Computer Science and Information Systems 19(3):1463–1480 https://doi.org/10.2298/CSIS220120036Z

A Novel Deep LeNet-5 Convolutional Neural Network Model for Image Recognition

Jingsi Zhang, Xiaosheng Yu, Xiaoliang Lei, and Chengdong Wu*

Faculty of Robot Science and Engineering, Northeastern University
Shenyang 110819, China
newmansuper@163.com
yuxiaosheng@mail.neu.edu.cn
xiaolianglei@stumail.neu.edu.cn
wuchengdongnu@163.com

Abstract. At present, the traditional machine learning methods and convolutional neural network (CNN) methods are mostly used in image recognition. The feature extraction process in traditional machine learning for image recognition is mostly executed by manual, and its generalization ability is not strong enough. The earliest convolutional neural network also has many defects, such as high hardware requirements, large training sample size, long training time, slow convergence speed and low accuracy. To solve the above problems, this paper proposes a novel deep LeNet-5 convolutional neural network model for image recognition. On the basis of Lenet-5 model with the guaranteed recognition rate, the network structure is simplified and the training speed is improved. Meanwhile, we modify the Logarithmic Rectified Linear Unit (L_ReLU) of the activation function. Finally, the experiments are carried out on the MINIST character library to verify the improved network structure. The recognition ability of the network structure in different parameters is analyzed compared with the state-of-the-art recognition algorithms. In terms of the recognition rate, the proposed method has exceeded 98%. The results show that the accuracy of the proposed structure is significantly higher than that of the other recognition algorithms, which provides a new reference for the current image recognition.

Keywords: CNN, image recognition, feature extraction, deep LeNet-5, L_ReLU.

1. Introduction

With the development of science and technology, computer vision has been widely used in various fields. The core technologies of these applications are image processing [1], image recognition [2] and classification tasks [3]. The recognition technology is to calculate the characteristics of the samples and apply them to the classifier to generate classification for different calculated values.

Since 1980s, research on optical character recognition methods has always been a hot topic in pattern recognition [4]. It is not easy for a computer to correctly recognize a large number of handwritten fonts because different people have different habits of writing numbers. Therefore, it is of great significance to study an accurate and efficient number recognition method.

* Corresponding author

For image recognition methods, the traditional recognition methods such as support vector machine, traditional neural network, the K-nearest neighborhood method (KNN), have some shortcomings. The minimum distance classification algorithm is a traditional recognition algorithm, but it is not suitable for handwritten fonts. The recognition method of KNN is derived from statistics [5]. The principle is to calculate the features of the image and measure the distance between the calculated results of different features for classification. Its advantage is that it is insensitive to abnormal data collection. SVM has been successfully applied to image recognition [6]. In machine learning, SVM can avoid the complexity of high-dimensional space, and it is very prominent in small sample, high-dimensional space calculation and nonlinear problems. However, in the classification problem, the storage space occupied by solving the function is large. These traditional recognition algorithms mentioned above have very poor expression ability for more complex mathematical functions, poor generalization performance, and usually fail to reach the expected effect of data prediction and accuracy.

The emergence of convolutional neural network (CNN) provides the possibility to solve the generalization ability of image recognition [7,8]. Convolutional neural network, as a successful model in deep learning, has been widely applied in the field of image recognition.

CNN can extract the hidden features of human face by using hardware acceleration technology and massive face image data training. This feature is highly invariant to scale changes such as translation, zoom and tilt, and has certain robustness to complex fonts. Therefore, the research on image recognition based on CNN is very active. For example, the reference [9] achieved the fusion of high and low level features by improving the structure of AlexNet. Reference [10] integrated binary tree with ResNet convolutional neural network, and put forward a binary tree CNN information fusion model for image recognition.

In this paper, we propose a novel deep LeNet-5 convolutional neural network model for image recognition. The network structure of LeNet-5 is improved, and a new activation function L_Relu is used to solve the over-fitting phenomenon in the training process. The experiment of network structure is carried out through MINIST database to improve the operation speed of network structure and the recognition accuracy of the proposed algorithm.

This paper is organized as follows. In section 2, we give the related works including CNN and LeNet-5. Section 3 presents the proposed image recognition method in detail. Experiments and analysis are conducted in section 4. We make a conclusion for this paper in section 5.

2. Related works

2.1. Structure of CNN

The traditional CNN is generally composed of five parts: the input layer, the convolution layer, the down-sampling layer, the fully connection layer and the output layer [11]. The network structure is shown in figure 1.

For general multi-layer neural networks, the first layer is the eigenvector [12]. In general, the image is processed manually to obtain the feature vector, which is used as the

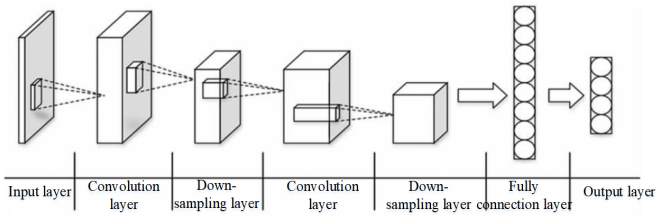


Fig. 1. The Structure of CNN

input of the neural network. The convolution neural network is different from general multi-layer neural network, and the whole image is as the input of the network. For example, the experimental object of this paper is the handwritten digital image in the MINIST database, and the size of the image obtained after processing is 28×28. In order to facilitate the call and reading of data, the image can be expanded according to the pixel number.

