

PAPER • OPEN ACCESS

Improvement of MNIST Image Recognition Based on CNN

To cite this article: Yifan Wang *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **428** 012097

View the [article online](#) for updates and enhancements.

You may also like

- [Variational quantum one-class classifier](#)
Gunhee Park, Joonsuk Huh and Daniel K Park
- [Relative stability toward diffeomorphisms indicates performance in deep nets](#)
Leonardo Petrini, Alessandro Favero, Mario Geiger et al.
- [MNIST-MIX: a multi-language handwritten digit recognition dataset](#)
Weiwei Jiang



245th ECS Meeting • May 26-30, 2024 • San Francisco, CA

Present your work at the leading electrochemistry & solid-state science conference.

Network with academic, government, and industry influencers!

Submit abstracts by December 1, 2023

[Learn more & submit!](#)



Improvement of MNIST Image Recognition Based on CNN

**Yifan Wang^{1,3}, Fenghou Li², Hai Sun⁴, Wenbo Li¹, Cheng Zhong⁵, Xuelian Wu¹,
Hailei Wang¹, Ping Wang⁶**

1 Institute of Technology Innovation, Hefei Institute of Physical Science, Chinese Academy of Sciences, Hefei, China

2 State Grid Hulunbeier Power Supply Company, Hailaer, China

3 Anhui University, Hefei, China

4 China Telecom Shanghai Ideal Information Industry (Group) Co., Ltd, Shanghai, China

5 Beijing Aerospace Flight Control Center, Beijing, China

6 Department of Airborne Remote Sensing, National Disaster Reduction Center of China, Beijing, China

Email: 158020280@qq.com

Abstract. At present, great progress has been made in the field of image recognition, especially in convolutional neural network. Lenet-5 convolutional neural network has been able to identify handwritten digit MNIST database with high precision. In this paper, experiments show that different activation functions, learning rates and the addition of the Dropout layer in front of the output layer will make the convergence speed different, weaken the influence of the initial parameters on the model, and improve the training accuracy. It is proved that the modified LeNet-5 model has a better improvement in handwritten digit recognition. This method is an efficient recognition method.

1. Introduction

In recent years, with the rapid development of cloud computing, big data, artificial intelligence and other technologies, the application of computer technology has led to the explosive growth of data scale [1]. Since it was proposed, deep learning theory has been highly valued and widely applied in artificial intelligence technologies such as image recognition, voice judgment and intelligent decision-making [2].

Handwritten digital data sets are a technique for automatically recognizing Arabic numerals. The technology is widely used and can be applied to the automatic identification of handwritten data such as postcode, bank notes, accounting. With the development of the times, this identification technology is more and more widely used. Due to the close relationship between the application field and people's vital interests, the requirements and the accuracy of identification must be higher. Therefore, it is of great significance to know how to improve accuracy.

Lenet-5 [3] model is used to identify the traditional MNIST [4-5] (Mixed National Institute of Standards and Technology database) handwritten digital data sets and it is proved that the accuracy will be able to reach 98% after training.

2. MNIST

MNIST is an entry-level computer vision dataset in a neural network.



2.1. MNIST Data Set

In machine learning, in order to achieve a certain function, it is required to put a large amount of data into the established neural network input. The user is different, and the purpose of the input data is also different. The performance of different neural network models is well shown by using the MNSIT data set. MNIST contains a huge number of digital images, encapsulated by 70,000 digital images. In the MNSIT data set, it is further divided into 60,000 training data sets and 10,000 test data sets. Each picture is a single number from 0 to 9, as shown in Figure 1.



Figure 1. MNIST Numbers

2.2. Data Set Classification

The MNIST data set has a total of four files, namely training set, training set label, test set, test set label, as shown in Figure 2.

