ISSN: 2414-1895 DOI: 10.6919/ICJE.202102_7(2).0055

MNIST Data Set Recognition Research based on TensorFlow Framework

Jie Liu[1,a], Chenghua Fu[1,b]

[1]College of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin, Sichuan 644000, China.

[a]1138091221@qq.com, [b]1920266248@qq.com

Abstract

TensorFlow is one of the current mainstream frameworks for deep learning. As an end-to-end open source machine learning platform, it has a comprehensive and flexible ecosystem, including rich libraries and community resources. Based on the TensorFlow framework, this paper uses MNIST as the data set to build a model and make prediction, and obtains a convolutional neural network (CNN) to recognize grayscale handwritten numbers. At the same time, compared with traditional classification, the research shows that the CNN has a higher recognition rate for images, and the experimental results show that the recognition rate of the CNN reaches 99.38%.

Keywords

Deep Learning; Tensorflow; Softmax Classification; MNIST; Convolutional Neural Network.

1. Introduction

Recently, deep learning has been widely used in various fields, especially in image recognition and classification with high accuracy that traditional recognition methods cannot achieve. Due to its efficient performance, deep learning is widely used in face recognition, industrial detection, license plate recognition and other aspects. CNN, as the mainstream recognition and feature extraction network for deep learning, has become a research hotspot of artificial neural network. The early CNN model is LeNet-5 proposed by LeCun et al. [1] in 1998. The model has been successfully applied to the recognition of handwritten numbers and has realized the recognition of 0 to 9 numbers. The AlexNet network model proposed by Alex Krizhevsky et al. [2] in 2012 has greatly exceeded the accuracy of image classification on ImageNet data set by traditional methods, and won the championship of ILSVRC 2012 competition at one stroke. The VGG series models proposed by Simonyand et al. [3] in 2014 have won the second place in classification tasks and the first place in location tasks on ImageNet Challenge. Up to now, VGG and other series models still appear frequently as image recognition and feature extraction networks. He KaiMing. [4] proposed the ResNet network model in 2015, which greatly improved the fitting ability of the neural network and further improved the performance of the CNN. Since then, researchers have continuously improved CNNs according to different tasks, which makes them successfully applied in target detection, semantic segmentation, natural language processing and other fields. In this paper, a network model of MNIST handwritten digit recognition is constructed based on TensorFlow. Compared with the traditional Softmax classification, high-precision handwritten digit recognition is realized.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202102_7(2).0055

2. Introduction to MNIST Datasets Section Headings

MNIST data set is a very classic data set in the field of machine learning. It consists of 60,000 training samples and 10,000 test samples. Each sample is a pixel grayscale handwritten digital picture, as shown in Figure 1.

Figure 1. Grayscale Handwritten Numbers

Download on the official website contains four files, training set, training set label, test set and test set label, as shown in Table 1:

Table 1. Data sets

File Name	Size	Content
train-images-idx3-ubyte.gz	9,681 kb	55000 training sets, 5000 verification sets
train-labels-idx1-ubyte.gz	29 kb	Label corresponding to training set picture
t10k-images-idx3-ubyte.gz	1,611 kb	10000 test sets
t10k-labels-idx1-ubyte.gz	5 kb	Label corresponding to the test set picture

3. Softmax Principle

For two classifications, such activation functions $sigmoid, tanh, relu$ can be selected, but in real life it is often necessary to carry out multiple classifications for a certain problem. The main application of the Softmax algorithm is multiple classifications, so Softmax can also be understood as an extension of the above activation functions, and its calculation formula is as follows:

$$softmax(X_i) = \frac{\exp(X_i)}{\sum_{i=1}^{n} \exp(X_i)} \tag{1}$$

$\exp(X_i)$ is an exponential function with a single input and $\sum_{i=1}^{n} \exp(X_i)$ is the sum of all exponential functions with inputs, so the sum of the outputs of the $Softmax$ function is 1, which can also be considered as a probability output between 1.