The convolution layer is also the feature extraction layer. Convolution operation is the soul of convolutional neural network, and convolution kernel is the tool of convolution operation[13]. The principle of convolution operation is shown in figure 2.

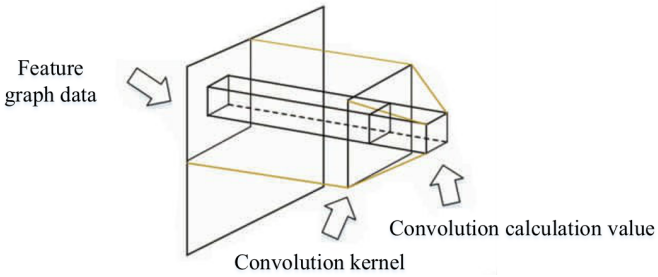


Fig. 2. Principle diagram of convolution operation

The down-sampling layer is also known as the pooling layer. During the pooling of feature graphs, the depth of the image does not change, but the size of the image can be reduced. Pooling can be viewed as converting a high resolution image to a low resolution image with a smaller size. Through multiple pooling layers, the number of parameters in the final fully connection layer can be gradually reduced to reduce the parameters of the whole neural network and improve the training speed. The pooling layer has no parameters that can be trained. The fully connected layer is the same as the normal fully connected layer. Its input layer is the previous feature graph, which will transform all neurons in the feature graph into data in the fully connection layer.

2.2. Structure of LeNet-5

The LeNet-5 convolutional neural network model for image recognition is shown in figure 3, which consists of input layer, hidden layer and output layer. The input layer is a 32×32 single channel target image. The hidden layer is responsible for the extraction and classification of object features. The output layer outputs an integer representing the category.

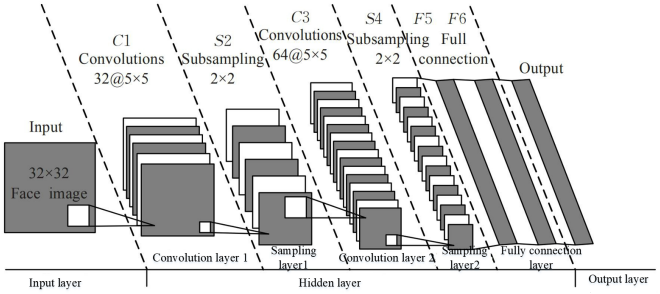


Fig. 3. The LeNet-5 convolutional neural network model

The hidden layer of the convolutional neural network is generally composed of Convolutions, Sub-sampling and Fully connection. The model in figure 3 contains two convolution layers (C1 and C3). C1 and C3 have 32 5×5 convolution kernels and 64 5×5 convolution kernels respectively. In the process of image recognition, the convolution kernel and the input image convolve each 5×5 region to extract the feature graphs that are highly robust to scale changes. When the convolution step size is 1, the convolution operation is shown in equation (1):

$$H^s_{i,j} = \sum_{m=0}^{k} \sum_{n=0}^{k} W^s_{m,n} X_{m+i,n+j} + b^s. \tag{1}$$

In equation (1), X represents an image with a depth of 1 and a size of $u \times v$. W represents the convolution kernel of $k \times k$. S represents the feature graphs of different convolution kernels. $H^s_{i,j}$ represents the element in row i, column j of the feature graph H^s matrix. b^s is the Bias value corresponding to the convolution kernel.

S2 and S4 are pooling layers. Through multi-layer sampling, CNN can reduce the dimension of feature graph and eliminate repeated features [14], so that features have certain translational invariance. F5 and F6 are fully connection layers similar to traditional multi-layer perceptron neural networks, with 1024 and 67 neurons respectively. The image features obtained by the convolution layer and the sampling layer are dot product with the weight of the multi-layer perceptron neural network, and then the feature classification is completed by the Sigmoid function.

Recently, many researchers developed LeNet-5-based methods to process the images. For example, In reference [15], the recognition of haze images was performed by adjusting the parameters and structure of the classic LeNet-5 model. The image recognition

technology was applied to a haze image field, which showed good performance. Zhang et al. [16] proposed an improved Lenet-5 algorithm for traffic sign recognition. The picture noise elimination and image enhancement on selected traffic sign images were performed. Then, Gabor filter kernel was adopted in the convolution layer for convolution operation. In the convolution process, the normalization layer Batch Normality (BN) was added after each convolution layer and reduced the data dimension. Zhang et al. [17] studied the detection of hyperthyroidism by the modified LeNet-5 network.

The accuracy statistics result is shown in figure 4. After nearly 10000 times training, the recognition accuracy of the network is still far less than 1, indicating that the convergence speed of this network model is slow and its learning ability is poor. The accuracy of the training set is always much higher than that of the test set, which indicates that the network is over-fitting and the generalization ability is poor. Through the above analysis, it can be seen that the LeNet-5 model-based convolutional neural network has poor image recognition effect, and the network needs to be improved.