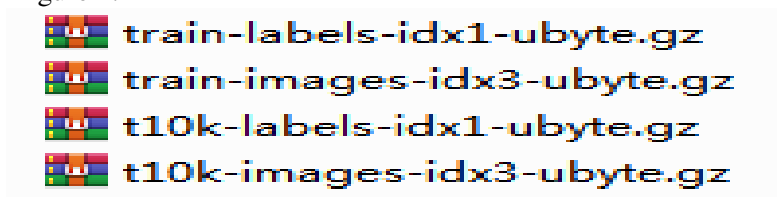


Figure 2. MNIST Dataset File Classification

MNIST data is divided into two parts: mnist.train and mnist.test. This division has great symbolic significance as it shows how data can be used in machine learning. In the process of training, we must keep a separate data for machine training as verification data, in order to ensure that the training results can be generalized. Each piece of MNIST data consists of images and tags. The training image is called mnist.train.images and the training label is mnist.train.labels.

2.3. Reading MNIST

OpenCV is used to read data and labels. Figure out the image pixel and label values, the pixel matrix can be saved as an image file using cv2 module, and the image data with the following form can be parsed out, as shown in Figure 3.

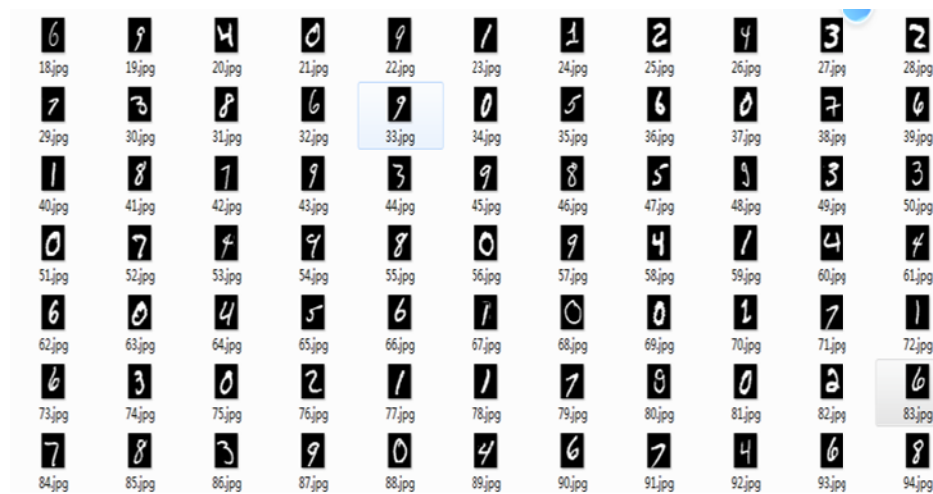


Figure 3. Analyzed Image Data

The first image sample data is shown in Figure 4. This number is 5, and each resolved image is 28×28 in size, with the pixel value ranging from 0 to 255.

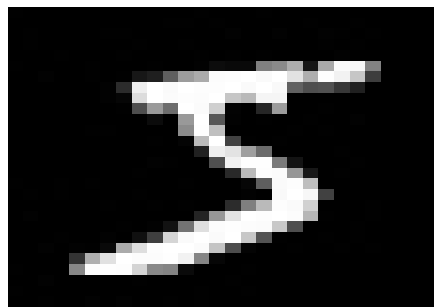


Figure 4. First Image Sample

3. Traditional Lenet-5 Model Structure

The convolutional neural network [6] proposed by LeCun et al, like the neurocognitive machine, is a model based on the structure of the sensory field in human visual cortex. Lenet-5 [7] algorithm is a convolutional neural network proposed by LeCun. It is a special multi-layer neural network. Like other neural networks, it is also trained by backpropagation. The difference is its network structure. It differs from ordinary neural networks in the local receptive domain and weight sharing [8], this let Lenet-5 to reduce the number of parameters in the process of building the network and speed up the learning process.

