The Study of Handwriting Recognition Algorithms Based on Neural Networks

Barak Finkelstein[1] and Kaplan Kuncan

Athabasca University , Canada
[1]finkelstein@athabascau.ca

Abstract

Handwriting Identifies basic graph-like problems and has a high real-world value in areas such as cloud accounting, finance, and postal administration. Due to the unrestricted problem of handwritten numbers when writing, it is relatively difficult to achieve rapid and effective recognition. With the emergence of deep learning-related algorithms and the rapid development of computer hardware technology, image classification methods based on Convolutional Neural Network (CNN) have gradually become a research hotspot. Because the convolutional network has a strong letter numbering ability and network generalization ability, the recognition rate can often exceed the traditional graph classing method. Therefore, the study of hand-written word recognition should be implemented using CNN through the network. Handwriting Word Recognition is the key technique for self-identification. Therefore, summarizing and analyzing the existing handwritten digit recognition algorithms, two handwritten digit recognition algorithms based on Convolutional Neural Network (CNN) are proposed. To improve the recognition performance of the CNN model, this article proposes a handwriting recognition algorithm based on the change to CNN. To extract the image feature information more fully, this paper proposes a handwriting recognition algorithm based on feature fusion and SVM. First, using the modified CNN model and the Gabor filter that introduces curvature systems, extract the CNN and Gabor characteristics of the character image; Second, the characteristics of its progress are fused to obtain more effective new features; Finally, the fusion feature is entered into the SVM classifier into the line number of words to recognize. The results of the experiment show that the algorithm can effectively improve the recognition effect of handwritten words

Keywords: Handwritten digit recognition, Convolution neural network, Character segmentation, Gabor filter, SVM classifier

1. Introduction

1.1. The background and significance of the study

Handwritten Numeral Recognition (HNR) focuses on how machines or computers can be used to automatically recognize Arabic numerals written on paper, which is an important branch of Optical Character Recognition (OCR) technology. OCR is a comprehensive

technology, which involves pattern recognition, image processing, artificial intelligence, statistical decision-making theory, information theory, computer, and other disciplines,

Handwriting digital recognition is mainly divided into online handwriting digital recognition and offline handwriting digital recognition. The former is easier to identify than the latter but is limited by the input device. The latter can be any kind of image acquisition device, such as a scanner, digital camera, CCD, etc. Because it is not restricted by the input device, so it is more widely used. At present, the main application of handwriting digital recognition has the following three aspects.

(1) Large-scale statistics. For example, in applications such as industry annual inspections, financial statements, and censuses, large amounts of data need to be entered into a computer. If manually entered, it will not only consume a lot of manpower and material resources but also bring a great risk of error. To solve this problem, the researchers proposed the intelligent and efficient automatic machine identification scheme to automatically, accurately, and efficiently enter information and process it. However, most practical systems require users to fill in the specification squares, which makes the recognition of handwritten numbers relatively easy and reduces the requirements of recognition algorithms.

(2) Ticket identification. For example, in bank checks, personal vouchers, and other applications, because of the importance and rigor of entering data information, its recognition requires not only high precision but also high algorithm requirements.

(3) Mail sorting. With the development of the economy, there will be more and more private and business letters, and if sorted manually, the efficiency will be very low. Combined with handwriting digital recognition, manual auxiliary identification, and other methods, the completion of mail sorting automation will become a major trend.

The study of handwriting digital recognition not only has the above important practical significance but also because the pattern of handwritten numbers is relatively simple. This opponent writes digital recognition research has the following important theoretical significance.

(1) Arabic numerals are the most common digital language in the world, and the methods of their research are universal. Scholars can discuss and compare them with each other.

(2) The small pattern category of numbers can be used as an experimental tool to evaluate the effectiveness and feasibility of various research methods, which can help to analyze and validate some new theories.

(3) The method of handwritten digital recognition can be easily extended to other character recognition problems and provide a reference for other character recognition. At the same time, in the context of artificial intelligence, the world is promoting and developing advanced technologies such as big data, cloud computing, and data mining. But the application of handwritten digital recognition technology, input a large amount of data into the computer, which helps the daily word processing, statistics, prediction and will undoubtedly improve productivity and competitiveness.

