ISSN: 2664-9640

Chinese calligraphy style recognition based on vgg16 convolutional neural network after dimension reduction

Meng Tao *, Guoliang Yuan

College of Information Engineering, Shanghai Maritime University, Shanghai 200120, China;

* Corresponding Author

Abstract

Chinese calligraphy has unique art, attracting many scholars to study and learn. There have been various calligraphy styles in China for 5000 years. The most influential styles are seal script, clerical script, standard script,semi-cursive script and cursive script. They are also the styles that many scholars want to imitate and sublimate.Now we are using standard simplified Chinese characters, many scholars are difficult to identify the long history of calligraphy style, so this paper proposes a method through deep learning and machine learning to help them identify calligraphy style.We first use vgg16 convolutional neural network to extract depth features, then reduce the extracted high-dimensional features to low dimensional space by dimension reduction, and use support vector machine classification method to recognize style in low dimensional space.The experimental results show that the accuracy of the proposed method is higher than that of vgg16, and the method has a good effect on style recognition.

Keywords

VGG16 neural network,PCA dimension reduction,Support Vector machine, Calligraphy style recognition.

1. Introduction

We all know that the first thing to do in Chinese New Year is to post couplets. A neat and beautiful couplet will be refreshing. Even ordinary people who do not have a deep knowledge of Chinese calligraphy will applaud when he reads a beautiful calligraphy couplet.From this,we know that the influence of calligraphy in Chinese traditional culture.After thousands of years of development, Chinese calligraphy not only expresses the writer's message, but also develops into an art, just like the ancient proverb "The word is like the person", it deeply expounds the value of calligraphy. Therefore, many scholars and ordinary people are attracted to study calligraphy. At present, there are five main styles of Chinese calligraphy: seal script, clerical script, standard script, semi-cursive script, and cursive script. For some beginners, when appreciating calligraphy works, they will not only lament the author's superb skills, but also further think and explore based on the inherent font and style of the calligraphy works, hoping to appreciate calligraphy works with similar styles in a style-guided way. This brings some urgent scientific research problems to the analysis, recognition and style classification of digital calligraphy works, which should be solved. In recent years, many researchers have devoted themselves to the study of handwritten Chinese character recognition[1][2], using traditional methods of manually extracting features[3][4], and then using some conventional machine learning algorithms such as nearest neighbor classification algorithm to recognize Chinese characters. With the development and application of artificial intelligence, more and more researchers use deep learning based on convolutional neural networks [5] [6] to recognize different Chinese characters, but there are not many researchers study on calligraphy. Research on style recognition.

ISSN: 2664-9640

Figure 1:Different writing styles of the word "匕"

As shown in Figure 1, they are quite different at first glance, but they are all different writing styles of words that express the same meaning, namely "匕".From left to right are the standard script, semi-cursive script, cursive script, clerical script, and seal script styles. For ordinary people who have not studied deeply or have not studied it, it is still very difficult to distinguish which style they are. So how do us make the machine recognize it? Wang Xiao, Zhang Xiafen, etc. [7] proposed Calligraphy Style Identification Based on Visual Features. The research method of the article is to extract features from the single-character image and stroke structure to obtain a total of 24 types of features as feature vectors, and calculate the probability of the sample to be tested and the five types of styles to obtain the style category. This method is manual feature extraction, which is not only computationally intensive and cannot achieve real-time results, but the recognition rate is not very high. Zhang Jiulong et al. [8] performed style recognition by extracting local and global features of the image. The extraction of local features uses the SIFT algorithm, and the extraction of global features uses the GIST algorithm, and then uses support vector machines for classification, although the recognition rate is somewhat improved, but the amount of calculation is not small. In order to improve efficiency, Mao Tianjiao et al. [9] proposed an improved KNN algorithm, which extracts more accurate feature points from the key points obtained by using the SIFT algorithm, and then uses the improved KNN algorithm as a classifier, and a good result is obtained. The recognition rate also reduces the computational complexity. With the development of artificial intelligence, more and more researchers apply neural networks to the recognition of calligraphy styles. GaoPengcheng et al. [10] used convolutional neural networks to recognize calligraphy styles and sent character images to the AlexNet network and use the DeCAF algorithm to extract the fully connected layer of the sixth layer as a feature, and design MQDF for recognition. This method can reach the accuracy of AlexNet, and it can also greatly reduce the amount of calculation and increase the speed. JiulongZhang et al. [11] added the squeeze-and-extract module after the convolutional layer to enhance important features while suppressing useless features, and then use softmax to classify and recognize through a layer of wavelet transform layer, which has high accuracy rate.