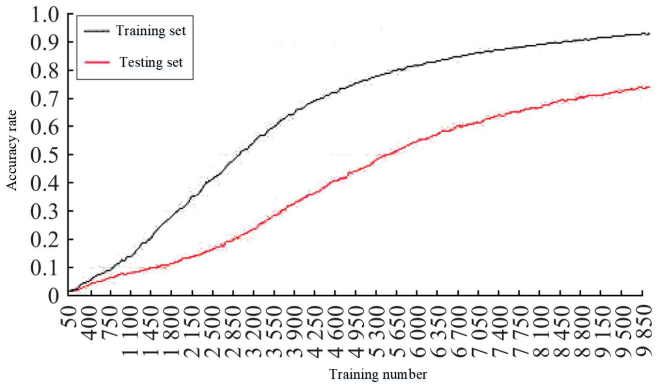


Fig. 4. Training results with LeNet-5 for image recognition

3. Proposed LeNet-FC convolutional neural network

3.1. Activation function optimization

The reason why CNN based on LeNet-5 model has a slow convergence speed in image recognition training is that the Sigmoid activation function appears the gradient disappearance phenomenon when the network is trained by gradient descent method. Sigmoid function is shown in equation (2):

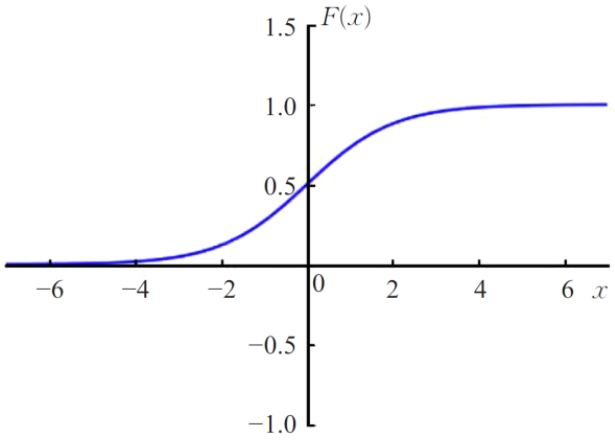
$$Sigmoid(x) = \frac{1}{1+e^{-x}}. \tag{2}$$

In the gradient descent training method, the parameter updating is mainly based on the gradient value to achieve training optimization. This gradient is calculated by backward

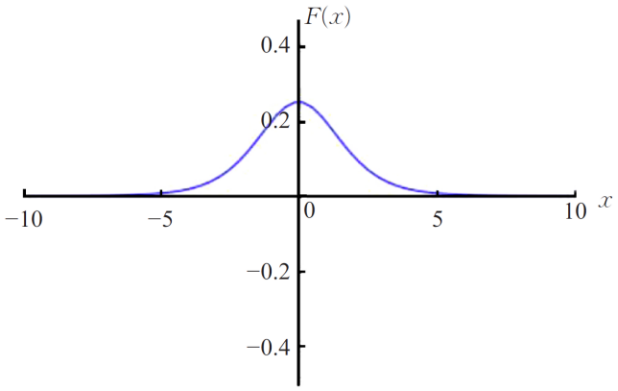
recursion from the output layer based on the error between the recognition result and the real result as shown in equation (3):

$$Grad = Error \times Sigmoid'(x) \cdot x. \tag{3}$$

In equation (3), Grad represents the gradient of the current layer. Error denotes the error. x is the input value of the current layer. $Sigmoid'(x)$ represents the first derivative of the Sigmoid function. According to figure 5(a), the Sigmoid function is double-ended saturated. As shown in figure 5(b), when x value is too large or too small, the first derivative of Sigmoid function will approach 0. This will cause the gradient value to be greatly attenuated in the backpropagation calculation, or even disappear to 0. Therefore, the network parameters are updated with a minimal gradient value in the training, and the convergence speed of the network is slow or even unable to converge.



(a) Sigmoid function graph



(b) First derivative of Sigmoid function

Fig. 5. Sigmoid function and its first derivative graph

In this paper, we present the statistatized Rectified Linear Unit (L_ReLU) in the activation function optimization. Its expression is shown in equation (4):

$$L_ReLU(x) = ln(\frac{1+e^x}{2}) + 0.1x. \tag{4}$$

According to the expression of L_ReLU and the function in figure 6(a), it can be seen that L_ReLU has five basic properties which is necessary to become an activation function:

1. Non-linearity. L_ReLU function is nonlinear, and it can play a good role of nonlinear mapping in CNN.
2. Differentiability. The derivative of L_ReLU function is shown in equation (5):

$$L_ReLU'(x) = \frac{e^x}{1+e^x} + 0.1. \tag{5}$$

   Therefore, the training method based on gradient can be adopted.
3. Monotonicity. $L_ReLU'(x) > 0$ shows that L_ReLU function is monotonically increasing. This can guarantee that each layer of network in CNN is convex function.
4. $f(x) \approx x$. L_ReLU function satisfies this condition when $x > 0$. The network can be initialized with a small random value to obtain a good training effect.
5. The output value is infinite. The output value of L_ReLU function is infinite. When the model is trained with a small learning speed, it can obtain a higher training efficiency.

