Research on application of an improved deep convolutional neural network in handwritten character recognition

Lan Wu¹, Xueying Jia², Canshi Zhu³

¹XiJing University, Xian, Shaanxi, 710123, China
²Zhongnan University of Economics and Law, Wuhan, Hubei, 430073, China
³XiJing University, Xian, Shaanxi, 710123, China

Corresponding author and e-mail: Lan Wu, 287295865@qq.com

Abstract. The classical convolutional neural network has been widely used for handwritten digit character recognition with high accuracy. However, due to its small convolutional layer, fixed size of convolution kernel and few extracted features, the recognition accuracy of complex handwritten characters are reduced. In this paper, an improved deep convolutional neural network model is proposed, which can allocate different convolution kernels according to the different information amount in the handwritten character image area for convolution, so as to better extract the effective information of the image and is more suitable for complex handwritten character recognition applications. Experiments show that the recognition rate can be higher.

1. Introduction

Image recognition is a key research area in artificial intelligence, involving many aspects such as handwritten character recognition, face recognition, speech recognition, object recognition, etc. Among them, handwritten character recognition is a field that has been studied more and is widely used to automatically read bank checks information, the postal code on the envelope and the recognition of some documents, etc. The recognition rate of handwritten characters directly affects the judgment and next processing of the computer. Good recognition technology will make the computer intelligent and bring convenience to our life. In recent years, deep neural networks have been widely used in computer vision processing, and are very practical in the face of massive data [1]. They have shown the most advanced in many important fields such as handwritten character recognition, face recognition, target detection, speech recognition, machine translation, etc [2]. The performance has therefore rapidly developed and become a mainstream learning method [3]. Among them, the deep convolutional neural network is particularly suitable for processing models with spatial characteristics, such as two-dimensional images, and has become an important technical means in the field of image processing and recognition.

2. Classical convolutional neural network

Convolutional Neural Networks (CNN) is a kind of multi-layer Neural network that can capture the characteristics of the network structure. It is considered as the first deep learning method that adopts the multi-layer hierarchical network successfully [6]. Among them, the convolutional layer is the core layer of the convolutional neural network and has strong characterization capability. It uses convolution operation to extract the features of 1-D and 2-D signals, and uses a unit with the same
weight vector distributed at different positions of the image to obtain the features of the image and form a feature map. Pooling layer is also called downsampling layer. Through local averaging and downsampling operations, the resolution of feature map is reduced and the sensitivity of network output to displacement and deformation is reduced, which is mainly used to compress feature dimension, reduce network scale, reduce overfitting and improve model fault tolerance.

The most classic deep convolutional neural network is LetNet-5 network, which is an earlier CNN model. It has three convolutional layers (C1, C3 and C5), two pooling layers (S2 and S4) and a full connection layer. It attracts the attention of the world because of its high recognition accuracy in handwritten digit recognition. The handwritten digital image of 32*32 pixels is input into letnet-5 network in the format of 32*32 pixels. After the convolution action of multiple alternative convolutional layers and pooling operation of pooling layer, the data features can be obtained. Finally, a multi-dimensional feature vector is output on the full connection layer as the sample classification standard, and the probability of 0~ 90 digits is output. The network model achieved an error rate of less than 1% at that time, and was successfully used to identify the zip code. It can be said that letNet-5 network is the first commercially valuable NN model.

3. Improved deep convolutional neural network

Although classical convolutional neural network can well fit the image topology without image preprocessing and has strong adaptability, but because of it’s small number of convolutional layers, fixed size of convolution kernel and few extracted features it is more applicable to simple images with fewer eigenvalues, such as handwritten digital image recognition. If the handwritten character image that needs to be recognized has a large amount of information and weak categorization, it needs to be further improved.