2. Convolutional Neural Network

2.1. LeNet-5 neural network

The earliest convolutional neural network is the LeNet-5 neural network proposed by Yann LeCun[12], as shown in Figure 2 below.

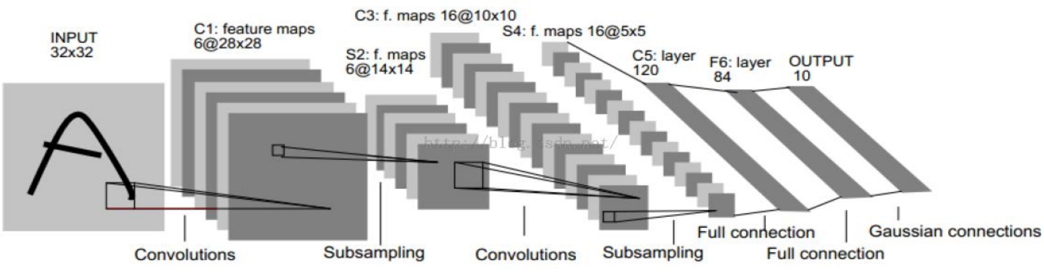

Figure 2:LeNet-5 Convolutional Neural Network

A convolutional layer and a pooling layer are added to the traditional neural network. The convolutional layer mainly implements feature extraction to obtain features that are easy to

ISSN: 2664-9640

distinguish between different styles. The pooling layer can reduce the dimensionality of high-dimensional features to ensure effective classification. In the case of reducing secondary information, reducing the amount of calculation and increasing the phase rate. The model includes 4 convolutional layers, 2 fully connected layers and a Gaussian connected layer, and finally the classification results are output through the softmax layer. After the input image is subjected to convolution operation, the feature map is obtained, and then the pooling operation is performed to obtain the output image, and then the previous steps are continued on the output image, until the fourth layer is connected with the fully connected layer, and the fully connected layer is same with the traditional neural network by designing different deep networks and using the training set to train the parameters of the network to automatically recognize the style. Because more layers are added, the network has better fitting ability. At the same time, because each layer of convolution kernel weights are shared, the parameters are reduced and the network is prevented from overfitting. Compared with traditional image classification methods, it no longer requires manual feature description and extraction of target images[13], but autonomously learns features from training samples through neural networks, and these features are closely related to the classifier. This solves the problem of manual feature extraction and classifier selection, and has greatly promoted the development of artificial intelligence [14].

2.1.1 Convolution operation

Image convolution operation[15] realizes the convolution operation by using the convolution kernel to multiply and add the image, which is defined as:

$$g = f * h \Rightarrow g(i,j) = \sum_k \sum_l f(k,l)h(i-k, j-l) \tag{1}$$

In the formula, f(i,j) is the input image pixel, h(i,j) is the convolution kernel pixel, and g is the output image.