The research of the digital recognition of the opponent's writing is not only of great practical significance but also has a very broad application prospect. Although there are some difficulties in its research. Once the accuracy rate reaches the application requirements, it will certainly produce great social and economic benefits. So, it has a certain practical significance. Of course, handwriting digital recognition is also of great theoretical significance. First of all, for the ten Arabic numerals 0 to 9, they are closely related to each country, nation, and region, almost all countries in the world use them, and it can be said that they are symbols of uniform use. Therefore, the study of handwriting digital recognition is not isolated but can communicate and explore with each other. Besides, digital recognition has only ten numbers, the variety is relative

to English and characters. The number of kinds is small, so it is convenient to use some new algorithm models to conclude quickly. These new algorithms and models can be further optimized based on the results of theoretical analysis and research. Eventually, better algorithms and models will be applied to handwriting numerical recognition, even to the recognition of handwriting such as English letters or Chinese characters.

1.2. Hand-written word recognition research status quo

The research of handwriting digital recognition based on various algorithms is emerging, and the recognition rate is gradually increasing, but these handwritten number recognition algorithms are based on different data sets introduced by different agencies. Handwritten digital recognition dates back to 1929, researchers use optical template matching for numeric recognition, projecting 10 numeric characters from 0 to 9 onto the corresponding 10 templates, numeric characters obscure the light-transmission portion of the template, and when the light is at its latest, you can get the numeric characters to be recognized. In the early 1970s, Parks et al. proposed a feature extraction method for extracting character topological features and a method for identifying multi-level structure links [1]. Beginning in the 1980s, scholars applied methods such as logical reasoning, Support Vector Machines, and artificial neural networks to OCR technology which achieving better results

In the late 1970s, the Chinese Academy of Sciences carried out further research on digital recognition, successfully developing China's first zip code recognition sample. At the same time, scholars from Fudan University began to study the recognition of printed digital characters, and successfully developed China's first printed digital character recognition machine. At the end of the '80s, many colleges and universities in China carried out a systematic study on digital character recognition, and then digital recognition entered the climax of research. Scholars have proposed several methods for handwriting digital recognition, which can be broadly divided into the following categories.

(1) Template matching method. A template is defined for each category, and characters are compared to all templates when characters are recognized. The category to be recognized belongs to the same category as the template with the smallest difference. This method generally applies only to print characters or regular characters, and is not ideal for randomly written handwritten characters [2].

(2) Logical reasoning. Each pattern class is described first with the composed rules of the characters to be recognized; Each character has a corresponding pattern class. Then using a series of rules in the knowledge base to reason about it, to get the recognition result. The disadvantage of this method is that it is difficult to get the composition examples of the recognized characters

(3) Statistical pattern method. Based on mathematical decision-making theory, a statistical character recognition model is established.   And the characteristics of characters are extracted by statistical analysis which treats recognition characters for classification recognition. The difficulty of this method is that it is difficult to extract the structural characteristics meeting the requirements.

(4) Structure statement method. The complex pattern is first decomposed into several simple sub-pattern combinations, and then the multi-sub-pattern combinations are broken down into multiple sub-elements. The method identifying complex patterns by identifying the base and sub-patterns. Because the extraction of the base is generally difficult, the recognition effect of this method is not very good [3].

(5) Fuzzy judgment. The fuzzy set is divided into multiple subsets with the same number of pattern categories as the membership function and then classified according to the principle of closeness. This method allows for some noise and distortion in the sample to be identified, but it is difficult to establish a suitable membership function [4].

(6) Support vector machine method. Based on the principle of structural risk minimization and the VC dimension theory, the best compromise is sought between the learning ability and complexity of the model through limited learning samples, to obtain a better identification effect and stronger generalization ability. However, this method is difficult to be implemented for large-scale training samples.

(7) Neural network law. The neural network is a distributed and parallel information processing algorithm that imitates the behavior characteristics of an animal neural network. It has the ability of self-learning and self-adaptation. Neural networks include feed-forward neural networks [5], BP neural networks [6], deep belief networks, and convolutional neural networks [7]. Among them, CNN not only has the traditional neural network's strong fault tolerance, self-learning, and adaptive ability but also has the characteristics of automatic extraction characteristics, local connections, and weight sharing.

The simulation results show that this method can improve the classification performance of classical reel neural networks. Alani [8] proposed a handwritten digital recognition method that uses RBM to extract useful features from the original data using a combination of a Restricted Boltzmann Machine (RBM) and a CNN structure, enter the extracted features into CNN, and monitor the training and testing process. The experimental results show that the method can significantly improve the accuracy. Darmatasia, etc. [9], proposed a handwritten digital recognition method that combines CNN's L1 loss function with SVM and L2 normalization, using CNN as a feature extractor. And SVM as a classifier gets better recognition results than CNN [10].