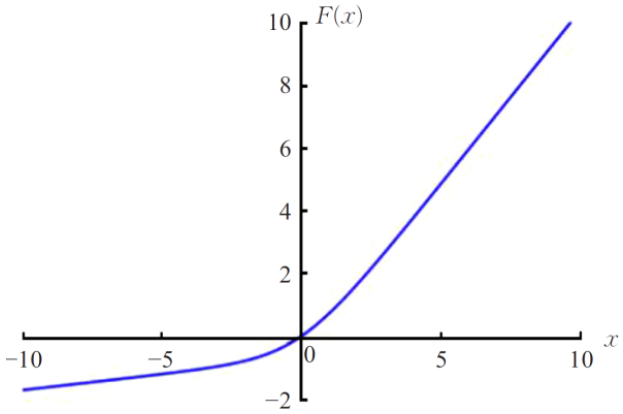
Compared with Sigmoid function, L_ReLU activation function does not appear gradient disappearance in gradient descent method. By computing the limit of the derivative of L_ReLU in equation (5), it can be seen that when x tends to positive infinity and negative infinity, the limits are 1.1 and 0.1 respectively. As shown in figure 6(b), when x is too large, the derivative value of L_ReLU is close to 1.1. When x is too small, its value will be close to 0.1, and will not be 0. Therefore, L_ReLU activation function can be used in CNN to carry out effective gradient descent training.

The Rectified Linear Unit (ReLU), Softplus [18] and the proposed L-ReLU activation function are compared and analyzed. The curves of the three functions are shown in figure 5. The expressions of ReLU and Softplus are in equations (6) and (7) respectively, while the expression of L_ReLU is shown in equation (4).

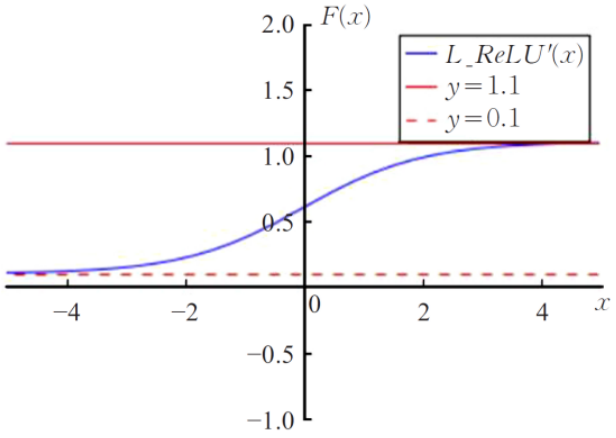
$$ReLU(x) = max(0, x). \tag{6}$$

$$softplus(x) = lb(l + e^x). \tag{7}$$

According to figure 7, it can be seen that the ReLU function has the characteristics of one-sided suppression of negative input values (single-ended saturation, and the output is 0 when negative values are input) and wide excitation boundary (linear mapping for positive input), so the nonlinear mapping is sparsity. According to equation (6), the calculation amount of ReLU is far less than Sigmoid function, and its first derivative is 1, so it will not cause the gradient disappear. The Softplus activation function is also single-ended saturation, so it converges faster than the Sigmoid function. However, it is only a smooth approximation of ReLU, so it does not have sparse activation, and its activation



(a) L_ReLU function graph



(b) First derivative of L_ReLU function

Fig. 6. L_ReLU function and its first derivative graph

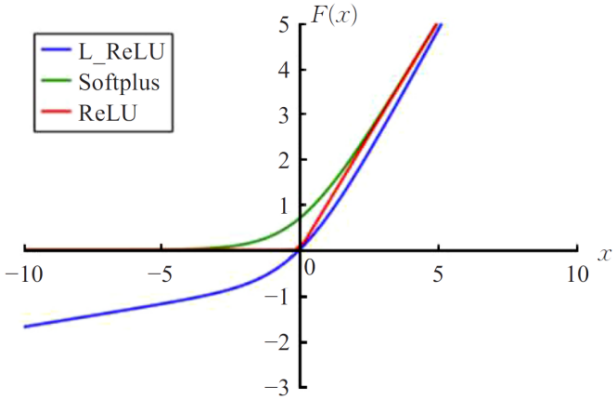


Fig. 7. Three activation function curves

performance is worse than that of ReLU activation function. L_ReLU activation function is double-ended unsaturated, and the gradient will not disappear in the CNN training, so the convergence speed of the network will be faster than the Sigmoid activation function. Moreover, by comparing the L_ReLu in figure 7, it can be seen that L_ReLU is also similar to a smoothing of ReLU, but different from Softplus, and it also has a certain sparse mapping feature, that is, it has an appropriate inhibitory effect on negative input, and an approximate linear mapping for positive input.

In this paper, MNIST data training experiments are carried out on CNNs with the above three activation functions respectively. The convergence of the network is shown in table 1. In table 1, the network convergence speed of L_ReLU and ReLU is higher than that of Softplus, but the convergence speed of L_ReLU is slightly lower than that of ReLU. By comparing equations (4) and (6), it can be seen that the convergence speed of L_ReLU is slightly slower than that of ReLU, because it requires a large amount of calculation.