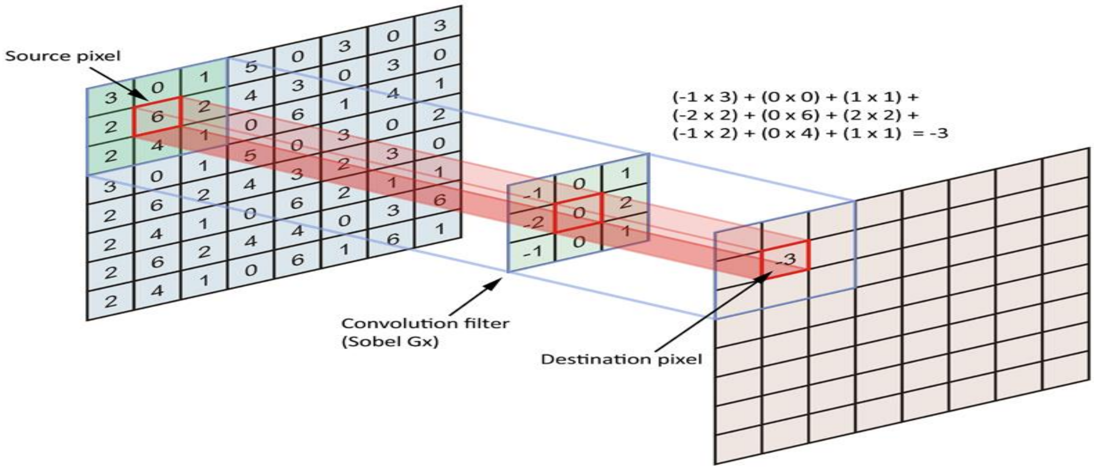

Figure 3: Convolution operation process

As shown in Figure 3, the input image pixels are on the left, the convolution kernel template is in the middle, and the processed output image is on the right. Let the convolution kernel move according to a certain rule on the input image, calculate the product of the convolution kernel coefficient and the corresponding input pixel below it, and then add the product and sum, and finally output the result of the operation to obtain the characteristic image. The detailed operation is shown in the figure Shown. The operation of convolution can enhance the features, extract the features useful to the result, remove the secondary features, and reduce the amount of data. In a convolutional neural network, the coefficients of the convolution kernel are exactly the parameters we need to train. Using convolution kernels with different numbers of channels in each layer can produce different output feature maps, and the dimensions of these feature maps are also relatively large, so a pooling layer is needed to extract the main features.

ISSN: 2664-9640

2.1 .2 Pooling layer

The pooling layer also uses a window to slide in the image to process the corresponding pixels. Pooling has maximum pooling and average pooling. In general convolutional networks, maximum pooling is used. There are also documents that show that there is not much difference between maximum pooling and average pooling. Here is an introduction to maximum pooling. Maximum pooling is the area corresponding to the size of the filter on the image, and the maximum value of the pixel is taken in this area to obtain the feature data. Generally speaking, the feature data obtained by this method better retains the texture of the image feature.

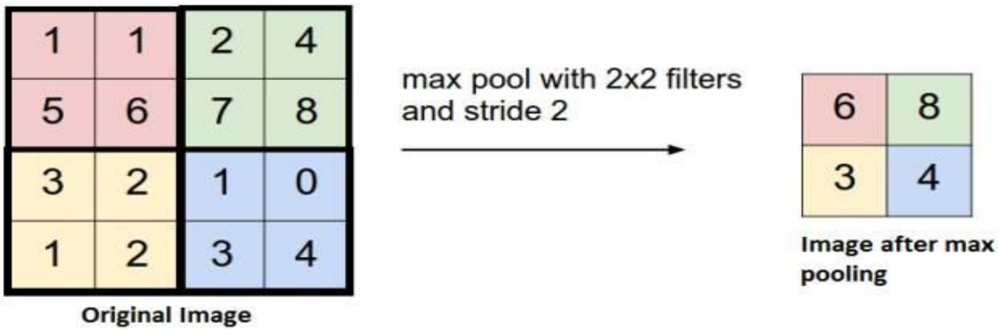

Figure 4: Pooling process

The pooling process is shown in Figure 4, using a window with a size of 2 and sliding on the input image with a step length of 2, selecting the maximum value of the corresponding input pixels below the window as the output, and continuing to slide to get the maximum value in all sliding windows as output result graph, this graph is the characteristic graph after pooling.

2.2. VGG16 neural network

VGG[16] is a new neural network proposed by Simonyan and Zisserman on the basis of convolutional neural network, and used this model to participate in the 2014 ImageNet image classification and positioning challenge, and obtained excellent results. For neural networks, the greater the number of layers, the stronger the fitting ability, which can improve the accuracy of recognition. The VGG neural network is designed to verify this result. Experiments show that when the depth of the network is 16 layers, the accuracy is the best Good, that is, the following VGG16 neural network.

