Intelligent Computing Technology and Automation
Z. Hou (Ed.)

doi:10.3233/ATDE231286

Handwritten English Character Recognition Method Based on Deep Learning

Panpan DU[a,1], Yan LI[b]
[a] *Department of Foreign Languages, Gingko College of Hospitality Management, Chengdu, 611743, China*
[b] *Sichuan University of Media and Communications, Chengdu, 611745, China*

Abstract. In order to solve the problem of low accuracy of computer intelligent recognition of handwritten English in practical application, a method of font feature extraction and recognition using deep learning is proposed. Considering the serious differences of off-line handwriting styles of different people, the scheme uses the preprocessed image data to train the improved CNN model, reduces the interdependence of parameters. Then the improved LeNet-5 model is applied to the research of handwritten numeral recognition. The simulation results show that our method can effectively realize the automatic recognition of handwritten English, and its character recognition accuracy can reach more than 95%.

Keywords. character recognition; CNN; LeNet-5; handwriting; deep learning

1. Introduction

Character recognition technology is to realize the intelligence of information recognition. Through intelligent means, the docking and exchange of information can be completed smoothly. Therefore, the research direction of character recognition has great development prospects in information intelligence. At present, convolutional neural network (CNN), support vector machine (SVM), deep belief network (DBN) and k-nearest neighbor algorithm are commonly used character recognition methods [1]. Under the background of the continuous development of computer technology and artificial intelligence application, it is of great significance to study the vision method based on deep learning. However, handwritten English fonts include the writer's personal writing style and writing habits, which makes font recognition difficult to a certain extent. Traditional image recognition methods cannot analyze the attributes such as the writer of characters, so it is necessary to use deep learning algorithm to further study the high-level problems such as the writing style of characters.

Aiming at the recognition of handwritten characters, especially handwritten English, this paper constructs convolution neural networks with different depths by using deep learning technology, and then proposes a character image recognition algorithm based on convolution neural network. Based on the optimization principle of convolution neural network, a temporary output layer is added after the convolution layer of lenet-5 convolution neural network, which makes the network fit the processed image features better than the real label. We also set the corresponding calculation parameters for each character to obtain the output vector of the whole network. The

[1] Corresponding Author: Panpan DU; Department of Foreign Languages, Gingko College of Hospitality Management, Chengdu, 611743, China; panpan_du1027@163.com

simulation results show that the recognition accuracy of the deep learning network structure is significantly improved compared with the traditional deep learning network.

2. Deep Learning Theory

2.1 Principle of Deep Learning

Deep learning is a branch of machine learning: it is a new method of learning representation from data, which emphasizes learning from continuous layers, which correspond to more and more meaningful representations. "Depth" in "deep learning" refers not to the deeper understanding obtained by this method, but to a series of continuous presentation layers. Other names in this field include hierarchical representations learning and hierarchical representations learning [2]. In the neural network, the specific operation of each layer on the input data is saved in the weight of the layer. Its essence is a string of numbers. The transformation realized by each layer is parameterized by its weight, as shown in figure 1.

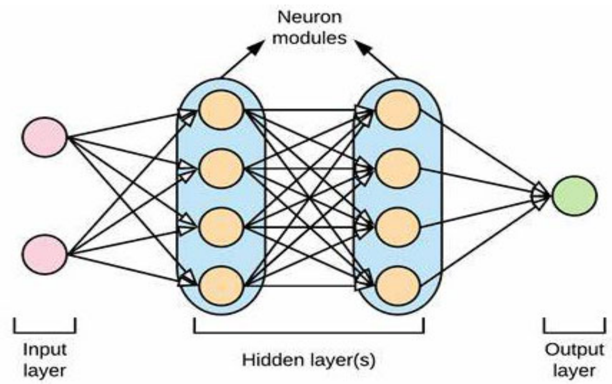

Figure 1. Typical depth network model structure.