3.1. Improved deep convolutional neural network structure

The improved deep convolutional neural network is composed of i independent sub-convolutional neural networks (2 ≤ i). The size of i can be determined by the complexity of the picture to be identified. The larger and more complex the information of the picture, the more i can be appropriately increased. The improved deep convolutional neural network structure is shown in Figure 1.

![Improved deep convolutional neural network structure diagram.](image)

When i = n, the improved deep convolutional neural network contains 3n convolutional layers (C1, C2, ..., C3n), 2n pooling layers (S1, S2, ..., S2n), n full connection layers (F1, F2, ..., Fn) and the output layer. Where C3, C6, ..., C3n is connected to the full connection layer.
3.2. Improved deep convolution neural network convolution kernel parameter setting method

The convolutional neural network extracts image information through the convolution kernel of the convolutional layer. Feature extraction is performed on the feature map of the previous layer through convolution operation. Multiple convolution kernels extract different features, so the parameters of the convolution kernel determine the quality of the feature map produced by this convolutional layer. In the classic convolutional neural network LeNet-5, the size of the convolution kernel is $5 \times 5$. Such fixed setting has obvious effect on simple handwritten digital images, but it cannot flexibly and effectively extract features and classify complex images. The information of complex images is generally not uniformly distributed. The improved deep convolutional neural network differentiates the effective information contained in the image with the gray value, sets a threshold according to the gray level, and flexibly sets the convolution kernel size according to the threshold.

Set the image size as $a \times b$, and divide it into $c \times d$ sub-images according to the image gray level. The degree of dispersion of gray level represents the complexity of the image. The more sub-images, the more complex the image, and the more information it contains, namely, the more gray level. Set the gray value $Y$ of each pixel is in the range of $\{0, 1, \ldots, 255\}$, then the probability of each gray level appearing in the image is:

$$ P(Y = k) = P_k, \quad (k = 0, 1, \ldots, 255) \quad (1) $$

"Entropy" is a scale for measuring the degree of uncertainty. "Information entropy" is often used as a quantitative indicator of the information content of a system, which can be used as a goal of system equation optimization or a criterion for parameter selection. The improved deep convolutional neural network uses information entropy to measure the dispersion degree of image gray level.

The information entropy of the gray value $Y$ of each pixel in the picture can be expressed as:

$$ H(Y) = - \sum_{k=1}^{255} P_k \log(P_k) \quad (2) $$

Therefore, the information entropy of the sub-image composed of $Y$ can be expressed as:

$$ H(Y) = \begin{bmatrix} H(Y)_{11} & \cdots & H(Y)_{1c} \\
\vdots & \ddots & \vdots \\
H(Y)_{d1} & \cdots & H(Y)_{dc} \end{bmatrix} \quad (3) $$

The average information entropy of the sub-image composed of $Y$ can be expressed as:

$$ \overline{H(Y)} = \sum_{i=1}^{c} \sum_{j=1}^{d} H_{ij} \quad (4) $$

Use the average information entropy as the threshold to set the size of the convolution kernel

$$ C_n = \begin{cases} C_1 & H_{ij} > \overline{H(Y)} \\
C_2 & H_{ij} = \overline{H(Y)} \\
C_3 & H_{ij} < \overline{H(Y)} \end{cases} \quad (5) $$

Choosing different convolution kernels according to the entropy values of each sub-image of the image above can obtain more effective information of the image, avoid simplification of the convolution kernel to ignore the local information of the image, and have more flexibility.

4. Experimental simulation

In order to evaluate the recognition effect of the improved deep convolutional neural network, the recognition accuracy and loss rate are used as evaluation indicators to measure the accuracy and robustness of the model. The expression of recognition accuracy is as follows:
\[
A_c = \frac{TR}{TR + FA}
\]  

In the formula, \( TR \) represents the number of correctly classified samples and \( FA \) represents the number of misclassified samples.