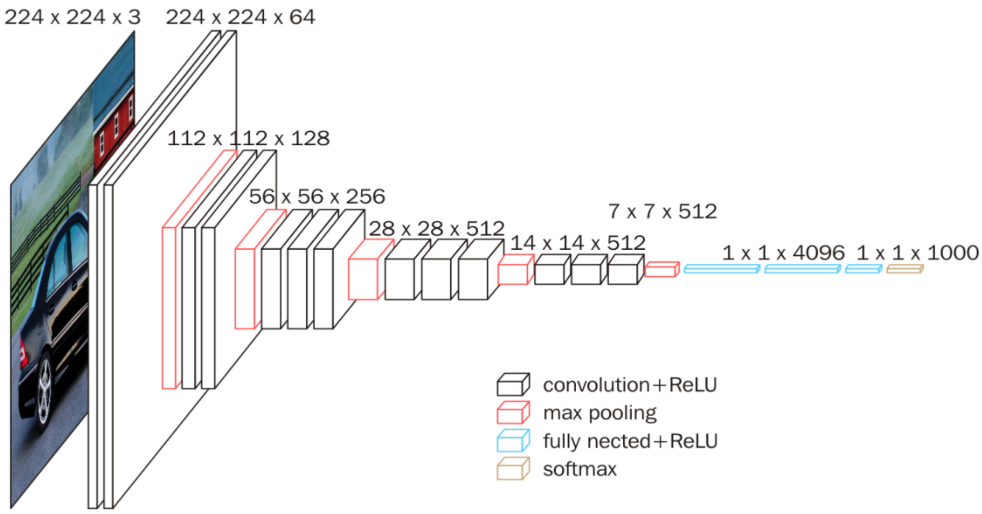

Figure 5: VGG16 neural network model

In Figure5, there are a total of 13 convolutional layers, 3 fully connected layers and 5 pooling layers. The convolution kernels in all convolutional layers are the same size, all of the same size, and the sliding step is 1.The pooling layer is a window template with a step size of 2.The size of both is the same, the sliding step size is 1, the pooling layer is a window template with a step size of 2, the first two fully connected layers have a dimension of 4096, and the third fully connected layer has a dimension of 1000, and softmax, which the dimension is also 1000, that is, there are 1000 categories of output. Since there are only five calligraphy styles we studied, the output of softmax is set to 5 in the experiment. VGG16 is to recognize and classify large samples, and our experiment is only small samples. The features extracted by this neural network will appear redundant, and the dimensionality reduction algorithm is needed to reduce the dimensionality. The characteristic of the convolutional neural network is that the nodes close to the input layer represent the abstraction of the image in low dimensions, while the nodes close to the output layer represent the abstraction of the image in higher dimensions. Low-dimensional abstraction describes the texture and style of the image, while high-dimensional abstraction describes the layout and overall characteristics of the image. Therefore, high-dimensional features can better represent the content of the image. Therefore, it is necessary to reduce the dimensionality of the features extracted by VGG16 in the high-dimensional.

3. Dimensionality reduction

3.1. PCA

Principal component analysis (PCA)[17][18] is often used to reduce the dimensionality of a data set, while retaining the feature that contributes the most to the variance in the data set. This is done by retaining low-dimensional principal components and ignoring unnecessary high-dimensional principal components. Such low-dimensional components can often retain the most important part of the data. PCA is the simplest method to analyze multivariate statistical distributions with feature quantities. Usually, this kind of operation can be seen as revealing the internal structure of the data, so as to better show the variability of the data. If a multivariate data set is represented by the coordinate system of a high-dimensional data space, then PCA can provide a lower-dimensional image, which is equivalent to a projection of the data set on the angle with the most information. In this way, a small amount of principal components can be used to reduce the dimensionality of the data and lose some secondary information to obtain low dimensionality. The mathematical principles of PCA are as follows:

The number of input samples is m , and each vector represents $x^i (i=1,2,\ldots n)$, The samples can be combined into a matrix of size $m \times n$, and the hope is by follows:

$$z = W^T x \tag{2}$$

linear transformation to obtain the transformed z , its size is $m \times k$, and k is much smaller than n , and the effect of dimensionality reduction can be achieved. The transformation matrix W is what we want. In order to obtain W , let the transformed variance be the largest of as the following formulas:

$$Var(z) = \frac{1}{N}\sum_z (z-\bar{z})^2 = (W)^T \frac{1}{N}\sum (x-\bar{x})(x-\bar{x})^T W \tag{3}$$

In the formula, z is the transformed data, $\bar{z}$ is the average value, x is the input data, and W is the required matrix, which is a matrix composed of all feature map vector row vectors, and this matrix retains important information of the original data. Let:

$$S = \frac{1}{N}\sum (x-\bar{x})(x-\bar{x})^T \tag{4}$$

then S is the covariance matrix of the original data. So get a linear optimization function:

$$\max(W^T)SW \quad s.t \| w^i \| = 1 \tag{5}$$

ISSN: 2664-9640

After calculation, w^i is the eigenvector corresponding to the eigenvalues of the covariance matrix S arranged by size. If we take the characteristic matrix W composed of the eigenvectors corresponding to the first k eigenvalues ,it is possible to project the n-dimensional vector of the original data in the k-dimensional feature space, so as to achieve the effect of dimensionality reduction. Often the input sample dimensions are very large, which makes the calculation of the covariance matrix S very complicated, and it is not easy to obtain the transformation matrix W. Usually, the singular value decomposition SVD of the matrix is needed to solve it.