When using the Softmax method, its function is to infer the probability of which number a picture is. Since the data set contains 0-9 ten numbers, the output of Softmax is a probability array of 1*10. For example, if a picture with 3 is processed by the Softmax algorithm, the probability of getting the number 9 is 5%, the probability of representing 8 is 10%, the probability of representing 7 is 1%... the probability of representing 3 is 70%, and the probability of representing other numbers is only

ISSN: 2414-1895 DOI: 10.6919/ICJE.202102_7(2).0055

14%. Taking the number corresponding to the maximum probability is the classification of this graph. In the test MNIST data set, the results obtained by Softmax algorithm in five tests are shown in Table 2.

Table 2. Test Results Table

Number of tests	Training dataset size	Test dataset size	Recognition rate
1	60000	10000	0.9182
1	60000	10000	0.9213
1	60000	10000	0.9195
1	60000	10000	0.9216
1	60000	10000	0.9196

4. Improved Model Based on TensorFlow Framework

4.1 Introduction of CNN

CNN is one of the most classical models in deep learning. CNNs can be found in all the classical models of deep learning today. It skillfully uses very little weight but achieves the effect that the fully connected network cannot achieve [5]. At the same time, the CNN uses the method of parameter sharing, which not only greatly improves the accuracy, but also greatly reduces the required parameters on the color map compared with the full connection layer. The structure of CNN is roughly divided into input layer, convolution layer, pooling layer, full connection layer and output layer. As shown in Figure 2

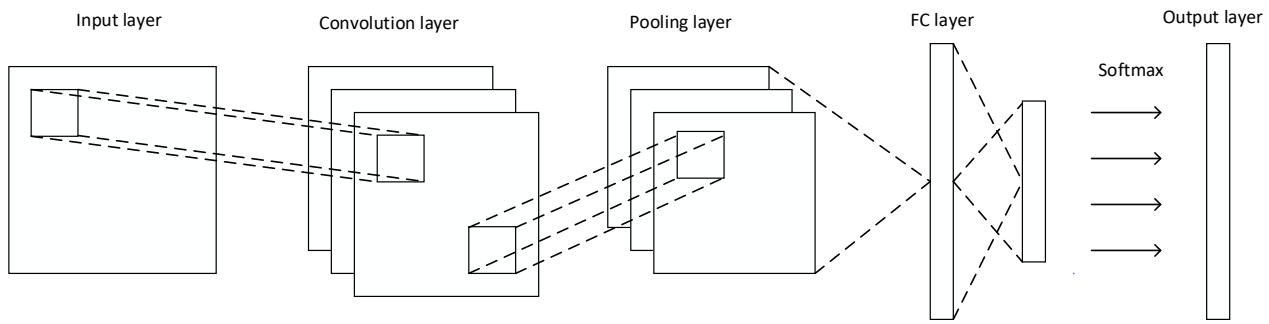

Figure 2. Convolution network structure diagram

4.2 Improvement of TensorFlow Deep Learning

Compared with Softmax method, convolution layer, pooling layer, global average pooling layer, etc. are added to the improved model. At the same time, batch normalization treatment is added after convolution. Batch normalization treatment effectively alleviates the problems of gradient explosion and disappearance in deep neural network training [6], and Softmax function is used in the final average pooling layer and Relu function is used as activation functions in other layers. Multiple convolution kernels are used in the convolution process, so the feature map generated by this method has diversity and richer features. In the training process, in order to prevent the phenomenon of over-fitting, dropout method is adopted, that is, in the training process, a part of nodes are randomly selected each time not to be trained. In this paper, the training season keep_prob=0.5, that is, half of the nodes are randomly ignored each time and do not participate in the operation. In order to speed up the operation during training, the convolution kernel is disassembled. The 5*5 convolution kernel is decomposed into 5*1 and 1*5 convolution kernels, and then the original input is convolved respectively. The improved model structure is shown in Figure 3.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202102_7(2).0055

