Research on Artificial Intelligence Machine Learning Character Recognition Algorithm Based on Feature Fusion

Dongdong He *, Yaping Zhang *
Quanzhou University of Information Engineering,Quanzhou,China
*307766723@qq.com
*Corresponding author’s e-mail: dongdonghe@quie.edu.cn

Abstract. With the continuous progress of science and information technology, people begin to study in the field of intelligence, and machine learning is one of the key contents. At present, human beings have made some progress in intelligent robot, speech recognition and network search. The method of character recognition based on machine learning is of great significance to information technology. In this paper, an improved CRNN algorithm based on feature fusion is proposed, which combines Gabor features and Zernike moment features into a new feature vector, and then uses generalized K-L transform to compress the new feature dimension to remove redundant information. After testing, the accuracy of CRNN based on feature fusion on training data set and test data set is as high as 0.99, which shows that the neural network model can perfectly fit the training set of Chinese character recognition.

1. Introduction
The development of science and technology and Internet technology provides the foundation for people to move towards intelligence, while machine learning is a very important component in the field of artificial intelligence, but its application needs further expansion. Among them, deep question answering technology and automatic driving technology, which are widely concerned, are based on artificial intelligence machine learning. In the development process of these fields, it has surpassed the intelligent operation function of human beings to a certain extent. In this case, it is of great significance to actively strengthen the research of character recognition methods based on artificial intelligence machine learning.

In the field of artificial intelligence machine learning and pattern recognition, feature is a very important research object. The problems such as target classification, recognition and image retrieval can be attributed to the process of extracting features from targets and comparing them with prior information by using different similarity definitions [1]. Similarity reflects the relationship between two objects or two features, which is an important index to measure the similarity between samples [2]. At present, the application of deep learning in the technical fields of image classification, recognition and target detection is becoming more mature. Literature [3]
proposed VGG network, Literature [4] proposed GoogleNet, Literature [5] proposed typical deep convolution neural networks. Literature [6] applied deep learning algorithm in the field of character recognition, which caused the whole world to conduct in-depth research on it. The research results provide a new way of thinking for realizing Chinese character recognition and blind reading, and it is a research direction with potential application prospect. Traditional character recognition methods are based on the intuitive morphological characteristics of characters. By statistically analyzing the morphological differences between characters, a set of approximate optimal statistical parameters that can represent the differences of characters can be found to screen and recognize characters, so as to achieve the purpose of computer character recognition and automatic entry and preservation [7]. However, the recognition results are always unsatisfactory, and it is still difficult to obtain a good recognition rate even for English with few character sets. In this paper, a parallel feature fusion technology is proposed for printed Chinese character recognition, and K-L transform is used for feature selection to obtain combined optimized features. This method can not only improve the recognition rate by combining the information contained in two sets of features, but also avoid a sharp decrease in speed.

2. Development and deficiency of character recognition
Character recognition system can be divided into three types according to objects: handwritten character recognition, printed character recognition and world scene character recognition. Among them, handwritten character recognition is the most mature application of character recognition at present. From the domestic open handwriting recognition software test, it has a high recognition rate, and some excellent software even has a similar recognition rate. Recognized as excellent recognition software in the market, the recognition rate of printed text images in common environments is basically kept between. For images with obvious pollution of text images, the recognition rate will drop obviously, but it can still be accepted by people. Word recognition of world scenes refers to the recognition of images with literal meaning from images containing a large number of natural backgrounds, which is the most difficult object in the word recognition system. The reasons can be summarized as follows: for handwriting, the input is usually done in the fixed area of the input terminal, and the input text pixels are binary images. However, the printed text image is a gray image, which will inevitably increase the computational complexity. The background interference of the world scene image far exceeds that of the printed text, which increases the computational burden.

Handwriting recognition represents a clear meaning of words in every input area. However, printed text images need to be segmented by some means in order to obtain areas with clear text meaning. Chinese characters in the world scene are a more generalized expression, and even some characters are used as the background of other characters. This complex and changeable combination makes it difficult for the computer to know which area in the image has a fixed and clear meaning through some algorithm.