The loss rate adopts the cross-entropy loss function, which is expressed as follows:

\[
L_Q = -\frac{1}{N} \sum_{p=0}^{N-1} \sum_{q=0}^{M-1} y_{pq} \log H_{pq}
\]  

In the formula, \( N \) represents the number of samples; \( M \) represents the number of categories in the multi-classification task; \( y \) represents the true category of the sample; \( y_{pq} \) represents that the \( p \)-th sample is predicted to be the \( q \)-th category; \( H_{pq} \) represents the probability that the \( p \)-th sample is predicted to be the \( q \)-th category.

Select the handwritten character pictures of the cifar10 data set for experiment. The input picture size is 32 \( \times \) 32, which can compare the LetNet-5 network parameters and the improved deep convolutional neural network parameter values \((n=2, n=3)\), as shown in Table 1.

**Table 1.** Comparison table of two network parameter configurations.

| Types | Convolution kernel size | Number of feature maps | Number of neurons |
|-------|------------------------|------------------------|------------------|
|       |                        | LetNet-5 | Improved \((n=2)\) | Improved \((n=3)\) | LetNet-5 | Improved \((n=2)\) | Improved \((n=3)\) |
| C1    | Convolional layer      | 5 \( \times \) 5 | 6 \( \times \) 6 | 8 \( \times \) 8 | 6   | 16   | 32   | 4704 | 4032 | 6096 |
| S1    | Pooling layer          | 2 \( \times \) 2 | 2 \( \times \) 2 | 4 \( \times \) 4 | 6   | 16   | 32   | 1176 | 1008 | 4024 |
| C2    | Convolional layer      | 5 \( \times \) 5 | 5 \( \times \) 5 | 5 \( \times \) 5 | 16  | 32   | 64   | 1600 | 480  | 612  |
| S2    | Pooling layer          | 2 \( \times \) 2 | 2 \( \times \) 2 | 2 \( \times \) 2 | 16  | 32   | 64   | 400  | 256  | 348  |
| C3    | Convolional layer      | 5 \( \times \) 5 | 4 \( \times \) 4 | 4 \( \times \) 4 | 120 | 148  | 186  | 120  | 128  | 120  |
| F     | Fully connected layer  | -      | -               | -               | -   | 84   | -    | -    | -    | -    |
| C4    | Convolional layer      | -      | 5 \( \times \) 5 | 6 \( \times \) 6 | -   | 16   | 32   | -    | 16128| 16824|
| S3    | Pooling layer          | -      | 2 \( \times \) 2 | 3 \( \times \) 3 | -   | 16   | 32   | -    | 4032 | 6096 |
| C5    | Convolional layer      | -      | 4 \( \times \) 4 | 5 \( \times \) 5 | -   | 32   | 64   | -    | 1536 | 1804 |
| S4    | Pooling layer          | -      | 2 \( \times \) 2 | 4 \( \times \) 4 | -   | 32   | 64   | -    | 384  | 512  |
| C6    | Convolional layer      | -      | 3 \( \times \) 3 | 3 \( \times \) 3 | -   | 148  | 186  | -    | 124  | 256  |
| F     | Fully connected layer  | -      | -               | -               | -   | -    | -    | -    | 376  | -    |
| C7    | Convolional layer      | -      | 4 \( \times \) 4 | -               | -   | 32   | -    | -    | 14032|
| S5    | Pooling layer          | -      | -               | 2 \( \times \) 2 | -   | -    | 32   | -    | -    | 1008 |
| C8    | Convolional layer      | -      | 3 \( \times \) 3 | -               | 64  | -    | -    | -    | 480  | -    |
| S6    | Pooling layer          | -      | 2 \( \times \) 2 | -               | 64  | -    | 256  | -    | -    | 124  |
| C9    | Convolional layer      | -      | 2 \( \times \) 2 | -               | 186 | -    | -    | -    | -    | 472  |

