Adaptive preprocessing of character recognition image based on neural network

Fang Wang*  
School of Information Engineering, Tianjin College, University of Science and Technology Beijing, 301830, China  
*Corresponding author: morningwf@163.com

Abstract. Words in natural scenes contain abundant information. Automatic acquisition of text information in images can help people understand images more effectively and further process images such as storage, compression and retrieval. Pattern recognition technology is to use machines to simulate people's various recognition abilities. At present, it mainly simulates people's visual and auditory abilities. Based on BP neural network in deep learning, a text positioning BP neural network is proposed, which automatically extracts the text features in the image and avoids the defect of using manual design features. This paper studies the preprocessing algorithm of Chinese character recognition in natural scenes, including binarization of text regions, extraction of character colors, clustering analysis of colors, segmentation of characters, etc., and designs the rules for merging text regions after positioning. Experimental results show that the algorithm can solve the problem of insufficient recognition ability for text images and long-sequence text images with complex background noise between characters when the background and text gray level are similar.

Keywords: BP neural network; Character recognition; Natural scene image

1. Introduction

Text, as a product of human communication, is highly abstract, and at the same time, it enables people to understand the higher-level semantics it expresses. As the visual basis of human perception of the world, image is the main carrier of information transmission. It is precisely because of the relevant information that images bring to people that image recognition technology has made great progress with the rapid development of computer technology and multimedia technology.

Traditional image recognition technology [1], many of which are based on large-scale computation, has irreconcilable contradiction between computational complexity and computational accuracy. After Zhou Tianyi [2] introduced spread spectrum technology, the robustness of watermarking algorithm was obviously improved compared with spatial algorithm. Detailed solutions and experimental verification are given for anti-noise addition of watermark, low-pass filtering, JPEG compression and various algorithms. Hamouda et al. [3] constructed a fast text locator, which is actually a cascade Adaboost

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
classifier. Each weak classifier in the cascade Adaboost classifier is trained by using a set of features, which include gradient histogram, vertical difference, horizontal difference, etc. [4-5]. In addition to solving the traditional problems of large number of pattern recognition and complex distortion invariance recognition, the challenges faced by image recognition begin to be more complex and emotional recognition, such as athlete training and posture correction, recognition of people by walking posture, recognition of people's mouth shape changes and even facial expressions, etc.

An important characteristic of neural network is its fuzzy operation and learning ability. By automatically adjusting its connection weight, the desired input-output mapping is finally obtained. Therefore, in this paper, BP neural network is used for adaptive preprocessing of character recognition images.

2. Neural network classifier

Artificial neural network has achieved good results in handwritten Chinese characters and character recognition because of its good ability of self-learning, self-organizing and imitating human brain intelligence. Among them, BP neural network is one of the most studied and commonly used neural networks. Artificial neural network is a theoretical abstraction, simplification and simulation of the structure, function and some basic characteristics of the physiological real human brain biological neuron network. In fact, it is a complex information processing system, which is an adaptive nonlinear dynamic system composed of a large number of neurons through extremely rich and perfect connections.

The earliest embryonic form of deep learning appeared in cybernetics. The earliest deep learning is a simple artificial neural network model from the perspective of neuroscience. These models are designed to use a group of \( n \) inputs \( x_1, \ldots, x_n \) and associate them with an output \( y \), while the model hopes to learn a group of weights \( w_1, \ldots, w_n \) and complete the calculation according to formula (1) [6].

\[
f(x, w) = x_1 w_1 + \cdots + x_n w_n \quad (1)
\]

BP neural network is a typical feedforward neural network. The transfer function of neurons is S-shaped, and the output is continuous from 0 to 1. It can realize arbitrary nonlinear mapping from input to output. Because the weight adjustment adopts back propagation learning algorithm, it is often called BP network [7-8]. The basic structure of its network is shown in Figure 1, which includes input layer, hidden layer and output layer. The hidden layer can be one or more layers, and the neurons on each layer are called nodes or cells.

![Fig.1 Structure diagram of BP neural network](image)

Distance classifier and neural network classifier are different in principle and have their own advantages in performance, but generally speaking, each single classifier cannot achieve high recognition rate and reliability. In this paper, starting from the application of measurement layer, a set of voting rules is designed. The voting rules fully consider the output results of samples. Considering the result state of samples and the weights of classifiers comprehensively, the classifier's confidence about the samples is obtained, and the final recognition results are obtained by integrating the outputs.