In view of the current development status of character recognition system, how to improve the recognition rate of printed characters is still the current research focus, and how to recognize characters in the world scene will be a development direction of character recognition system. In addition, how to build a character recognition system with the characteristics of automatic layout analysis, strong fault tolerance, high recognition rate, error self-learning and self-correction, and easy expansion is the research goal of character recognition automation.
3. Character recognition related technologies

3.1. Principle of character recognition
Character recognition is a part of pattern recognition, which belongs to the problem of super-large category pattern recognition. The essence of character recognition is classification. Chinese character recognition is an interdisciplinary comprehensive technology, including pattern recognition, graphics and image processing, statistics, computer science, artificial intelligence and other disciplines. The basic principle of Chinese character recognition is shown in Figure 1:

![Figure 1. Basic schematic diagram of Chinese character recognition](image)

Firstly, Chinese characters are converted into images that can be stored by computer through optical recording equipment such as scanner or digital camera. Then carry out relevant preprocessing operations to remove noise, analyze the text layout and split it into single text images; After feature extraction, it is sent to the corresponding classifier for classification and recognition; After a certain post-processing work, the original text is sorted out; Finally, the result is output.

3.2. Text image preprocessing technology
In the process of forming a computer-recognizable picture from text information by optical equipment such as scanner or digital camera, it will be interfered by various factors, resulting in the deviation between the image and reality, which will have a certain impact on text recognition. How to deal with these deviations and interferences to the greatest extent is the main work of pretreatment technology. According to different situations, the most common method of binarization is to set an appropriate threshold $T$, and compare the gray values of all pixels in the image with the threshold $T$. All pixels larger than the threshold are determined as the points used to form characters in the image, and are set to 1 (true); Other pixels whose gray value is less than the threshold value $T$ are considered as background, which is set as 0 (false). A certain pixel after binarization is expressed as follows:
The key of binarization lies in the selection of threshold $T$.

3.3. Feature extraction
Recognition algorithm is the core of the whole character recognition system. After preprocessing the original image, a binary image is obtained which is complete and clear enough [8]. After the processing of feature extraction, the information with more Chinese characters has been compressed into features with less dimensions, which can be used as the basis for classification. After classification by classifier, the process of identifying and judging a single character can be completed. Finally, post-processing is carried out according to the context, and the recognition results are output in the initial order, thus completing the whole character recognition process.

Neural network models such as Hopfield neural network, ART network and cognitive model can be used for character recognition. These methods are mainly used in feature extraction and selection, learning and training, classifier design, word recognition post-processing and so on. Compared with statistical methods, neural network is independent of the model, and has the advantage that the output can approach any target in feature space through adjustment. However, the mathematical explanation of neural network is very complex, and the experimental work is very heavy. Combining the neural network method with the traditional recognition method, we can learn from each other's strengths.

4. Character recognition algorithm

4.1. Feature selection
Since the information of each Chinese character is stored in the 32×32 matrix space, it is necessary to calculate the 1024-dimensional vector when calculating the similarity between the two matrices, which is acceptable for single character recognition, but it will take a lot of time to do this calculation for each Chinese character classification. Therefore, in order to shorten the recognition time, this paper adopts the two-level classification method in similarity calculation [9], that is, the binary matrix normalized by the sample to be measured is selected according to the whole row and column units, and the intersection node of the row and column units is taken as the feature unit for similarity calculation. When the similarity result is greater than a specified threshold, the classified Chinese characters are retained, otherwise, they are discarded.

The difference between primary classification and secondary classification is that the feature units for similarity calculation are different, and primary classification is based on local features [10]. The number of selected feature units is relatively small, which can filter out Chinese character classifications with great similarity with the samples to be tested; However, the two-level classification is based on the global features, and the selected feature units are relatively large, so the ability to distinguish similar Chinese characters is stronger.

Over-fitting is a common problem in machine learning and deep learning. If over-fitting occurs, the accuracy of Chinese character recognition generated by training is very low. In order to improve the generalization ability of the network and avoid the phenomenon of over-fitting, this paper preprocesses the training image by means of data augmentation. The data augmentation methods used in this paper include distortion transformation, noise addition and so on. The sine function is

\[
P_a(x, y) = \begin{cases} 
1 & g(x, y) \geq T \\
0 & g(x, y) < T 
\end{cases}
\]
used to translate, rotate and ripple the original data set. In this paper, taking the Chinese characters as an example, the expansion of Chinese characters is realized by changing the parameters such as amplitude and period of the positive selection function.