2. Hand-written body number word recognition technique analysis

In the traditional handwriting digital recognition method, once any link of the processing deviation, such as improper pre-processing of the image distortion, etc., will lead to the final extraction of the characteristics of the ability to reduce. In addition, even if each module can obtain its own most accurate characteristic vector expression, the eventually integrated features are not guaranteed to be optimal. This not only wastes a lot of adjustment time but also the recognition effect and network generalization ability are not satisfactory. Other than that, these methods, due to a large number of parameters, make the training and recognition process very time-consuming and complex. And the function of complex classification problems and network generalization ability is limited, which often cannot achieve a high recognition accuracy and rapid convergence of the target stage, hand-written number of words recognition of the entire process that has long matured, as shown in [Figure 1].

Using a broom and other electronic devices to scan the number of words written by a two-dimensional hand, turn it into a black-and-white slice and add some pre-operation. After that, features are extracted from the images in various ways. This step and subsequent classifier recognition are the core of determining the performance of the system.

To judge the quality of the identification algorithm or system studied, we need several parameters to evaluate the merits of the identification and classification performance. However, the final indication is that the following three indicators are required for identifying class performance. For the result of identifying, set Q as the sample total; a indicates the number of samples correctly identified; b indicates the number of misunderstood samples; c represents the

number of samples that have been rejected; A indicates the identification rate; M indicates the misconception rate; R represents the rejection rate.
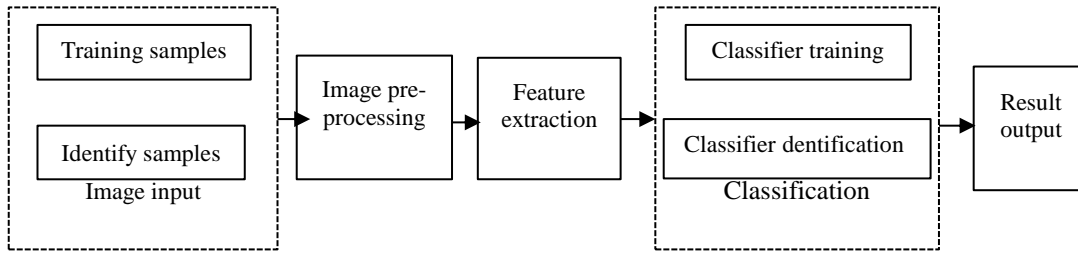

Figure 1. The handwriting digital recognition process

$$A = \frac{a}{Q} \times 100\% \tag{1}$$

$$M = \frac{b}{Q} \times 100\% \tag{2}$$

$$R = \frac{c}{Q} \times 100\% \tag{3}$$

It's obvious that A+M+R=100%. Recognition Rate A is a measure of the ability of a system to correctly recognize several words, representing accuracy; Misrecognized rate M is a measure of the system's ability to incorrectly identify the number of words category; Rejection rate R is a measure of the ability to refuse to recognize several words.

2.1. Hand-written body number word recognizes the process

2.1.1. Picture pre-processing

In some practical situations, the acquisition of an image is influenced by a variety of external objective factors, which may eventually result in lower clarity and completeness of the input image, as shown in [Figure 2]. This significantly reduces the accuracy of the identification. Instead of entering images directly, the sample set is usually preprocessed as an image, which is a series of processing images that contain numbers. The primary goal of the whole pre-processing is to discard the invalid information in the picture, and at the same time to normalize and unify the image, which will facilitate the later feature extraction and classification learning process. Grayscale's conversion, di-valued, smooth noise reduction, skew correction, positioning segmentation, and normalization of centralization are many commonly used pictures processing methods.