It can be seen from the table that the classic LetNet-5 network has a total of six layers except for the input layer and the output layer, including three convolutional layers, two pooling layers, and a fully connected layer; When \( n=2 \), the improved deep convolutional neural network has two sub-convolutional networks. In addition to the input layer and the output layer, there are 12 layers in total, including six convolutional layers, four pooling layers, and two fully connected layers; When \( n=3 \), the improved deep convolutional neural network has three sub-convolutional networks. In addition to the input layer and the output layer, there are a total of 18 layers, of which nine convolutional layers, six pooling layers and three fully connected layers; The size of the classic LetNet-5 network convolution kernel is fixed at 5 \( \times \) 5, the maximum number of feature maps is 120, and the improved deep...
The evaluation indicators of the improved deep neural network and LetNet-5 are shown in Table 2.

| network          | recognition accuracy (Ac) | loss rate (L_o) |
|------------------|---------------------------|-----------------|
| LetNet-5         | 86.62%                    | 0.127           |
| Improved(n=2)    | 92.43%                    | 0.098           |
| Improved(n=3)    | 94.01%                    | 0.087           |

The experiment shows that the recognition accuracy of the n=2 improved deep convolutional neural network can reach 92.43% and the loss rate is 0.098; the recognition accuracy of the n =3 deep convolutional neural network can reach 94.01% and the loss rate is 0.087; while the highest recognition accuracy of the classical LetNet-5 network is 86.62% and the loss rate is 0.127.Under the of the same scale training conditions, the improved deep convolutional neural network has higher recognition accuracy and lower loss rate than the classical LetNet-5 network. At the same time, the more subimages, the higher the recognition rate, but the network is relatively complex and the amount of calculation is high.

5. Conclusions

This study presents an improved deep convolutional neural network. Compared with the classic LetNet-5 network, the improved network can allocate different convolution kernels to perform convolution according to the different amount of information in the image area, which can better extract the effective information of the image. This network solves the problem of letNet-5 network with fewer convolutional layers, fixed convolution kernels and fewer extracted features. Experiments show that this network has a higher recognition rate and a lower loss rate than the LetNet-5 network.
At the same time, the more the number of sub-images, the higher the recognition rate, so it is more suitable for the recognition of complex handwritten symbols.

References

[1] Ramirezcortes J M, Gomezgil P, Alarconaquino V, et al. A Biometric System Based on Neural Networks and SVM Using Morphological Feature Extraction from Hand-Shape Images[J]. Informatica, 2011, 22(2):225-240.

[2] A.K. Sampath, N. Gomathi. Handwritten optical character recognition by hybrid neural network training algorithm [J]. The Imaging Science Journal, 2019, 67(7).

[3] Fitehalew Ashagrie Demilew, Boran Sekeroglu, Ancient Geez script recognition using deep learning[J]. SN Applied Sciences, 2019, Vol.1 (11), pp.1-7

[4] Theljani F, Laabidi K, Zidi S, et al. A New Kernel-Based Classification Algorithm for Systems Monitoring: Comparison with Statistical Process Control Methods [J]. Arabian Journal for Science and Engineering, 2015, 40(2):645-658.

[5] Dan Cire,D.Meier U.Schmidhuber J. Multi-column deep neural networks for image classification [C]/ / Providence, USA: 2012 IEEE Conference on Computer Vision and Pattern Recognition,2012: 3642—3649.

[6] Ourabloo A, Liu X. Pose-Invariant Face Alignment via CNN-Based Dense 3D Model Fitting [J]. International Journal of Computer Vision, 2017, 124(2):187-203.

[7] Wafaa Wardah, Abdollah Dehzangi, Ghazaleh Taherzadeh, Mahmood A. Rashid, M.G.M. Khan, Tatsuhiko Tsunoda, Alok Sharma. Predicting protein-peptide binding sites with a deep convolutional neural network [J]. Journal of Theoretical Biology, 2020,496.

[8] RUSSAKOVSKY O, DENG J, SU H, et al. Image Net large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211—252.