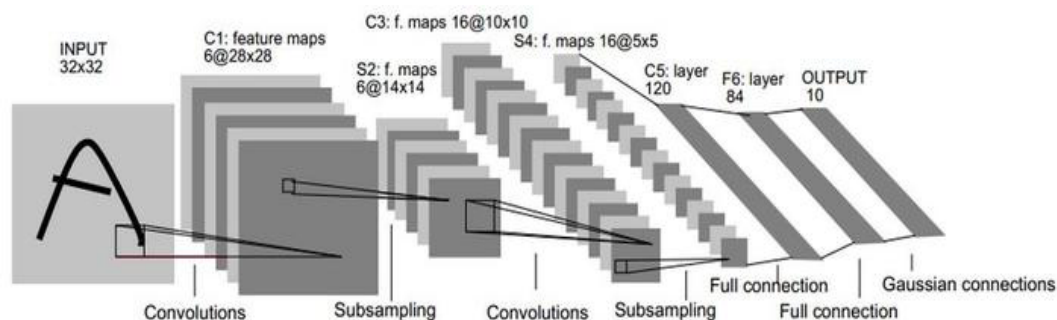


Figure 5. Lenet-5 Network Structure

The network structure of Lenet-5 is shown in Figure 5. Lenet-5 is a typical convolutional neural network model. The network consists of 8 layers of input and output layers. It is composed of convolution layer, pooling layer and full connection layer. The convolution layer and pooling layer cooperate to form multiple convolution groups, which are extracted layer by layer and classified through several full connection layers. The operation of the convolution layer can be thought of as inspired by the concept of local receptive field. On the other hand, the pooling layer is mainly to reduce the data dimension. The output layer has 10 nodes representing the Numbers 0 to 9. Therefore, it is very intuitive and significant to build handwritten numeral recognition model with lenet-5 model.

4. Model Improvements

In the traditional Lenet-5 model, the Sigmoid function is used. Experimental results show that Sigmoid function is not as effective as ReLU and Elus. So, improve the model the learning rate is modified to accelerate the convergence speed and make the model more accurate.

4.1. Improvement of Activation Function

4.1.1. Sigmoid Function

Sigmoid function is a commonly used activation function. The main function of the activation function is to add non-linear factors to solve the defect of inadequate model expression ability, which plays a crucial role in the whole neural network. The expression of the Sigmoid function is:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

The function image is shown in Figure 6.

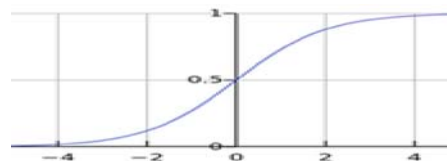


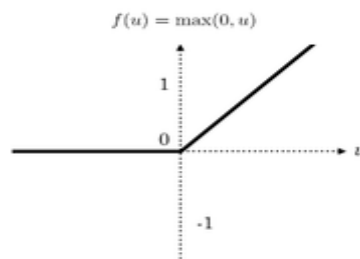
Figure 6. Sigmoid Function Image

Sigmoid function is a kind of saturation function which is one with finite upper and lower bounds. So does Tanh function. As you can see, the usual choice of transfer function is a sigmoidal (S-shaped) response function, such as the logistic and hyperbolic tan functions, which transform an input in the range $(-\infty, +\infty)$ to the range $(0, 1)$. Certain functions have limits. As output values approach this limit the resulting curve starts to level out (becomes horizontal). Thus, when using Sigmoid and Tanh, keep input absolute value under the limits so that the generated results would be effective.

The domain of x is $(-\infty, +\infty)$, and the range of $f(x)$ is $(0,1)$. The Sigmoid function compresses the input value between 0 and 1, infinitely close to 0 or 1. This is called saturation. It means that $x = x = 100$ and 1000 reflect the data is the same, resulting in data loss, the impact of the final results of the experiment. Tanh function and Sigmoid have the same defect of saturation. Therefore, when using Sigmoid and Tanh, the input value should not be too large, otherwise the model training effect is not good or even unable to train.