As shown in fig. 2, after linearly normalizing a single character, the obtained character is outlined to obtain the outline of the character, and then the directional line element features in the outline are extracted, and the features are sent to two categories for recognition, and then processed in parallel to
obtain the recognition results of two groups of candidates, which are sent to the voting algorithm module for voting operation and output to obtain the final character recognition results.

![Block diagram of multi-classifier fusion word recognition algorithm based on directional line element features](image)

**Fig.2** Block diagram of multi-classifier fusion word recognition algorithm based on directional line element features

### 3. Image recognition technology based on BP neural network

#### 3.1 Extraction of image feature vector

Feature extraction is one of the key problems in pattern recognition research. For image recognition, extracting effective image features is the primary task of image recognition, and its effectiveness is very important for subsequent processing, so how to find suitable feature vectors has become the core problem of neural network application [9]. In the process of watermark detection, it is necessary to filter the watermark image after geometric transformation, and then extract the watermark information from the watermark image, which requires the original image. According to the color information and spatial position information of characters. In order to get the color of characters, it is necessary to get the exact position of characters in natural scene images.

If the number of hidden layer units is too small, the network may not be trained, or the accurate characteristics of the samples may not be extracted, and the samples that have not been trained before cannot be identified, and the fault tolerance is poor; However, the number of hidden layer units is too large, which leads to too large network scale and complex structure, thus increasing the training time of the network, and sometimes even not converging. In addition, the feature space will be divided too finely, and the decision surface of the network will only contain training samples, so that the network has no generalization ability, and the recognition rate of samples other than training samples may decrease. The following empirical formula [10] can be referred to in practical application:

\[
L = \sqrt{M + N} + a \quad (2)
\]

\[
L = \log_2 N \quad (3)
\]

In which \(M\) is the number of neurons in the output layer, \(N\) is the number of neurons in the input layer, and \(a\) is a constant between 1 and 10.

Selecting some of the most effective features from a group of features to reduce the dimension of feature space is called feature selection. The normalized prediction frame information is restored to the prediction frame information matching the size of the original input image according to the size of the original input image, and it is judged whether the confidence information in the prediction information is greater than the set threshold, if so, the information is retained, otherwise, it is discarded. At this time,
it is necessary to judge whether the binarized result needs to be reversed, in order to correctly get the processing target with white background and black characters. Therefore, in general, we should first consider increasing the number of neurons in the hidden layer, and when a single hidden layer can not meet the requirements, we can consider increasing the number of hidden layers.

3.2 Character extraction

When merging text positioning areas, it is necessary to use the color information of characters, and judge whether to merge according to the color information and spatial position information of characters. In order to get the color of characters, it is necessary to get the exact position of characters in natural scene images. Therefore, we consider a method of extracting image features. When extracting feature vectors, we first cut out a small image of appropriate size from the defective area of the defective image, and then cut out a small image of appropriate size from the non-defective image.

First, the color image is transformed into gray image by formula (4), and then it is transformed into black and white two-pixel image by OTSU dynamic processing.

\[
i_{\text{Grey}} = 0.29 \times r + 0.587 \times g + 0.144 \times b
\]  

OTSU method uses an exhaustive search method, and uses gray histogram to determine the image segmentation threshold dynamically with the maximum variance between target and background. This method has achieved good performance in practical application.

Images in natural scenes are generally color images composed of RGB colors, which contain abundant information such as colors and edges. However, it should be noted that it is difficult to select too many neurons in the hidden layer, otherwise the recognition rate will drop sharply and the anti-noise ability of the network will be reduced [11]. In a week of positioning frame, there will be a range of pixel number ratio between character color and background color. Set a threshold value, and count the number of white pixels and black pixels in a week of the original positioning frame. When the low-pass filtered noise image in the watermark template is superimposed with the original image, the watermark template can get better invisibility under the action of visual masking effect because of the local similar gray level, so the watermark information can be well hidden.