3.2. SVD decomposition

Singular value decomposition[19] is a commonly used method for high-dimensional decomposition to low-dimensional space. It is widely used in many machine learning algorithms. The specific decomposition method is as follows:

For any matrix A of $m \times n$, its SVD is:

$$A = U \Sigma V^T \tag{6}$$

U is a unitary matrix of $m \times m$, Σ is an $m \times n$ diagonal matrix, usually a matrix composed of the singular values of A. The singular value σ_i has the following relationship with the eigenvalue λ_i of matrix A :

$$\sigma_i = \sqrt{\lambda_i} \tag{7}$$

V is a unitary matrix of $n \times n$,by:

$$(AA^T) u_i = \lambda_i u_i \tag{8}$$

to find each eigenvector of the matrix U, where l is the eigenvalue and u is the eigenvector. In the same way, we can use:

$$(A^T A) v_i = \lambda_i v_i \tag{9}$$

to find the matrix V.In this way, the SVD of matrix A can be obtained.

3.3. The relationship between PCA and SVD[20]

In the low-dimensional space formed by w, high-dimensional data can be represented by the following formula:

$$x \approx c_1 w^1 + c_2 w^2 + \cdots + c_k w^k + \bar{x} \tag{10}$$

In order to find these k vectors, you can make the following formula take the minimum value:

$$L_{error} = \sum \| (x - \bar{x}) - (\sum_{k=1}^{K} c_k w^k) \|_2 \tag{11}$$

For (10), it can be disassembled as:

$$\begin{aligned} x^1 - \bar{x} &\approx c_1^1 w^1 + c_2^1 w^2 + \cdots \\ x^2 - \bar{x} &\approx c_1^2 w^1 + c_2^2 w^2 + \cdots \\ &\vdots \end{aligned} \tag{12}$$

That is:

$$X = x - \bar{x} \approx U \Sigma V \tag{13}$$

x is the $m \times n$ matrix, U is the $m \times k$ matrix, Σ is the $k \times k$ matrix, and V is the $k \times n$ matrix. For the decomposition of X , the SVD method is used, and the matrix U is the eigenvalue composed of the first k eigenvectors obtained through the covariance S ,which is the PCA decomposition. After the feature matrix U is obtained, formula (13) can be used to reduce the dimensionality of the original data.

4. Experiment and analysis

4.1. Data set and experimental environment

China Academic Digital Associate Library(CADAL) is the largest digital library in China. It collects a large number of digital style fonts, which is very convenient for scholars to use. In CADAL, a five-style dictionary contains more than 100,000 characters, including seal script, clerical script, standard script, seni-cursive script and cursive script. This article selects five thousand characters for each style for experiment, and divides them into training set and test set according to 4:1.The software environment of the experiment is the Windows operating system and the TensorFlow neural network framework. The programming language uses Python 3.6. The hardware is a PC equipped with CUDA-supported NVIDIA GPU graphics card GeForceGTX285, Xeon quad-core processor and 32GB memory. The flow chart of the experiment is shown in Figure 6.

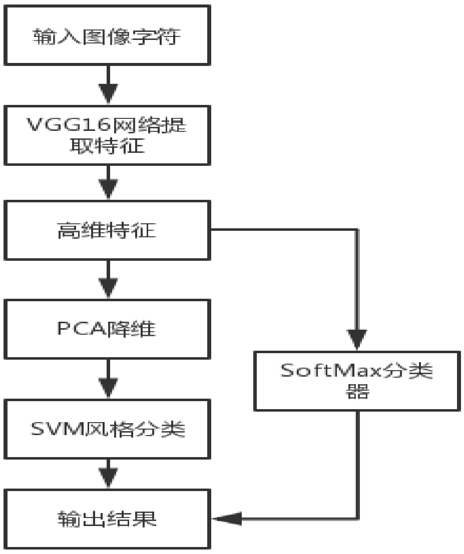

Figure 6: Experimental flowchart

4.2. Experimental results

The experiment first sends the character image into the VGG16 network to train the model, and the experimental results obtained are shown in Figure 7 below.