In deep learning, the activation value of the $(l+1)_{th}$ layer is computed as:

$$z^{(l+1)} = W^{(l)}a^{(l)} + b^{(l)} \quad (1)$$

$$a^{(l+1)} = f(z^{(l+1)}) \quad (2)$$

where l is the layer number; $z_i^{(l)}$ is the weighted sum of the i_{th} unit at l layer; $a_i^{(l)}$ is the activation value of the i_{th} unit at l layer.

2.2 Convolutional Neural Network

Convolutional neural network(CNN) is a variant of multilayer perceptron (MLP). Its advantages are more obvious when the input of the network is an image, so that the image can be directly used as the input of the network, avoiding the complex process of feature extraction and data reconstruction in the traditional recognition algorithm [3,4]. It has great advantages in the process of two-dimensional image processing. For example, the network can extract the features of the image by itself, including color, texture, shape and image topology, especially in the problem of two-dimensional image processing, especially in identifying displacement [5-8]. It has better robustness and computational efficiency in the application of scaling and other forms of distortion invariance. The general CNN models are summarized as shown in table 1.

Table 1. Typical CNN models

Model	Feature
LeNet-5	The architecture is very simple
AlexNet	Relu and Dropout are introduced. Data enhancement and pooling cover each other. One maximum pooling + three full connection layers are introduced
VGGNet	The convolution kernels of 11 and 33 and the maximum pooling of 2 * 2 make the number of layers deeper
Google Inception Net	While controlling the amount of calculation and parameters, a better classification performance is obtained
Microsoft RESNET residual	The introduction of highway structure can make the neural network very deep

3. English Character Recognition Based on Deep Learning

3.1 English Character Recognition Process

The character recognition process mainly includes three modules: target detection, character recognition and matching judgment. The target detection module uses CNN to detect and locate different pixels in the image. The character recognition module recognizes the positioned character area [9]. The main problem to be solved is what each character is, and converts the character area in the image into character information. The matching judgment module compares the input data with the fragments in the library one by one, and then outputs the final result in the form of template. The process is described in figure 2.

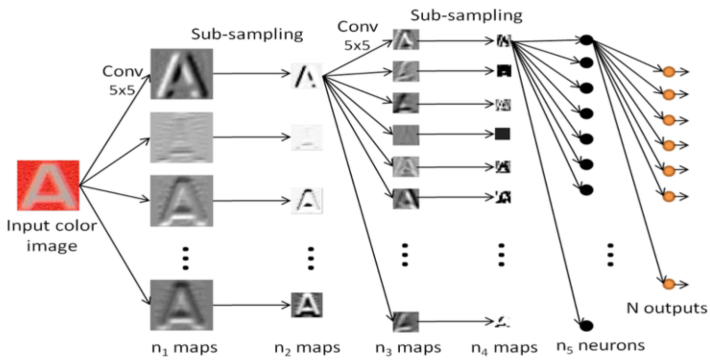

Figure 2. English character recognition process based on CNN.

3.2 Data Preprocessing

(1) Image binaryzation

Set the gray value of the pixels on the image to 0 or 255, that is, the whole image will show an obvious visual effect of only black and white. After image binary segmentation according to the threshold obtained by Otsu method, the inter class variance between foreground and background image is the largest [10]. It is suitable for processing the image with large difference between foreground image and background image The original test paper samples are directly converted into gray images through the Changegray() function in Python. Count the number of pixels with the gray value corresponding to each gray value and the specific formula is depicted as follows:

$$f(i,j)=\begin{cases}0, f(i,j)<k\\1,\ f(i,j)\geq k\end{cases} \tag{3}$$

Assuming k is known binarization threshold, when $f(i,j)<k$, the pixel points of image are unified as 0; when $f(i,j)\geq k$, the pixel points of image are unified as 1. Then the final results are binary image.