According to the characteristics of noise, the recognition image is de-noise processed, the accuracy of recognition processing is improved, and the noise reduction algorithm has a great influence on feature extraction. The median and adaptive Werner filters are generally used to reduce noise in the picture. At the same time, artificial handwriting uncertainty can also lead to overlap between numbers. Strokes are not continuous which will greatly affect the final results of the recognition system. So, each number needs to be positioned and cuts into separate numbers by a series of numbers in the picture. Tilt correction restores the tilted number, through normalization processing. Individual digital images are integrated into the same size and size. And then the recognition algorithm is used for unified processing.
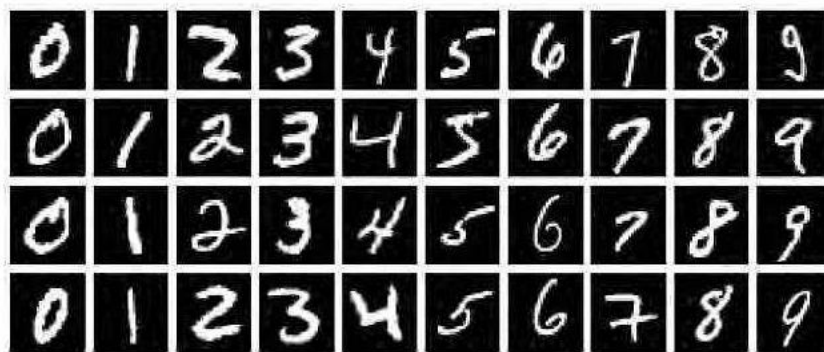
Figure 2. The original image of the handwritten number

2.1.2. Feature extraction

The key to handwriting digital recognition is feature extraction of numbers. That is, after image pre-processing, the current image feature is extracted as a template feature vector by a series of methods, and then fed into the classifier for training and recognition. In the first chapter, we also mention the current commonly used extraction characteristics method. That is generally divided into statistics, structural feature extraction, and neural networks. The former uses the closest method, histogram method, random selection method, recursive feature rejection method, and other screening characteristics, many effective statistical and structural features are proposed. However, the application scope of these feature extraction and selection methods is narrow and the generalization ability is poor. But to carry out the operation of dimensional reduction, the efficiency is too low. The latter, especially the coil neural network makes full use of the advantages of simple feature extraction method, feature extraction, and classifier classification. This makes the whole feature extraction process very simple and much more efficient.

2.1.3. Classifier

The last step of handwritten number recognition is classifier recognition. At present, the classification algorithms used in handwritten number recognition are rough: proximity algorithm, SVM algorithm, BP neural network method, and so on. The first algorithm has the lowest accuracy and takes a long time to classify. The second SVM algorithm is mainly used to identify the two-classification problem, which is not suitable for the multi-classification problem of handwriting numerical recognition. If you want to use the SVM algorithm, you must build a multi-level classifier that consumes a lot of saving space and computing time when storing and calculating samples. If the sample set for training is too large, it cannot be implemented in detail. For traditional handwriting digital recognition methods, most digital pictures cannot be entered directly. And you must first remove the structure or statistical features from the original picture, and then pass them into the classifier to do the training recognition work. This must-do tedious pre-processing of pictures, while the separation of feature extraction and training and learning process will also lead to low recognition efficiency and correctness. The module functions such as feature extraction and classifier recognition of the convolutional neural network are assembled in the same model. And by feeling the characteristics of wild local connection and weight sharing between layers and layers, the whole network model greatly simplifies the complexity, and the number of parameter weights between layers is greatly reduced, which eventually leads to the network's training and learning

knowledge is easier, but also improves the generalization ability and accuracy of classification. Common classification algorithms for constrains neural networks are logical regression, SoftMax regression, and the most classic SVM.

3. The principle of the experiment

3.1. CNN's co-core initialization method

The nature of the training process for convolutional neural networks is to update network parameters, i.e., weight vectors and bias vectors. This first requires initializing the parameters of each layer and then using optimization algorithms such as gradient drop for the target function. $k_{ij}^{(l)} b_j^{(l)}$. However, for the training model, nonlinear functions are heavily superimposed which resulting in a non-convex $J(K, b)$ function. How to choose the initial value of parameters makes the target function easy to be optimized has become a question worth exploring.

The essence of the random initialization method "Xavier" is to ensure that the inputs and outputs of neurons have the same variance to avoid all output values tending to 0. They took the weights from a Gaussian with a mean of zero and a variance of $2/(n_{in} + n_{out})$, where n_{in} is the number of incoming neurons and n_{out} is the number of outgoing neurons. Therefore, the Xavier method automatically determines the size of the initialization based on the number of input and output neurons and keeps the signal within a reasonable range of values through multiple layers.