Table 1. The error of three activation functions in MNIST data training

activation	200	400	600	800	1000	1200	1400
softplus	13.5	13.4	10.1	7.2	6.8	6.1	6.1
L_ReLU	7.7	4.6	3.8	3.7	3.5	3.4	3.4
ReLU	2.8	2.7	2.6	2.5	2.4	2.4	2.4

Although the ReLU has excellent performance, it inputs all negative values into the sparsity map, and two problems are likely to occur in the actual training of CNN: (1) "dead neurons" phenomena appears in the network, that is, the gradient value of the network parameter in this neuron is 0, and it cannot be trained and updated again. (2) it results in the loss of some characteristic information quantized by negative values. By comparing the curves of L_ReLU and ReLU in table 1, it can be seen that L_ReLU realizes the non-zero sparse mapping through the compression of the negative input similar to the

exponential function, which can avoid the above two problems. Therefore, the optimized activation function used in this paper is L_ReLU.

3.2. Structure optimization

It is found that artificial neural networks with appropriate multiple hidden layers can learn the essential features of data and classify data effectively. However, increasing the network depth too much will reduce the performance of the neural network. Therefore, one direction of improving CNN network structure is to properly increase the convolutional layer, while the other direction is to adjust the size of the convolutional kernel. This paper conducts performance testing on CNN with nine different structures, and the results are shown in table 2.

Table 2. CNN performance comparison with nine different structures

No.	Convolutional layer number	Kernel	Training rate	Testing rate
1	2	2×2	1.00	0.89
2	2	3×3	1.00	0.91
3	2	4×4	1.00	0.93
4	3	2×2	1.00	0.84
5	3	3×3	1.00	0.92
6	3	4×4	1.00	0.92
7	4	2×2	0.95	0.61
8	4	3×3	0.98	0.83
9	4	4×4	1.00	0.86

According to table 2, under the condition of the same size of convolution kernel, CNN with three convolution layers has the best accuracy in the testing set. For CNN with the same number of convolutional layers, 3×3 convolutional kernel CNN has better performance. Considering that the performance difference is not large and the scale of CNN is small, the network structure is improved with three convolution layers in this paper. The convolution kernel of each layer is 3×3.

By comparing the accuracy of the network in the training set and the testing set in table 1, it can be seen that although the improvement of network structure has improved the performance of CNN, the problem of over-fitting still exists in the network. The reason for the over-fitting is that the parameters in the fully connection layer are updated completely according to the feature recognition results of the training data. The classification of training data is "overlearned", which leads to the failure to accurately classify the test data. An effective solution is the Dropout technology. Therefore, Dropout technology is adopted in this paper, where the parameter is set as 0.7, that is, in the training process of deep learning network, the parameter is temporarily dropped from the network with a probability of 0.7 without training update, so as to improve the network generalization ability.

In addition to the above improvements, the new CNN model also carries out the following changes.

1. Single channel image input with a high resolution of 68×68 is adopted, so that CNN can extract deeper, high-scale invariant and strong robustness implicit features of the image.
2. During the convolution, zero padding with the value of 1 is carried out first, and a layer of boundary with the value of 0 is added to the input image to enhance the extraction of image edge and contour features.
3. The sampling layer is improved by using 2×2 maximum pooling to enhance sparse expression of features.
4. The output layer is improved by computing the Softmax function as shown in equation (8):

$$p_j = \frac{e^{f_{y_i}}}{\sum_j e^{f_j}}. \tag{8}$$

In equation (8), y_i is defined as the label of the i-th input feature. f_j represents the j-th element of the output vector f in the output layer. P_j represents the probability that the input feature belongs to the j-th class.

3.3. Performance analysis of LeNet-FC model

The proposed LeNet-FC model is shown in figure 8. The image recognition training of LeNet-FC model adopts sparse data labels, that is, the labels (categories) of the original training data set and test data set conducts one-hot coding according to table 3 before training [19]. Then, the Adam optimization algorithm is used to minimize the value of cross entropy loss function, as shown in equation (9):

$$Loss = -\frac{1}{N}\sum_{i}^{N} lb(p_i + \varepsilon), \varepsilon = 1 \times 10^{-10}. \tag{9}$$

In equation (9), N is the number of input samples during each training. p_i is the Softmax function output value corresponding to each sample. The function of ε is to prevent the occurrence of $L = \pm\infty$ when $p_i = 0$ leading to the termination of training.

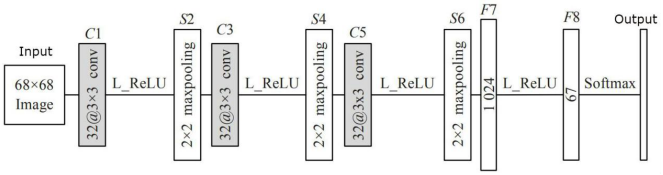


Fig. 8. Proposed LeNet-FC model

This paper conducts experiments on the LeNet-FC model and the LeNet-5 model. The results are shown in figure 9. According to figure 9, the convergence speed of LeNet-FC convolutional neural network is significantly faster than that of CNN based on LeNet-5 model. At the same time, the accuracy change curves of the new LeNet-FC model in the

Table 3. Font sizes

Category label	0	1	2
Coding	$[1,0,0,\cdots,0,0,0]$	[0,1,0,$\cdots$,0,0,0]	[0,0,1,$\cdots$,0,0,0]
$\cdots$	64	65	66
$[0,0,0,\cdots,1,0,0]$	$[0,0,0,\cdots,0,1,0]$	$[0,0,0,\cdots,0,0,1]$	$[0,0,0,\cdots,1,0,0]$