4.2. Parallel feature fusion
In this paper, firstly, Gabor features and Zernike moment features of Chinese character samples are extracted, then two sets of feature sets in sample space are combined by parallel feature fusion method, and finally, features are extracted in the obtained feature space by K-L transform.
Specific practices are as follows:
Let $ZM, GM$ be Zernike feature and Gabor feature on Chinese character sample space $\Omega$. Because the measurement methods of Zernike moment eigenvector and Gabor eigenvector are different, it is necessary to unify the measurement of these two eigenvectors before using the feature fusion method.
The specific algorithm is as follows:
$$ZM_{\text{norm}} = \frac{ZM - \mu_{ZM}}{\sigma_{ZM}}, GM_{\text{norm}} = \frac{GM - \mu_{GM}}{\sigma_{GM}} \tag{2}$$

In which $\mu$ represents the mean value of training sample eigenvectors and $\sigma$ represents the variance of training sample eigenvectors. For any sample $\xi$, its corresponding two feature vectors are $\alpha \in ZM_{\text{norm}}, \beta \in GM_{\text{norm}}$. Parallel combination features represented by complex vector $\gamma = \alpha + \beta_i$ ($i$ is imaginary unit). Note that if the dimensions of the two groups of features are not equal, the low-dimensional feature vector is complemented by 0.
Combined feature space on sample space $\Omega$ is defined as:
$$C = \{\alpha + \beta_i \mid \alpha \in ZM_{\text{norm}}, \beta \in GM_{\text{norm}}\} \tag{3}$$

Obviously, the space is an $n$-dimensional complex vector space, where
$$n = \max\{\dim ZM_{\text{norm}}, \dim GM_{\text{norm}}\} \tag{4}$$

Because the combined feature is a complex vector, the K-L transformation must be carried out in unitary space. Taking the generation of matrix $S_i$ as an example, it is assumed that the feature vector of $S_i$ is $\xi_1, \cdots, \xi_n$ (orthogonal to each other), and the corresponding feature value is $\lambda_1, \cdots, \lambda_n$, which satisfies $\lambda_1, \cdots, \lambda_n$. The generalized K-L transformation with $S_i$ as the generating matrix is:
$$Y = \Phi^H X, \Phi = (\xi_1, \cdots, \xi_d) \tag{5}$$

In which $(\xi_1, \cdots, \xi_d)$ is the first $d$ eigenvectors. This transform is used to extract combined features. Because the transformed feature vectors are orthogonal vectors, K-L transform can effectively eliminate the correlation between the original feature vectors, remove redundant information and improve the recognition speed.

4.3. Design of CRNN algorithm based on feature fusion
The CNN (Convolutional Neural Network) part of CRNN (Convolutional Recurrent Neural Network) uses a VGG-16-based network structure [11]. In recent years, with the development of
convolutional neural networks, many network structures with deeper layers and more complex structures have appeared [12]. In this paper, based on the original CRNN, the Incision network, which has fewer parameters and stronger feature expression ability, is selected, and the feature image extracted by the Incision network is fused with the feature image extracted by the original convolution layer, and the fused feature image is used as the input of RNN. The network structure diagram of CRNN based on feature fusion is shown in fig. 2.

![Figure 2](image2.png)

**Figure 2.** CRNN network structure diagram based on feature fusion

Because feature images can be fused only when they are the same size, this paper adjusts its network structure based on InceptionV3, so that it has the same output size as CRNN's original feature extraction network. Padding is used in convolution and pooling in the Inception module, so that after passing through the Inception module, the size of feature graph is unchanged, only the number of channels is changed. The structure of the Inception module is shown in Figure 3.

![Figure 3](image3.png)

**Figure 3.** Inception module structure diagram
5. Analysis of experimental results

5.1. Test on a test set
This experiment is completed under the caffe deep learning framework, and the operating system is Ubuntu 16.04; CPU is Intel i7; Memory 32G; 11G; multi-block NVIDIA 1080Ti GPU memory; Python2.7 environment. The experimental data uses script to generate printed Chinese characters with various fonts, and adds random noise to improve the generalization ability of the model. The verification results of CRNN model based on feature fusion after final training on the test set are shown in Table 1:

Table 1. Verification results of network model by test set

| Training set | 145,730 pieces of image data |
|--------------|-----------------------------|
| Precision    | 0.99932                     |
| Sample size  | Number set of 28527 images  |

It can be seen from Table 1 that the accuracy rate of the network model for the whole test set is as high as 99.932%, which shows that the model is feasible to complete the character recognition function.

In order to reconfirm the accuracy of the model, the same method is used to generate test images of different fonts and randomly add noise interference to test the model, and the average accuracy of the model is 97.73%. The test results of the four fonts are shown in Table 2.

Table 2. Test result table of four fonts

| Font name | Song typeface | Boldface | Official script in Chinese | Chinese regular script |
|-----------|---------------|---------|---------------------------|-----------------------|
| Font style | 爱 | 爱 | 爱 | 爱 |
| Accuracy rate | 99.35% | 98.24% | 96.28% | 97.04% |
| Average | 97.73% |

5.2. Comparison of different models
The accuracy of migration learning for Chinese character data sets using Inception-v3, migration learning and fine-tuning for Chinese character data sets using Incision-resnet-v2, and training for data sets from scratch using CRNN based on feature fusion are shown in Table 3.