The Goss distribution with a zero average and a standard deviation of 0.01 initializes the weight of its network and sets the bias parameters for the second, fourth, and fifth-reel layers, as well as the full-connection layers to constants. Caffe1 uses the only n_{in} to initialize the co-product parameters, which makes it easier to implement. Ortho matrix initialization is better for linear networks than Gaussian initialization. And it is also suitable for networks with nonlinearity. A hierarchical unit variance processes scheme that can be considered a formal initialization of standards combined with batch standardization that is performed only on the first small batch. Although it is similar to batch standardization because they all use the unit variance normalization process, but it uses Orthogonalized normalization to initialize weights, which helps effectively unrude the correlation between layers. This initialization technique has been used in unsealed visual representing, and its performance has been significantly improved [11][12][13].

Table 1. The schematic diagram of the convolutional neural network structure used in this paper

The input layer
Reel layer C1: 6 5 x 5 co-product cores, sigmoid
Pooled layer S1: average, no bias, and activation functions
Reel layerC2：12 co-product nucleation's of 5 x 5, Sigmoid
FlattenLayerF6：Just "tile" all the elements of the pooled layer S2 output
Output layer: Sigmoid, no bias
SPP pooled layer S2:2-dimensional feature vector

The convolutional neural network structure designed in this chapter consists of 7 layers, consisting of 1 input layer, 2-convolutional layers, 2 sampling layers, 1 Flatten layer, and 1 output layer. In addition to the input and output layers, the middle five layers C1, S1, C2, S2, F6 can be considered hidden layers, and the structure is shown in [Table 1].

3.2. The basic course of handwriting recognition

For handwritten digital recognition issues, a single feature of a character image typically contains only limited information about the image. Different image features contain different information of the image, and the advantages of each feature are complementary through feature fusion technology. To extract the image feature information more fully, this chapter uses the improved CNN network model to extract the CNN characteristics of the character image by the ability of the convolutional neural network to automatically extract the image feature information; Using the improved Gabor filter to extract the Gabor features of the image, the two features will be extracted more effectively by feature fusion technology. Finally, it is entered into the SVM classifier for character classification recognition. The overall framework of the algorithm in this chapter is shown in [Figure 3].
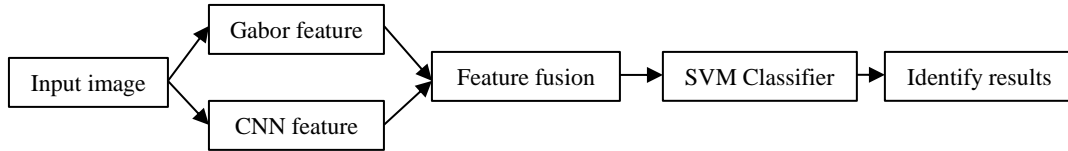
Figure 3. Gabor-CNN-SVM algorithm overall framework

CNN has good feature extraction capabilities, but the classification performance in general. After feature extraction is complete, RBF or SoftMax classifiers are typically used to enter the classification category, but with the non-stop development of shallow classifiers such as SVM, sparse representing, and fluid learning, their class performance has also improved greatly. Therefore, this chapter is improved for the CNN network model to extract the CNN characteristics of the character image. The Gabor feature of the character image is extracted using the Gabor filter that introduces the curvature coefficient. Then the Gabor feature and CNN feature are fused into the line features. And finally, the fused feature vector is transported into the SVM classifier implementation hand. The category recognition of the number of words written.

Handwriting Word Recognition is a multi-category problem that contains 10 Arabic words from 0 to 9. Common SVM multiclassing methods are "one-to-one," "one-to-many", and the forward-to-no-loop method. This chapter uses a one-to-one multi-category method to identify the characters to be recognized by the line classification class. The one-to-one multi-classification method is to divide a multi-classification problem into several binary classification problems for solving. Through training and learning of samples, the optimal classification function f (x) between every two categories is found, and the two categories are correctly identified with the maximum probability. For handwriting recognition issues that contain 10 categories, a pair of one-to-one SVM multi-classifiers can construct 45 class devices. By training two types of samples in each class of classes to achieve the division between any two classes, the last group of all the two classes get a pair of one SVM multi-classes Class.

To get the optimal classification function between i, j, and the two classifiers, $f(x) = sgn(w^T \cdot \varphi(x) + b)$, and achieve the correct classification between any two sample classes,

you need to maximize the classification interval between the two sample classes and minimize $\|w\|^2/2$ the value of. The optimal classification function f (x) can judge samples between the two categories so that the error rate is minimal. Selecting the appropriate nuclear $K(x, x) = (\varphi(x), \varphi(x))$ *the* function enables the SVM to linearly classify the samples between the two categories through the optimal classification function in high-dimensional nuclear space. Currently, the most commonly used nuclear functions in SVM classifiers are.