We first assume that there is a square error \( E \) between the result \( x^L \) obtained by the whole deep neural network after forward propagation and the actual result \( y \), as shown in formula (5), and the output of the \( l \)-layer in the deep neural network can be expressed as formula (6).

\[
E = \frac{1}{2} \| y - x^L \|^2
\]

\[
x^t = f(u^t) = f(x^{t-1}w^t + b^t)
\]

Then we calculate the error \( \delta^L \) produced by the last layer of neural network, which can be obtained according to formula (7). Now, we can calculate the error \( \delta^l \) generated by the \( l \)-layer neural network according to equations (6), (8) and (9).

\[
\delta^L = \frac{\partial E}{\partial u^L} = \frac{\partial E}{\partial x^L} \oplus \frac{\partial x^L}{\partial u^L} = f'(u^L) \oplus |y - x^L|
\]

\[
\frac{\partial w^{t+1}}{\partial x^L} = \frac{\partial (w^{t+1}x^L + b^{t+1})}{\partial x^L} = w^{t+1}
\]

\[
\delta^l = \frac{\partial E}{\partial u^l} - \frac{\partial E}{\partial u^{t+1}} \oplus \frac{\partial u^{t+1}}{\partial x^l} \oplus \frac{\partial x^l}{\partial u^l} = [w^{t+1}] \delta^{t+1} \oplus f'(u^l)
\]

Fig. 3 shows the algorithm flow chart of judging foreground and background after binarization of text area.
Grouping the image pixels and dividing the image plane into a series of "meaningful" regions greatly reduces the number of advanced processing stages such as image analysis and recognition, while retaining the information about image structural features. Conceptually, attention mechanism is to select important information from a large amount of information and ignore unimportant information. Or put enough hidden layer units at the beginning, and then according to the changes of different hidden layer units in the learning process, according to the pre-selected principle, gradually delete those hidden layer units with less influence, so as to get the optimal network structure. In the binary image corresponding to each located text area, find the black pixel located in the middle of the image area and find the black pixel connected with it. Then, on the original image, the color of each pixel in the connected domain is averaged as the color of the character.

3.3 Selection of initial weights
Before learning BP network, it is necessary to initialize the network. Because the system is nonlinear, the initial value has a great relationship with whether the learning reaches the local minimum, the length of training time and whether it can converge. When the calculated weight of a certain information is larger, the attention mechanism will focus on the information, and the information will retain more information than other unimportant information. When the calculated weight of a certain information is larger, then the attention mechanism will focus on the information, and compared with other unimportant information, this information retains more information [12]. Using the extracted character colors in this area, the foreground of this area is extracted, so as to remove the influence of shadow, illumination and other factors.

According to the principle that the product of horizontal projection value and vertical projection value is minimum when the tilt angle of character area is minimum, the text area is projected horizontally and vertically to determine the tilt angle of character area. The expected error value should also be determined as an appropriate value after comparison training, because the smaller expected error is obtained by increasing the nodes of hidden layer and training time. Here, initialization is mainly to assign the connection weights of all neurons in BP network model to random values less than 1. Randomization is particularly important in the implementation of many neural network models. If the weights connected in BP network are not randomly initialized, the learning process may not converge. Generally, the initial weights are random numbers between (-1, 1).

4. Experimental results and analysis
In MATLAB7.0, a two-layer network is created by using the function newff. The network has $16 \times 16$ inputs, 1 output neuron, 26 units in the hidden layer, and the number of nodes selected according to experience. When the network does not converge in the learning process, the number of neurons in the hidden layer can be appropriately increased. The excitation function of both layers is "logsig", because
its output range is 0–1, which is just suitable for outputting values after learning. After training, BP neural network can recognize 26 handwritten English capital letters. First of all, when the number of samples (26) is constant, although the number of hidden layer nodes can be determined by $\text{L M N a}+=+2\log \text{LN} =$, the recognition effect of different node numbers is different, as shown in Table 1.

| Number of hidden layer nodes | Training time /s | Training times | Recognition rate/% |
|-----------------------------|------------------|----------------|-------------------|
| 15                          | 2.3071           | 100            | 92.37             |
| 20                          | 2.0217           | 102            | 97.02             |
| 25                          | 2.6270           | 89             | 89.47             |
| 30                          | 3.1834           | 109            | 88.32             |

Using BP neural network, the 50-bit information is embedded into the spatial domain of Lena image, and the related test results are as follows: The bit error rate (BER) is the percentage of bits in the 50-bit information that cannot be recovered correctly. Testing tool is StirMark. For comparison, we selected the template without filtering and visual optimization [Scheme A for short] as reference. The attack type is StirMark's bending attack, the smoothing coefficient of automatic correlation function is $\delta = 0.3$, and the watermark embedding strength is 1.2.

![Fig.4 Comparison of bit error rate and attack strength](image)

It can be seen from fig. 4 that under the same attack intensity, the bit error rate of watermark information in this scheme is much lower than that in scheme a.

