method is increased, and the network oscillation is avoided, thus crossing the local minimum point and approaching the larger extremum point. Finally, it is proved by experiments that the value of batch_size is reduced to 64 o'clock, the network can converge at a faster speed, and the classification accuracy is high, which proves that the network also achieves the optimal solution, so this paper sets the Batch_size to 64 and the learning rate to 0.01.

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 handwriting characters are trained with four sub-networks of Cnnet respectively. For cnnet networks with higher complexity (more convolution nuclei and deeper network structures), the recognition rate of handwritten Chinese characters is higher, and their ability to express the sample features is much higher than that of the simpler convolution neural network Net2, that is, more sample categories can be classified.

Although the convolution network with shallow accuracy has increased more, the complexity of cnnet network is higher on the network convergence, and the following table 2 is the comparison between Net2 and cnnet networks in convergence.

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

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 each image size is larger, the network each training weight needs to be used in the sample batch that is batch_size also larger. Coupled with the need for all the test sample 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 Set 4, so that the time spent on each parameter is enormous, This is also a major disadvantage for deep convolution neural networks compared to simple convolution neural networks.

4. System design and implementation

4.1. System structure and design

1. System Framework and related modules

This system architecture adopts BS structure as shown in Fig.3, that is, users only need to have simple interaction with the browser, most of the core business of the system is on the server side. The system of this paper uses Python-based Tornado web framework:

Fig.3. BS architecture diagram

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

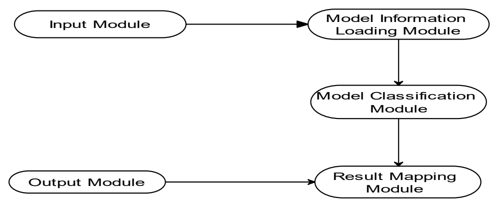

Fig 4 Overall module flow chart of the system

2. Core class Design

The main modules of this system are model loading, model classification and result mapping module, so each module corresponds to design a corresponding core class, complete the main functions of the module.

(1) Load model module; (2) Model classification module; (3) Result mapping module.

4.2. Data-related preprocessing operations

The model used in this system is CNNet3, the input image is a single channel grayscale image, and the test picture which has been processed does not need to be grayscale, but if the color image of RGB is entered, it needs to be preprocessed by grayscale.

There are many ways to grayscale, such as the component method, which uses the pixel value of one of the channels of RGB 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 components weighted, this paper uses the third Way, which represents the grayscale image, respectively, the input image of three sub-vectors, the grayscale formula is as follows:

$$f(x,y) = 0.30(x,y) + 0.59(x,y) + 0.11(x,y) \quad (1)$$

The G component weight is set to the highest, the lowest part of B component is the human eye to green sensitivity is the highest, and the blue sensitivity is the lowest. The input of the system is a multi-formatted character picture, and the correct characters in the picture are obtained by entering the picture. However, in the network model, the initial result of the output is the corresponding character category in the training sample data, and the corresponding transformation is carried out in the mapping module. Therefore, the mapping relationship between characters and categories must be established first in the implementation of this system, as shown in Fig.5.

Fig.5 Mapping relationships for datasets

The CNNet3 network model is used in the system, that is, the output node is 1000. You need to establish the first 1000 classes of the DataSet as shown in the map. As shown in the figure, class_000 represents the first classification of the sample set, in which the folder is the same character dataset, corresponds to the first character "丁" in the mapping file and marks it as n1001.

Acknowledgement

The research is partly supported by the "13th Five-Year" comprehensive investment and basic research business fee items of China (000040192).

References

1. Rong Zhang, Weiping Li, Tong Mo, Review of Deep Learning, *Information and Control*, 2018, 5: P385-P397.
2. Bing Shun, Dehui Kong, Wenhui Zhang, Wenhao Jia, Survey on Human Action Recognition From Depth Maps, *Journal of Beijing University of Technology*, 2018, 10: P1353-P1366.
3. Shengyi Zhou, Techniques Review of Deep Learning, *Information Technology Exploration*, 2018, 10: P116-P117.
4. Anqi Liu, Shuai Zhang, Xinfang Chang, Bingkun chen, Real time human action recognition and monitoring method based on deep learning, *Heilongjiang Science*, 2018, 10: P14-P16.
5. Kai Yang, Youchun Xu, XiangBi An, Yonglei Li, Kunpeng Liu, Vehicle Detection Method Based on Deep Learning, *Army Military Transportation Academy*, 2018, 7: P58-P61.
6. Youwen Huang, Chaolun Wan, Human behavior recognition algorithm based on deep learning, *Artificial Intelligence*, 2018, 10: P1-P5.
7. Qicheng Mei, Wenge Lv, Research on Commodity Image Recognition Based on Deep Learning, *Mechanical and Electrical Engineering technology*, 2018, 9: P27-P151.
8. Huailin Zhao, Zheng Wu, Xiaoxing Wang. A Multi-robot Rescuing System. *The 2018 International Conference on Artificial Life and Robotics* (ICAROB 2018), Feb.1-4, 2018, Beppu, Japan, P582-586.