(1) Linear core function:$K(x_i, x_j) = x_i \cdot x_j$

(2) Polyethnic nuclear function: $K(x_i, x_j) = [(x_i \cdot x_j) + 1]^q$

(3) Sigmoid core function:$K(x_i, x_j) = tanh(k(x_i \cdot x_j) + v)$

(4) RBF Letter Number: $K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$

The RBF function is selected as the core function of the SVM classifier. After the one-to-one SVM multi classifier is constructed between any two sample categories, the sample category is judged by the "voting method". Use the classification function to treat the identification sample x to vote, if i, j between the two classifications $f_{ij}(x) \geq 0$ functions, i class to get one vote; If the classification $f_{ij}(x) \leq 0$ functions, class j gets one vote and finally determines the category of sample x to be identified based on the number of votes obtained.

3.2. Gabor-CNN-SVM algorithm

Enter: Q of the image to be recognized
Output: Category label
Step 1 reels the preprocessed handwritten fractional character image with the 2D $I_{m \times n}$ Gabor filter to extract the Gabor characteristics of the character image $X = (x_1, x_2, \cdots, x_k)$
Step 2 will also enter an improved $I_{m \times n}$ CNN network model that extracts the CNN characteristics of character images using alternatingly connected "Volume and Downsampling" structure $Y = (y_1, y_2, \cdots, y_k)$
Step 3 calculates the feature vector V_{X+Y} after feature fusion
Step 4 fusion feature input V_{X+Y} SVM classifier, using one-to-one SVM multi-classifier for character classification recognition

4. Experimental results and analysis

This chapter also experiments with homemade sample sets and MNIST datasets in an experiment. Combined with the experimental parameters in the literature and the results of many experiments, the experimental parameters are set to Scale characteristics σ=2.75, image bending coefficient c=0.1, penalty factor c=128, core parameter δ=0.2. The experimental environment is Windows 10 operating system, the simulation platform is MATLAB R2014a. The experimental results of various algorithms are run 10 times on average. To verify the performance of the algorithm in this chapter, the influence of the relevant factors on the experimental results, and the recognition effect of the algorithm on the handwriting score of the paper, the following comparative experiments were carried out respectively.

Because different nuclear functions and classification methods have different adaptability and recognition effects, different multi-classification methods and nuclear functions can be used to construct different SVM classifiers. To study the effect of the SVM multi-classification method and nuclear function on recognition rate, common multi-classification methods and nuclear functions are used to compare experiments on homemade sample sets, and the results are shown in [Table 2].

Table 2. Compared with the experimental results of nuclear functions based on different multi-classification methods

Core function	Multi-class method					
	One to one		One to many		DAG	
	Recognition rate	Run time	Recognition rate	Run time	Recognition rate	Run time
Line core	85.5	93.2	83.5	152.1	83.9	123.4
Multiple cores	86.4	104.1	84.2	160.5	85.6	137.5
Sigmoid core	94.1	112.7	90.3	172.1	91.4	152.3
RBF core	97.6	75.9	95.3	132.3	93.5	115.4

As can be seen from [Table 2], the choice of different multi-classification methods and nuclear functions has an impact on recognition rate and run time. For classification methods, a one-to-many classification method has a higher recognition rate than a one-to-many and DAG approach. The recognition rate of one-to-many and DAG classification methods is not much different. The one-to-many classification method has some advantages in runtime. For nuclear functions with different classification methods, RBF core functions have the highest recognition rate and are significantly better than the other three nuclear functions. So SVM's strategy of combining one-to-one multi-classified and RBF nuclear functions works best.

5. Conclusion

Under the background of the comprehensive development of global inform ionization and the improvement of automation, the demand for digital recognition of rival writings has become very urgent. Given the limitations of the handwritten numeral recognition method, this paper based on artificial intelligence and big data technology enterprise project, in combination with convolution the advantage of neural network in image classification [14][15][16] processing, put forward how to use handwritten digital convolution neural network classification and recognition, and how to through improved convolution neural network model, made in the implementation of handwritten numeral recognition and a higher recognition rate and better performance. The main contents are summarized below.

(1) This paper analyzes and summarizes the research background and significance of digit recognition, introduces the research status of handwritten digit recognition, and summarizes the commonly used handwritten digit recognition methods combined with the literature.