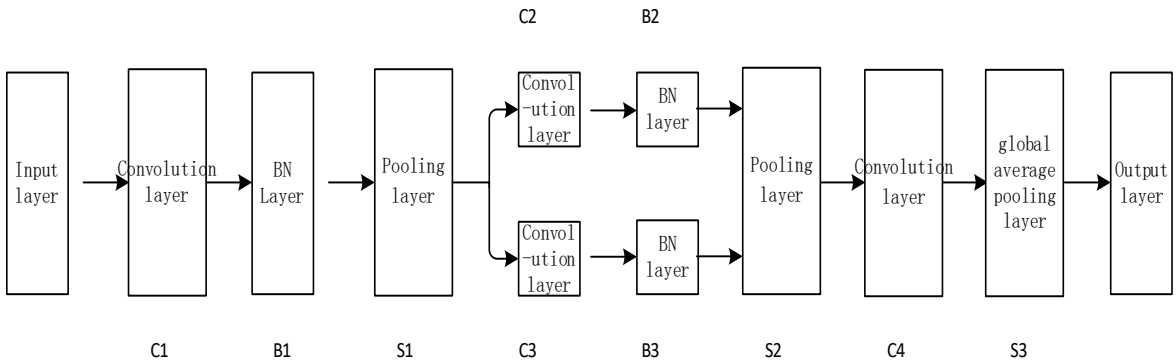

Figure 3. Structure diagram of improved model

The whole network structure consists of 10 layers (excluding input and output layers), 4 layers of convolution layer, 3 layers of pooling layer and 3 layers of BN layer. The result of the upper layer of the network structure is the input of the next layer. Input a single channel gray image of 28*28, and convolution operation is carried out on the image in C1 layer. The convolution kernel is 5*5, the step size is 1, and the B1 layer is batch normalization processing. The batch normalization converges very fast and has strong generalization ability. The results obtained from B1 layer are sent to S1 layer, and the maximum pooling process is adopted in S1 layer. The core size used in this layer is 2*2 and the step size is 2*2. The main purpose of pooling is to reduce the dimension, that is, to reduce the dimension of the array to the maximum extent on the basis of maintaining the original characteristics. Two convolution kernels are used in the C2 and C3 layers, For feature extraction, 5*5 and 7*7 convolution kernels are used respectively, and the step size is 1*1. The operation used in S2 layer is the same as S1. The convolution kernel with 5*5 and 1*1 step size is used in C4 layer, and the convolution kernel with 7*7 and 7 step size is used in S3 layer to obtain the average pooling layer of 10 nodes. Finally, the $Softmax$ activation function is used to obtain the output result. Subsequently, 5 trainings are carried out for 20,000 iterations each time. The experimental results are shown in Table 2.

Table 2 Test Results Chart

Number of trainings	Training dataset size	Test dataset size	Test accuracy
1	60000	10000	0.9903
1	60000	10000	0.9932
1	60000	10000	0.9931
1	60000	10000	0.9933
1	60000	10000	0.9938

5. Conclusion

TensorFlow is an open source software library that can use numeric classes to calculate, which runs stably in the Python locale [7]. It is often used in many fields of machine deep learning such as perception, language understanding, speech recognition, image recognition, etc. [8]. In this paper, CNN is built based on TensorFlow, and MNIST is used as data set for recognition research. Compared with Softmax traditional methods, two experiments are carried out. The recognition accuracy of the improved model reaches 99.38%, the effect is very ideal, and the overall performance is excellent.

References

[1] Lecun, Y, L. Bottou, Gradient-based learning applied to document recognition. Proceedings of the IEEE, Vol. 86(1998) No. 11, p. 2278-2324.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202102_7(2).0055

[2] Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Vol.25(2012), 1106–1114

[3] Simonyan K, Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition [EB/OL]. (2015-04-10). http://arxiv.org/abs/1409.1556.

[4] K. He, Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society.

[5] J. H. Li, TensorFlow of Deep Learning: Introduction, Principle and Advanced Actual Combat. Beijing: Mechanical Industry Press, 2018.

[6] Y. Wang. Research on Normalization Technology of Deep Neural Network. Nanjing University of Posts and Telecommunications, 2019

[7] M.M. Zhang, H.P. Xu, X.J. Wang, et al. Google Tensor-Flow Machine Learning Framework and Applications. Microcomputers and Applications, Vol. 36(2017) No. 10,p.958-60.

[8] R. Huang, X. M. Lu, Y.L. Wu. Handwritten numeral recognition and application based on deep learning of TensorFlow. Electronic Technology Applications, Vol. 44(2018) No. 10, p. 6-10.