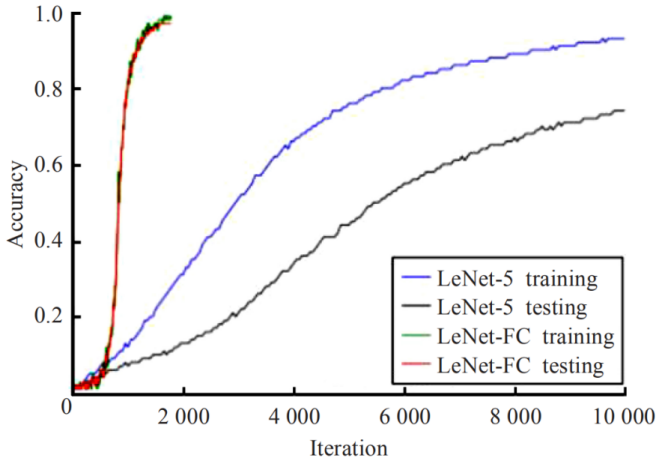


Fig. 9. Training comparison with the two CNN models

training data and test data almost coincide, which indicates that the CNN has a strong ability of image recognition generalization, and there is no over-fitting phenomenon.

As can be seen from table 4, compared with other improved CNN models, the proposed LeNet-FC in this paper has a good performance in image recognition. At the same time, because the other two models are based on AlexNet [20] and ResNet [21] respectively, their models have large size and many parameters, so they need to use hardware with stronger computing performance, such as GPU to perform well. The new model not only has good performance, but also has small scale and low requirement on hardware, and can even be applied to some embedded devices. Therefore, the application of LeNet-FC model is more extensive.

Table 4. Font sizes

Model	Accuracy/%
LeNet-FC	98.9
AlexNet[20]	98.1
ResNet[21]	95.6
DCGAN[22]	92.5
SCNN[23]	94.7
PAMSGAN[24]	95.1

3.4. Experiments and Analysis

The MINIST dataset is from the National Institute of Standards and Technology [25]. The training set is composed of numbers handwritten by 250 different people, some samples of which are shown in figure 10. It contains a total of 70000 images including 60000 training images and 10000 testing images. The images of training set and test set are not repeated. 50% are high school students, and 50% are workers at the Census Bureau. The test set is also handwritten digital data with the same proportion, which is 28×28pixel data set with labels.

Fig. 10. Some samples of MINIST

The CPU parameters of this experiment are Intel(R) Core(TM) i78700 3.20GHz, 8GB memory, Windows10 system, 64-bit operating system. Anaconda is used to simulate the development environment of TensorFlow. In the experiment, the learning rate is set as 1×10^{-4}, and the cross-entropy cost function is used to update the network parameters. During training, dropout is added to avoid over-fitting problems. Batch size=50. An iteration is defined that it traverses all the training sets one time. The initial variable is assigned with a normal distribution with a standard deviation of 0.1. The final prediction results are output after 50 training times in total, and then the model training data is analyzed and verified.

The structure of the image recognition based on Lenet-FC model consists of three parts: image preprocessing, feature extraction and Euclidean distance comparison. The image preprocessing includes the grayscale and normalization, and its purpose is to change the input image X into a single channel image with a resolution of 68×68, so that it is consistent with the input of LeNet-FC model.

Feature extraction is realized by the convolution-sampling layer in LeNet-FC model, and the obtained dimension of feature vector is 17776. The eigenvector of the measured number is denoted as Y. The average feature vector of N standard number images in the same category is denoted as the standard feature Y^S. The calculation formula is shown in equation (10):

$$Y^S = \frac{1}{N} \sum_{i=0}^{N} Y_i^S. \tag{10}$$

Where s stands for different numbers. By adding or deleting features in the library, the number of recognized categories can be changed.

The Euclidean distance comparison mainly calculates the Euclidean distance between the feature vectors to be recognized extracted from the measured number and all the feature vectors in the standard feature library. The expression is $d = \sqrt{(Y - Y^S)^2}$. Then, the

function $f(x) = argmid(d)$ outputs the index T corresponding to the minimum value in vector d, namely, the category. Threshold value η is an important parameter in Euclidean distance. If $d \leq \eta$, then the algorithm will output the result. Otherwise, it needs to re-input the number image.

In the test, the numbers of correct recognition, false recognition and rejection recognition are counted. The corresponding accuracy α, error recognition rate β, and rejection recognition rate δ are calculated [25], and the expressions are expressed as equations (11)-(13) respectively. The test results are shown in table 5.

$$\alpha = \frac{correctly \quad detected \quad and \quad d \leq \eta}{the \quad tested \quad total \quad number}. \tag{11}$$

$$\beta = \frac{error \quad detected \quad and \quad d \leq \eta}{tested \quad total \quad number}. \tag{12}$$

$$\delta = \frac{number \quad of \quad d > \eta}{tested \quad total \quad number}. \tag{13}$$