(2) Pre-processing is an important part of the handwriting digital recognition system, which designs and implements the pre-processing of the paper image for the actual working environment. It mainly introduces the extraction of the handwriting fraction area and the segmentation of the sticky hyphen;

(3) Given the problem of handwriting score recognition of the head of test paper, a handwritten digital recognition algorithm based on improving the courage neural network model is proposed. Combined with increasing the diversity of the sampling layer, this method can take into account the characteristic values obtained by the average sampling and the maximum sampling. So that the richer image features can be extracted. During the network training phase, the network training speed is improved by increasing the momentum item and adopting the adaptive adjustment strategy of learning rate. The superiority of the algorithm is proved by comparative experiments;

(4) Because the CNN network model is a good feature extractor and classification capability in general. To extract image feature information and improve the recognition effect, the improved CNN network model is used as the CNN feature of the feature extractor to extract character images. The Gabor feature of the character image is extracted by the Gabor filter that introduces the curvature coefficient. Then the two features are fused by feature fusion technology, and more effective fusion features are obtained. Finally, the fusion feature is entered into the SVM classifier for classification and identification, so the Gabor-CNN-SVM algorithm is proposed. Applying it to the handwritten number recognition of the test paper further verifies the validity of the algorithm.

References

[1] Hao M. L., Tian Q. C., and Bai X., "The study of feature extraction of handwritten digits based on ruler algorithm," Applied Mechanics and Materials, pp.1202-1207, **(2014)**

[2] Yu N. G. and Jiao P. N., "Handwritten digits recognition approach research based on distance and Kernel PCA," // Proceedings of the 5th International Conference on Advanced Computational Intelligence. Nanjing: IEEE, pp.689-693, **(2012)**

[3] Ma C. Y. and Zhang H., "Effective handwritten digit recognition based on multi-feature extraction and deep analysis," // Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery, Zhangjiajie: IEEE, pp.297-301, **(2015)**

[4] Wu X. H. and Zhou J. J., "Fuzzy discriminant analysis with kernel methods," Pattern Recognition, vol.39, no.11, pp.2236-2239, **(2006)**

[5] Robinson M. D., Manry M. T., Malalur S. S., et al., "Properties of a batch training algorithm for feedforward networks," Neural Processing Letters, vol.45, no.3, pp.841-854, **(2017)**

[6] Yu B. S., Wang S. L., Yang T., et al., "BP neural network constitutive model based on optimization with genetic algorithm for SMA," Acta Metallurgica Sinica, vol.53, no.2, pp.248-256, **(2017)**

[7] Sanni K., Garreau G., Molin J. L., et al., "FPGA implementation of a deep belief network architecture for character recognition using stochastic computation," // Information Sciences and Systems. Baltimore:IEEE, pp.1-5, **(2015)**

[8] Alani A., "Arabic handwritten digit recognition based on restricted Boltzmann machine and convolutional neural networks," Information, vol.8, no.4, pp.142-150, **(2017)**

[9] Darmatasia A., Fanany M., "Handwriting recognition on form document using convolutional neural network and support vector machines," International Conference on Information and Communication Technology, Malaysia: IEEE, pp.345-353, **(2017)**

[10] D. Mishkin and J. Matas, "All you need is a good in it," arXiv:1511.06422, **(2016)**

[11] Doersch C., Gupta A., and Efros A., "Unsupervised visual representation learning by context prediction," IEEE International Conference on Computer Vision, **(2015)**

[12] Wang X. and Gupta A., "Unsupervised learning of visual representations using videos," IEEE International Conference on Computer Vision, **(2015)**

[13] P. Agrawal, J. Carreira, and J. Malik, "Learning to see by moving," IEEE International Conference on Computer Vision, **(2015)**

[14] Sun M., Song Z., Jiang X., et al., "Learning pooling for convolutional neural network," Neurocomputing, vol.49, no.10, pp.224-235, **(2016)**

[15] Ma C. Y. and Zhang H., "Effective handwritten digit recognition based on multi-feature extraction and deep analysis," Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery, Zhangjiajie: IEEE, pp.297-301, **(2015)**

[16] Zufar Kayumov, Dmitrii Tumakov, and Sergey Mosin, "Hierarchical convolutional neural network for handwritten digits recognition," Procedia Computer Science, vol.171, pp.1927-1934, **(2020)** DOI: 10.1016/j.procs.2020.04.206