4.1.2. ReLU Function

ReLU function is very simple, if it is greater than 0, it remains. Otherwise, it is 0, as shown in Figure 7. ReLU ignores the characteristics of negative signals and is similar to the reflection of signals by human neuronal cells. Therefore, the ReLU function has achieved a good fitting effect in the neural network and has achieved wide application.

**Figure 7.** ReLU Image

The expression of the ReLU function is $f(u) = \text{Max}(0, u)$. Because of its simple form, it greatly improves the operating efficiency of the machine. The convergence speed of ReLU function is much faster than Tanh function, and only a threshold value is needed to obtain the activation value, which can save a lot of training time of gradient descent without a lot of complex calculations. ReLU function can make the output of some neurons 0, which leads to the sparsity of the network, reduces the interdependence of parameters, alleviates the over-fitting problem, and makes the whole model more efficient.

4.1.3. Elus Function

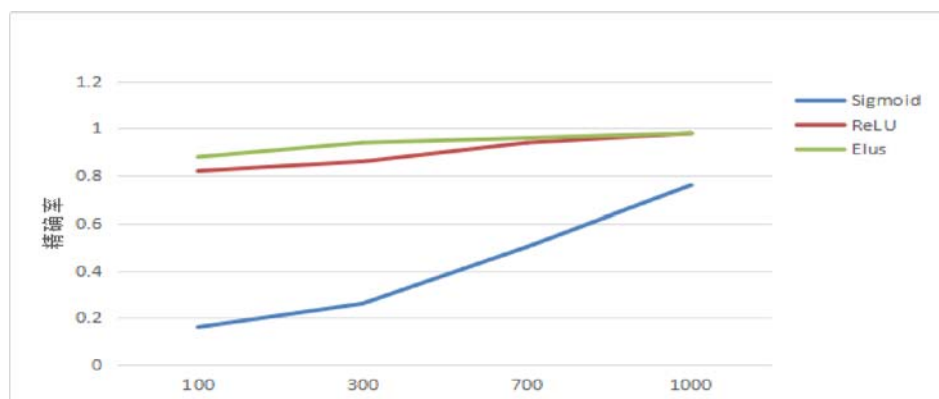
Its mathematical form is:

$$f(x) = \begin{cases} x(x \geq 0) \\ a(e^x - 1)(x < 0) \end{cases} \quad (2)$$

The form is the same as ReLU, with no parameters.

4.1.4. Experimental Results

Change the type of activation function to make the model optimized. Achieve greater improvements in accuracy, as shown in Table 1.

Table 1. Accuracy of Different Activation Functions

It can be seen from Table 1 that ReLU and Elus have higher accuracy than the Sigmoid function, for the Sigmoid function cannot solve the problem of neuron gradient dissipation. The ReLU function has unilateral suppression. The network is not prone to gradient dissipation during training, and has sparse characteristics, which make the network converge faster, about 6 times that of Sigmoid. Therefore, it has a better classification effect than the traditional activation function. As a function of the later improvement of ReLU function, Elus's network activation method is closer to the activation mode of biological neurons, and has achieved good results. The experimental results in Figure 8 also

verify this point. In practical applications, it is generally recommended to use ReLU on the classification problem, but when using Elus, model is significant to adjust the learning rate, otherwise the model is prone to shock.