In this paper, the ten candidate results identified by the two classifiers are multiplied by the corresponding parameters, and the candidate results of the two classifiers are reordered according to the calculated results to obtain a new candidate result sequence. The following gives the recognition results after integrating Euclidean distance classifier and neural network classifier:

Test set: 241 images in XD_ Text L& R Database, containing 2081 characters of test results.

After text positioning and character segmentation, the number of characters accurately located is 1872; The recognition result after classifier integration: the first candidate number is 1120, the first three candidates are 1183, and the first ten candidates are 1307. The recognition rates of Chinese character modules are 59.83%, 63.19% and 69.82% respectively. As shown in fig. 5, after the integration of classifiers, the recognition performance has been improved to some extent.

![Fig.5 Recognition performance of Euclidean distance classifier, neural network classifier and synthetic classifier](image)
Especially, the first candidate of recognition results is 950 for Euclidean distance classifier, 610 for neural network classifier and 1000 for synthetic classifier, which proves that the method given in this paper is feasible. Especially for the improvement of the preferred results, the results synthesized by classifiers have better performance.

5. Conclusion
In the preprocessing algorithm of this paper, because it is applied after the text positioning algorithm, the core content of this part of preprocessing is the rule design of text region merging after positioning, which is proved to be feasible by practical verification. The coordinate positioning information of the text in the natural scene image is predicted from the feature information output by the text positioning BP neural network method. Then, the features are extracted, and finally, the extracted features are sent to BP network, and the input capital English letters can be judged through the output of the network, so as to realize the recognition of English letters. The recognition result is displayed on the screen, and is also stored in a file and saved. At the same time, the reasons of insufficient experimental results are analyzed and some improvement methods are put forward.

References
[1] Zhang Xinfeng, Yan Kunpeng, Zhao Jun. Research on handwriting recognition technology based on bidirectional LSTM. Journal of Nanjing Normal University (Natural Science Edition), vol. 42, no. 03, pp, 64-70, 2019.
[2] Zhou Tianyi, Zhao Lei. Small sample handwriting recognition based on neural network. Journal of Shandong University of Technology (Natural Science Edition), vol. 33, no. 03, pp, 72-77, 2019.
[3] Hamouda M, Ettabaa K S, Bouhlel M S. Hyperspectral imaging classification based on convolutional neural networks by adaptive sizes of windows and filters. IET Image Processing, vol. 13, no. 2, pp, 392-398, 2018.
[4] Mars. Image processing technology of electronic information based on convolutional neural network. Electronic Production, vol. 000, no. 017, pp, 44-45, 2019.
[5] Xing yanfang, Duan hongxiu, he guangwei. the application of TensorFlow in image recognition system. computer technology and development, vol. 029, no. 005, pp, 192-196, 2019.
[6] Hurtik P, Molek V, Hula J. Data Preprocessing Technique for Neural Networks Based on Image Represented by a Fuzzy Function. IEEE Transactions on Fuzzy Systems, vol. 28, no. 7, pp, 1195-1204, 2020.
[7] Li Zhaotong, Sun Haoyun. Ship detection and identification system based on full convolution neural network. Computer and Modernization, vol. 000, no. 012, pp, 72-77, 2019.
[8] Ko J H, Kim D, Na T, et al. Design and Analysis of a Neural Network Inference Engine Based on Adaptive Weight Compression. vol. 38, no. 1, pp, 109-121, 2019.
[9] Du Lin, An Yuan. Design of handwritten character recognition platform based on fuzzy matching. Information and Computer, vol. 032, no. 002, pp, 101-103, 2020.
[10] Jiang Chongyu, Lu Tongwei, Min Feng, et al. Invoice text detection and recognition method based on neural network. Journal of wuhan institute of technology, no. 06, pp, 586-590, 2019.
[11] Chen Zhen' ang, Huang Zhou, Qin Zhongyuan. Text Verification Code Recognition Scheme Based on Image Processing and Convolutional Neural Network. Cyberspace Security, vol. 11, no. 08, pp, 79-84, 2020.
[12] Ki S, Bae S H, Kim M, et al. Learning-Based Just-Noticeable-Quantization- Distortion Modeling for Perceptual Video Coding. IEEE Transactions on Image Processing, o. 7, pp, 1-1, 2018.
[13] Fang Weiyu, Xu Shangzhi, Li Zhipeng. Automatic recognition algorithm for scanned charts of interlocking tables based on connected domain. Computer Science and Applications, vol. 10, no. 10, pp, 16, 2020.