Table 5. Font sizes

η	Correctly detected number	Error detected number	Rejected detected number	α	β	δ
0.05	130	0	70	0.66	0	0.34
0.06	150	0	50	0.76	0	0.24
0.07	165	0	35	0.84	0	0.16
0.08	177	1	22	0.89	0.01	0.11
0.09	182	1	17	0.92	0.01	0.09
0.10	186	2	12	0.94	0.01	0.06
0.11	190	2	8	0.95	0.01	0.04
0.12	190	6	4	0.95	0.03	0.02
0.13	191	7	2	0.96	0.04	0.01
0.14	191	8	1	0.96	0.04	0.01
0.15	191	8	1	0.96	0.04	0.01

Through the analysis of the test results in table 5, it can be seen that the image recognition based on LeNet-FC model has high recognition accuracy and relatively low error recognition and rejection rate. In order to test and compare the performance of the improved deep neural network, the traditional LeNet-5 structure is used to adjust the output structure, and the same data training set is also used for simulation comparison. In the traditional network model, a convolution layer and a fully connection layer are added, and the dropout method is not added to prevent over-fitting. The simulation results are shown in table 6 and table 7. We also make comparison with other recognition methods. It can be seen from tables 6, 7 that the recognition rates of LeNet-FC and traditional LeNet neural network are 98.6% and 97.8%, respectively. Compared with the traditional LeNet recognition rate, the recognition performance of LeNet-FC is better.

By comparing tables 6, 7, it can be seen that with the increase of iteration number, the accuracy rate is also continuously improved. Finally, the network gradually reaches

Table 6. Structure recognition rate of LeNet-FC and traditional LeNet neural network (%)

Iteration number	LeNet-FC	LeNet	DCGAN[22]	SCNN[23]	PAMSGAN[24]
10	96.8	96.2	94.8	95.7	95.9
20	97.1	96.9	93.1	95.7	94.6
30	98.6	97.3	94.1	95.7	95.5
40	98.6	97.5	91.2	95.6	96.7
50	98.6	97.8	96.6	94.7	95.8

Table 7. Average cross entropy error (%)

LeNet	LeNet-FC
1.42	0.89

the state of convergence. In terms of convergence effect, the convergence effect of the improved neural network structure tends to be stable, and the final recognition rate is 98.6% after 50 iterations. However, in the training of traditional LeNet, the recognition rate increases with the iteration number, presenting an unstable state with relatively large fluctuation of recognition. In the case of 50 iterations, the final recognition rate is 98.6%.

Meanwhile, this experiment also studies the impact of batch size input on the recognition rate. The single Batch is set as 50, 100 and 200, and other conditions in the experiment remain unchanged. The results are shown in table 8. The number of training iterations is 50, and every 5 iterations show the current training recognition accuracy.

Table 8. Test recognition rate of different batch conditions

No. 5	10	15	20	25	
50batches	98.11	98.69	98.89	98.97	99.03
100batches	97.70	98.38	98.68	98.81	98.92
200batches	96.84	97.99	98.38	98.55	98.67
No. 30	35	40	45	50	
50batches	99.12	99.11	99.28	99.22	99.25
100batches	98.88	99.09	99.13	99.09	99.15
200batches	98.79	98.79	98.95	98.97	98.88

As can be seen from table 8, after 35 iterations, the 50 batches, 100 batches can achieve recognition accuracy of more than 90%. With the increasing number of iterations, the change of recognition rate becomes stable, so it can be considered that the network model reaches the convergence state at this time. Where, the fastest convergence rate is at 50batch, and the slowest is at 200batch. After about 45 iterations, the recognition accuracy will fluctuate around 99%, and does not improve in 50 iterations. Overall, the recognition rate of 200batch model is lower that of 50batch and 100batch model in the preliminary data training. After 50 iterations of the 100 batches model, the recognition accuracy fluctuates around 99.1%. In the training process, the training speed of 50 batches is not the fastest.

4. Conclusion

On the basis of LeNet-5 neural network, the structure of the network is improved, which greatly reduces the number of neuron parameters, improves the training time, increases the number of feature extraction layers, and improves the recognition accuracy. The LeNet-FC model shows strong generalization ability in image recognition training. The comparative experiment shows that the proposed method greatly improves the recognition effect. In the future recognition training experiments, appropriate parameters should be selected according to the size of training batch, so as to improve the recognition rate and the training speed as much as possible.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant nos. U20A20197, 61973063, U1713216, 61901098, 61971118, Liaoning Key Research and Development Project 2020JH2/10100040, the China Postdoctoral Science Foundation 2020M670778, the Northeastern University Postdoctoral Research Fund 20200308, the Scientific Research Foundation of Liaoning Provincial Education Department LT2020002, the Foundation of National Key Laboratory (OEIP-O-202005) and the Fundamental Research Fund for the Central Universities of China N2026005, N181602014, N2026004, N2026006, N2026001, N2011001.

Availability of data and materials. The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests. The authors declare that they have no conflicts of interest.