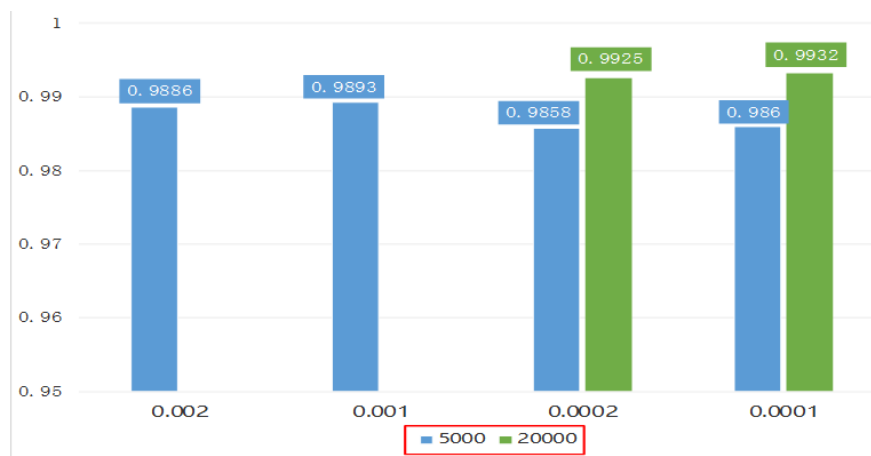
4.2. Change Learning Rate

Learning rate is used to determine the weight coefficient adjusting the degree of connection. Gradient descent is a type of linear regression. When using the gradient descent method, first initialize the parameter values, and then change these values until the global minimum is obtained. The first step is to set a learning rate θ . Learning rate setting is very important, if the setting is too large, the algorithm may not converge and may oscillate. Otherwise, the convergence speed may be too slow.

As the learning rate is adjusted every time, the data will be too large and take up too much memory. With the development of the time, a stochastic gradient descent algorithm is proposed. Each time only one sample is randomly selected to update the model parameters, so learning is very fast and can be updated online. But each update may not be carried out in accordance with the right direction, so you can bring to optimize fluctuations. Due to the fluctuation, the number of iterations will increase, that is, the convergence speed will slow down. But eventually it will gradient descent algorithm and the whole quantity, with the same convergence, the convex function converges to the global extreme value point, non convex loss function converges to local extremum points, fluctuations brought by the stochastic gradient descent one advantage is that, for similar basin area so the characteristics of the fluctuation may make optimization direction from the current local minima point to another better local minimum points, thus may for the convex function, finally converge to a local extremum points better, and even the global extreme value point. Both Adagrad, Adadelta, RMSprop and other algorithms have their own advantages and disadvantages, and the best optimization method is still self-tuning.

For the analysis of network recognition results, the influence of learning-ratio set learning-ratio respectively 0.0001, 0.0002, 0.001, 0.002. The results show as Table 2.

Table 2. Parameter Change



4.3. Introducing Dropout

To reduce over-fitting, add Dropout before the output layer. Use the `tf.nn.dropout()` function to implement Dropout, controlled by the `keep_prob` parameter.

In deep learning, over-fitting is improving the performance on the training data set, while the performance on the test data set is declining. The essential cause of over-fitting is due to the incompatibility of supervised learning. In deep learning, if the number is much smaller than the model space, over-fitting is likely to occur. Solve three ways of over-fitting:

- (1) Increase the training data set
- (2) Regularization
- (3) Dropout

Dropout is implemented by modifying the neural network itself, which is a technique used when training the network.

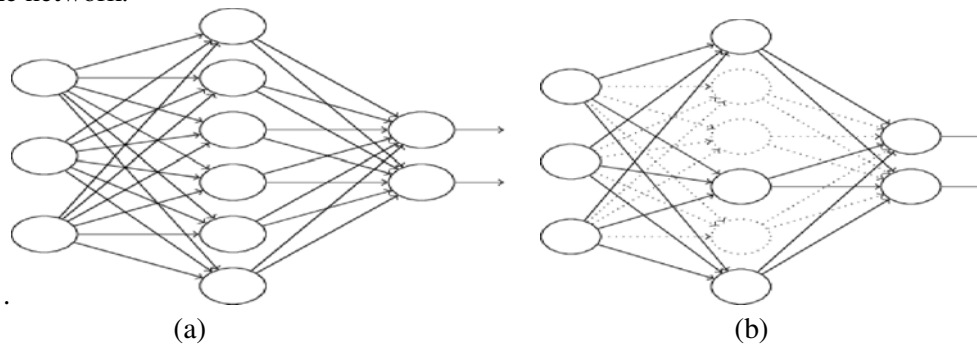


Figure 8. Dropout Iteration

When training the network of Figure 8(a), we randomly "delete" half of the hidden layer units and treat them as non-existent, resulting in the following network, as shown in Figure 8(b).