(2) Word segmentation

We use a pre-training font provided by Halcon to read the numbers in the figure [11]. Use create_ text_ model_ Reader creates the model and sets the parameter mode to 'auto'. Here, OCR classifier parameters must be passed. We can then use *set_text_model_Param* to specifie the segmentation parameters, and use *get_ text_ model_ Param query*. When finished, you can use find_ Textread text. The operator selects candidate characters according to the characteristics of region and gray value, and verifies them with a given OCR classifier. Part of the main codes of such process are described as follows:

```
threshold (Green, ForegroundRaw, 0, 220)
sub_image (RedReduced, GreenReduced, ImageSub, 2, 128)
mean_image (ImageSub, ImageMean, 3, 3)
binary_threshold (ImageMean, Cluster1, 'smooth_histo', 'dark', UsedThreshold)
difference (Foreground, Cluster1, Cluster2)
concat_obj (Cluster1, Cluster2, Cluster)
opening_circle (Cluster, Opening, 2.5)
read_ocr_class_mlp ('Industrial_0-9_NoRej', OCRHandle)
do_ocr_multi_class_mlp (FinalNumbers, ImageOCR, OCRHandle, RecChar, Confidence)
clear_ocr_class_mlp (OCRHandle)
…
```

3.3 Improvement of CNN Model

The model adopted in this paper is improved based on the network structure of Lenet5. The network structure of the improved Le net5 is shown in figure 3. In this paper, all neurons are used, but their outputs are multiplied by 0.5, which is a reasonable approximation to the geometric mean of the prediction distribution generated by multi exponential dropout network Layer C1, C3 and C5 are convolution layers. 6 convolution cores are used in this layer, and the size of each convolution core is 5×5. In this way, 6 feature maps are acquired; There is no overlap between pooling units, and new eigenvalues are obtained after aggregation statistics in the pooling area [12]; Change all activation functions from function Tanh to Prelu to reduce the computational complexity, that is, $y=\frac{e^x-e^{-x}}{e^x+e^{-x}} \to y=\max(0,x)$.

Figure 3. Improved LeNet-5 model structure

The recognition for English charactres include dataset containing m samples: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$, where the category is $y^{(i)} \in \{1, 2, \ldots, k\}$, $k = 26$. Since the output has 26 tags, supposing the output has 26 value for each x, that is, the formation of $h_\theta(x^{(i)})$ is:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 \mid x^{(i)}; \theta) \\ p(y^{(i)} = 2 \mid x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k \mid x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_{(i)}}} \begin{bmatrix} e^{\theta_1^T x_{(i)}} \\ e^{\theta_2^T x_{(i)}} \\ \vdots \\ e^{\theta_k^T x_{(i)}} \end{bmatrix} \tag{3}$$

where θ_1, $\theta_2, \ldots,$ $\theta_k \in R^{n+1}$ are model parameters. According to the training of tagged data, the parameters of all the network can be optimized. Then they are adopted to classify the untagged images.

4. Simulations

The data sets used in this paper include IAM English handwriting public set and collected student sample real handwriting set. The selection data were from 248 different authors, including 104650 handwritten texts. The sample data is randomly divided into two parts in the simplest way of data set division in advance. Then use the training set to train the model, and select the model and parameters on the test set. The recognition results of some data by the model implemented in this paper are shown in figure 4. We will load IAM dataset using fetch in sklearn fetchdata and assign images and labels to X and Y variables. Then a training/testing device will be created to test and analyze each number and letter respectively. With the deepening of the number of network layers, the parameters to be learned will also increase, which will more easily lead to over fitting. When the data set is small, too many parameters will fit all the characteristics of the data set rather than the commonness between the data. Therefore, in this case, in order to prevent over fitting, data enhancement came into being. In addition to data enhancement, there are regular terms/dropout and other methods to prevent over fitting. The results are represented by confusion matrix. It is found that the recognition effects of various numbers and letters are very different, as shown in table 2.

It can be seen that for characters with more strokes and special fonts, the error rate of handwriting recognition is high, indicating that the ability of the algorithm for edge detection needs to be improved.