References

1. Maruo S, Fujishiro Y, Furukawa T. "Simple autofocusing method by image processing using transmission images for large-scale two-photon lithography," *Optics Express*, vol. 28, no. 8, 2020.
2. Chen J, Zheng H, Xiong H, et al. "FineFool: A Novel DNN Object Contour Attack on Image Recognition based on the Attention Perturbation Adversarial Technique," *Computers & Security*, vol. 9:102220, 2021.
3. Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. "Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation," *Multimedia Tools and Applications*, vol. 79, pp. 31049-31068, 2020.
4. Khan M A, Rizvi S, Abbas S, et al. "Deep Extreme Learning Machine-Based Optical Character Recognition System for Nastalique Urdu-Like Script Languages," *The Computer Journal*, vol. 65, no. 2, pp. 331-344, 2022.
5. Murata M, Kanamaru T, Shirado T, et al. "Automatic F-term Classification of Japanese Patent Documents Using the k-Nearest Neighborhood Method and the SMART Weighting," *Information & Media Technologies*, vol. 14, no. 1, pp. 163-189, 2007.
6. Xia, B., Han, D., Yin, X., Gao, N. "RICNN: A ResNet & Inception Convolutional Neural Network for Intrusion Detection of Abnormal Traffic," *Computer Science and Information Systems*, vol. 19, no. 1, pp. 309-326, 2022.
7. Gorban A N, Mirkes E M, Tukin I Y. "How deep should be the depth of convolutional neural networks: a backyard dog case study," *Cognitive Computation*, vol. 12, no. 1, pp. 388-397, 2020.

8. Kim M J, Yi L, Song H O, et al. "Automatic Cephalometric Landmark Identification System Based on the Multi-Stage Convolutional Neural Networks with CBCT Combination Images," *Sensors*, vol. 21, no. 2, pp. 505, 2021.
9. X. Yu, W. Long, Y. Li, X. Shi and L. Gao. "Improving the Performance of Convolutional Neural Networks by Fusing Low-Level Features With Different Scales in the Preceding Stage," *IEEE Access*, vol. 9, pp. 70273-70285, 2021.
10. Wen L, Li X, Gao L. "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Computing and Applications*, vol. 32, pp. 6111-6124, 2020.
11. Kg A, Nc A. "Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN - ScienceDirect," *Procedia Computer Science*, vol. 167, pp. 878-889, 2020.
12. nan Güler a, B E B. "Expert systems for time-varying biomedical signals using eigenvector methods," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1045-1058, 2007.
13. Glorot X, Bordes A, Bengio Y. "Deep Sparse Rectifier Neural Networks," *Journal of Machine Learning Research*, vol. 15, pp. 315-323, 2011.
14. Gao S. "A Two-channel Attention Mechanism-based MobileNetV2 And Bidirectional Long Short Memory Network For Multi-modal Dimension Dance Emotion Recognition," *Journal of Applied Science and Engineering*, vol. 26, no. 4, pp. 455-464, 2022.
15. Fan Y, Rui X, Poslad S, et al. "A better way to monitor haze through image based upon the adjusted LeNet-5 CNN model," *Signal Image and Video Processing*, vol. 14, no. 2, 2020.
16. Zhang C, Yue X, Wang R, et al. "Study on Traffic Sign Recognition by Optimized Lenet-5 Algorithm," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 1, pp. 2055003.1-2055003.21, 2020.
17. Zhang Q, Hu X, Zhou S. "The Detection of Hyperthyroidism by the Modified LeNet-5 Network," *Indian Journal of Pharmaceutical Sciences*, vol. 82, 2020.
18. A. Senior and X. Lei. "Fine context, low-rank, softplus deep neural networks for mobile speech recognition," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7644-7648, 2014.
19. F. Jafarzadehpour, A. Sabbagh Molahosseini, A. A. Emrani Zarandi and L. Sousa. "Efficient Modular Adder Designs Based on Thermometer and One-Hot Coding," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 9, pp. 2142-2155, 2019.
20. Sarp, S., Kuzlu, M., Zhao, Y., Cetin, M., Guler, O. "A Comparison of Deep Learning Algorithms on Image Data for Detecting Floodwater on Roadways," *Computer Science and Information Systems*, vol. 19, no. 1, pp. 397-414, 2022.
21. Wu Z, Shen C, Hengel A. "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition," *Pattern Recognition*, vol. 90, pp. 119-133, 2019.
22. L. Sun, K. Liang, Y. Song and Y. Wang. "An Improved CNN-Based Apple Appearance Quality Classification Method With Small Samples," *IEEE Access*, vol. 9, pp. 68054-68065, 2021.
23. M. Zhang, M. Gong, H. He and S. Zhu. "Symmetric All Convolutional Neural-Network-Based Unsupervised Feature Extraction for Hyperspectral Images Classification," *IEEE Transactions on Cybernetics*, vol. 52, no. 5, pp. 2981- 2993, 2022.
24. Z. Zhang. "PAMSGAN: Pyramid Attention Mechanism-Oriented Symmetry Generative Adversarial Network for Motion Image Deblurring," *IEEE Access*, vol. 9, pp. 105131-105143, 2021.
25. S. B. Ahmed, I. A. Hameed, S. Naz, M. I. "Razzak and R. Yusof. Evaluation of Handwritten Urdu Text by Integration of MNIST Dataset Learning Experience," *IEEE Access*, vol. 7, pp. 153566-153578, 2019.
26. Chuang Bai, Xiang Chen. "Research on New LeNet-FC Convolutional Neural Network Model Algorithm," *Computer Engineering and Applications*, vol. 55, no. 5, pp. 105-111, 2019.

Jingsi Zhang is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Xiaosheng Yu is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Xiaoliang Lei is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Chengdong Wu is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Received: January 20, 2022; Accepted: August 29, 2022.