Keep the input and output layers unchanged, update the weights in the above-mentioned neural network according to the BP algorithm, and the units connected by the dotted lines are not updated because they are "temporarily deleted".

The above is an iterative process. In the second iteration, the same method is used, except that the half hidden layer unit deleted this time is definitely different from the last one, because we have every iteration. It is to delete half at random. It's been like this until the end.

Using the training process of dropout, it is equivalent to training a large number of neural networks with only half of the hidden layer units, also called "half network". Each such half network can give a classification result. Some of these results are correct. Some are wrong. As the training progresses, most of the half networks can give correct classification results, so a small number of misclassification results will not have a big impact on the final result.

5. Conclusions and Prospects

This paper improves the traditional MNIST image recognition method based on convolution neural network. Three activation functions have been tested successively. As can be seen from the figure, the effect of ReLU and Elus is obviously better than that of traditional Sigmoid. Its advantage is that it can retain the characteristic data to the greatest extent, and realize the data with sparse matrix with most elements of 0. From the data can also be found that Elus ReLU function is slightly better than others. By changing the learning rate, it can be found that if the learning rate is not set properly, the network will fall into a local minimum value, resulting in no convergence and over-fitting. Therefore, manual parameter adjustment is still needed in the experiment process to determine the best data. By modifying the activation function, learning rate and joining Dropout, the misidentification rate was reduced from 2% to 0.7%. It can be seen that through the improvement of lenet-5 model, the accuracy rate has been greatly improved. There are many areas for improvement based on lenet-5. Therefore, how to improve the recognition performance of the model remains to be further studied.

6. Acknowledgments

This paper is financially supported by the National Natural Science Foundation of China (No. 61773360, 41871302, 41701594) and the National Key Research and Development Program of China (No. 2018YFB1004600).

7. References

- [1] Marz N and Warren J 2015 Big Data: Principles and Best Practices of Scalable Realtime Data Systems. Greenwich, USA: Manning Publications Co.
- [2] Z. J. Sun, L. Xue, Y. M. Xu, and Z. Wang, 2012 Overview of deep learning *App. Res. of Comp.*, vol. 29(8) pp. 2806–2810 (in Chinese).

- [3] He K, Zhang X, Ren S and Sun J 2016 Deep Residual Learning for Image Recognition *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, IEEE Computer Society* pp 770-778.
- [4] Delahunt Charles B and Kutz J Nathan 2019 Putting a bug in ML: The moth olfactory network learns to read MNIST *Neural networks: The Official J. Int. Neu. Net. Society* pp 118.
- [5] Leon M D, Moreno-Baez A, Magallanes-Quintanar R and Valdez-Cepeda R D 2011 Assessment in Subsets of MNIST Handwritten Digits and their Effect in the Recognition Rate *J. Pat. Recog. Res* vol 2(2) pp 244-252.
- [6] Moeskops P, Viergever M A, Mendrik A M, de Vries LS and Benders MJ, Isgum I 2016 Automatic Segmentation of MR Brain Images with a Convolutional Neural Network *IEEE Trans. on Med. Imaging* vol 35(5) pp 1252-1261.
- [7] Lecun Y, Bengio Y and Hinton G 2015 Deep Learning *Nature* vol 521(7553) pp 436-444.
- [8] Ilmi N, Budi W T A and Nur R K 2016. Handwriting digit recognition using local binary pattern variance and K-Nearest Neighbor Classification *Int. Conf. on Info and Comm. Tech. IEEE* pp 1-5.