Table 3. Comparison of accuracy of different models

| Model            | Training mode        | Accuracy |
|------------------|----------------------|----------|
| Inception-v3     | Transfer learning    | 0.8832   |
| Inception-ResNet-v2 | Transfer learning   | 0.9014   |
| Inception-ResNet-v2 | Fine-tune           | 0.93     |
| CRNN based on feature fusion | Abrupt training | 0.99     |

It can be seen from Table 3 that the depth convolution neural networks of Inception-v3 and Incision-resnet-v2 have very strong ability to extract features, but the effect of feature extraction and classification directly using the pre-training model is not as good as that of fine tuning, indicating that the fine-tuned model can better express the features of Chinese character data sets.
However, the results of migration learning and fine-tuning may not fit the neural network perfectly. Moreover, the model of Inception-v3 and Inception-ResNet-v2 is very computationally intensive and takes more time.

In this paper, based on MNIST, a deep convolution neural network is designed, which is more suitable for Chinese character recognition data set. According to the experimental results, the parameters are adjusted and the model is improved, which makes the Chinese character recognition data set fit the neural network model better. Finally, the trained CRNN based on feature fusion has an accuracy of 0.99 on the training data set and 0.99 on the test data, which shows that the neural network model can perfectly fit the Chinese character recognition training set.

6. Summary
With the continuous development of artificial intelligence technology, people’s research on it has become more in-depth. Machine learning has become a new research core issue in artificial intelligence, which has been widely used in network search, language recognition and machine vision. However, Chinese character recognition with a large character set has always been a difficult problem to be solved urgently. Aiming at the problem of character recognition in artificial intelligence machine learning, this paper focuses on the improvement of CRNN algorithm, and proposes a character recognition algorithm based on feature fusion. Through experimental research, it is found that the accuracy of fine tuning is higher than that of transfer learning, which can better extract the features of the current data set and improve the accuracy.

References
[1] Yawalkar, P. M, Kharat, M. U. (2021) Automatic handwritten character recognition of Devanagari language: a hybrid training algorithm for neural network. Evolutionary Intelligence, 1-18.
[2] Rao, Z., Zeng, C., Wu, M., et al. (2018) Research on a handwritten character recognition algorithm based on an extended nonlinear kernel residual network. Ksii Transactions on Internet & Information Systems, 12(1):413-435.
[3] Li, W. X., Sun, J. F. (2018) A traffic signs' Chinese character recognition algorithm based on mixed optimized deep Boltzmann machine. Computer engineering and science, 040(001):79-85.
[4] Gao, R. X., Ren, X. D., Wu, X., et al. (2017) Research on Character Recognition Algorithm Based on Gaussian Probability Model. Measurement and control technology, 036(003):33-36.
[5] Sahlol, A. T., Elaziz, M. A., Al-Qaness, M., et al. (2020) Handwritten Arabic Optical Character Recognition Approach Based on Hybrid Whale Optimization Algorithm with Neighborhood Rough Set. IEEE Access, PP(99):1-1.
[6] Khaustov, P. A., Spitsyn, V. G., Maksimova, E. I. (2017) Structural model constructing for optical handwritten character recognition. IOP Conference Series: Materials Science and Engineering, 173:012006.
[7] Vazquez-Cervantes, A., Hernandez-Diaz, T., Cordova-Esparza, D. M., et al. (2018) Toward implementation of associative model in real time for character recognition: A hardware architecture proposal for embedded systems. International Journal of Circuit Theory and Applications, 46(9):1723-1732.
[8] Zhou, Y. F., Zhang, H. X. (2019) Anti-tilt Chinese text image file recognition technology . Computer System Application, 28(01):34-39.

[9] TEXT RECOGNITION IN PCBS: AN OBJECT CHARACTER RECOGNITION (OCR) ALGORITHM. International Journal of Development Research, 10(7):1-7 (2020).

[10] Research on License Plate Character Recognition Algorithm Based on MATLAB. Digital technology and application, 000(009):121-123 (2017).

[11] Liu, H. (2017) Application of Radial Basis Function Neural Network Algorithm in License Plate Character Recognition. Journal of Huaqiao University (Natural Science Edition), 038(001):113-116.

[12] Li, L., Gao, X., Jin. L. (2018) HCRCaaS: A Handwritten Character Recognition Container as a Service Based on QoS Guarantee Algorithm. scientific Programming, 2018(PT.2):1-16.