Figure 4. Partial recognition effect on IAM dataset

Table 2. Part of the letter confusion matrix

Test character	Recognized as a	Recognized as b	Recognized as c	Recognized as d	Recognized as e	…	accuracy
a	**250**	5	0	1	0	...	
b	0	**295**	1	0	0	...	
c	1	0	**289**	0	1	...	
d	0	0	0	**300**	2	...	
e	0	0	0	0	**289**	...	
…	...	...	...	...	...	...	...

At the same time, several traditional positioning methods are used to compare the test paper text. The experimental results are shown in table 3. It can be concluded that compared with the classical East, PSENET, CPTN, BP neural network and RESNET, the improved algorithm in this paper has good performance in accuracy, recall and F-value index [13]. If higher accuracy is required, we just need to increase the sample size and continue to fine tune the newly generated model according to the above steps. Moreover, the training of the improved LeNet5 model will be fast, and multiple epochs are not required. Since we have a model with an accuracy of 90%, we can use the prediction results of the model to request the characters to be recognized. More samples can be taken in time, which can also ensure the effectiveness of the detection rate.

Table 3. Comprehensive performance comparison of different model

Method	Accuracy(%)	Recall(%)	F-value(%)
EAST	67	81.5	76.2
CPTN	80	80	71.1
PSENet	72	87.5	77.5
BP	82	93.1	89.1
Improved LeNet5	85	96	90.2

5. Conclusion

This paper uses the feature extraction algorithm of deep learning to recognize handwritten English characters. Using the trained letnet-5 model, the input handwritten English character pictures are output as the extracted feature vector. The recognition layer is improved on the basis of LeNet-5. The softmax layer uses one-dimensional features to obtain the probability of belonging to each category. Then the whole flow of the algorithm is given through pytorch, and tested and analyzed in the general font library. The experimental results show that the overall recognition error rate of the deep learning English recognition scheme proposed in this paper is low, and has better comprehensive performance compared with similar algorithms.

References

[1] Li C. Research on Character Recognition Based on Depth Learning. Techniques of Automation and Applications, 2018, 37(11): 120-125
[2] Ren B, Wang L T, Deng X, et al. An Improved Deep Learning Network Structure for English Character Recognition. Journal of Chengdu University of Information Technology, 2017, 32(3):259-263
[3] Harizi R, Walha R, Drira F, et al. Convolutional neural network with joint stepwise character/word modeling based system for scene text recognition. Multimedia Tools and Applications, 2021(12):1-16
[4] Ye Z, Dai F, Jin X, et al. The recognition and implementation of handwritten character based on deep learning. Proceedings of International Conference on Artificial Life and Robotics, 2019, 24:276-279
[5] Wang T, Xie Z, Li Z, et al. Radical aggregation network for few-shot offline handwritten Chinese character recognition. Pattern recognition letters, 2019, 125(7):821-827
[6] Deore S P, Pravin A. Devanagari Handwritten Character Recognition using fine-tuned Deep Convolutional Neural Network on trivial dataset. Sādhanā, 2020, 45(1):1-13
[7] Habib K N, Awais A. Urdu Optical Character Recognition Systems: Present Contributions and Future Directions. IEEE Access, 2018, 6:46019-46046
[8] Demilew F A, Sekeroglu B. Ancient Geez script recognition using deep learning. SN Applied Sciences, 2019, 1(11):1-3
[9] Li Z, Y Xiao, Wu Q, et al. Deep template matching for offline handwritten Chinese character recognition. The Journal of Engineering, 2020, 4:120-124
[10] Ananth E. Handwritten Text Recognition using Deep Learning and Word Beam Search. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, 12(2):2905-2911
[11] Li Z, Wu Q, Xiao Y, et al. Deep Matching Network for Handwritten Chinese Character Recognition. Pattern Recognition, 2020, 107:107471
[12] Mushtaq F, MM Misgar, Kumar M, et al. UrduDeepNet: offline handwritten Urdu character recognition using deep neural network. Neural Computing and Applications, 2021, 1:1-24
[13] Ma X, Xu H, Zhang X, et al. An Improved Deep Learning Network Structure for Multitask Text Implication Translation Character Recognition. Complexity, 2021, 1:1-11.