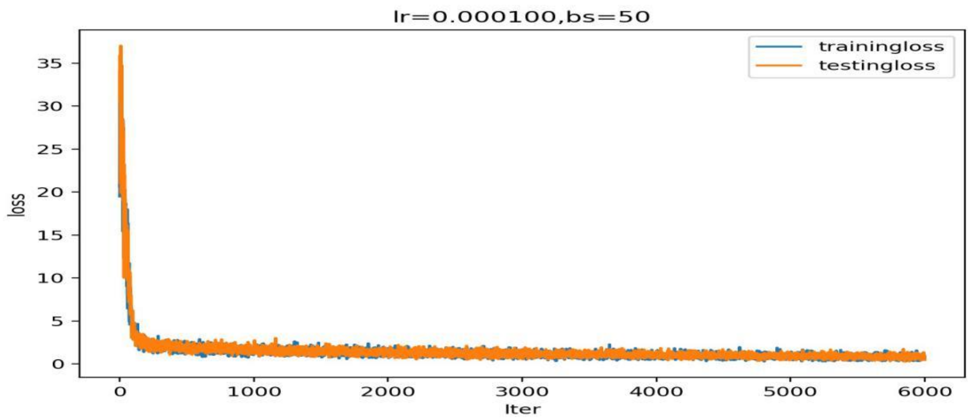

Figure 7: Training results of VGG16 neural network

It can be seen from the figure that when the training is about 6000 times before, the error between the training set and the test set is 9.2%, that is, the correct rate of the five styles of calligraphic character recognition using VGG16 is 90.8%. For the VGG16 network, the fully connected layer has a dimension of 4096, and there is more redundant information for the five style classifications, and the dimensions are relatively large and consume memory, so dimensionality reduction processing is required. Use PCA dimensionality reduction processing to reduce high-dimensional vectors to low dimensionality, and then use SVM[21][22] for classification. The experimental results are shown in Table 1.

Table 1: Accuracy of style recognition in different dimensions

dimension	4096	1024	256	32	16	8
accuracy	92.3%	95.4%	97.2%	89.4%	87.6%	73.6%

4.3. Experimental analysis

In Table 1, it can be seen that when the extracted high-dimensional features are directly classified using SVM without dimensionality reduction, the recognition accuracy is higher than that of the original VGG16 network, which proves that the SVM classifier can improve the network performance. Accuracy; with the use of PCA to gradually reduce the high-dimensionality to the low-dimensionality, the recognition rate first increases and then decreases. The analysis may be due to the fact that there are fewer types of styles, so there is redundant information in high-dimensional features, and when it is reduced When the dimension is very low, the useful information is drastically reduced, resulting in a very low recognition rate.

5. Conclusion

By combining the VGG16 convolutional neural network with the SVM classifier, not only can excellent machine learning algorithms be added to the neural network, but also the computational complexity can be reduced, and the recognition accuracy can also be improved. First, the trained neural network is used to extract high-dimensional features, and the high-dimensional features are reduced by PCA, and then the reduced features are sent to the designed SVM classifier to identify different styles. By comparing the accuracy of style recognition directly with the VGG16 neural network, it can be obtained from experiments that the recognition rate of our method reaches 97.2%, which is 6.4% higher than that of VGG16, which proves that the combination of machine learning and deep learning The method is feasible.The VGG16 network has many parameters and consumes relatively large resources. The next step is to try to improve its network structure, and then combine with machine learning to recognize calligraphy styles to obtain higher performance.

References

[1] ZhaoJiYin,ZhengRuiRui,WuBaoYing,LiMin.A Review of Off-Line Handwritten Chinese Character Recognition[J].Electronic Journal,Vol.38(2010)No2,p.405-415.

[2] Soselia D , Tsintsadze M , Shugliashvili L , et al. On Georgian Handwritten Character Recognition[J]. IFAC-PapersOnLine, Vol.51(2018)No30,p.161-165.

[3] WangMing,ZengBaoYing,YaoChenHong,TianXiangYuan,SunXiangNan.Feature extraction and recognition of Chinese calligraphy[J].Information and Communication,Vol.151(2015)No7,p.19-20.

[4] Yuan Lin,Jiangqin Wu,Pengcheng Gao,Yang Xia,Tianjiao Mao. LSH-based large scale chinese calligraphic character recognition[P]. Digital libraries,Vol.10(2013)No1145,p.323-330.

[5] Xiao X , Jin L , Yang Y , et al. Building Fast and Compact Convolutional Neural Networks for Offline Handwritten Chinese Character Recognition[J]. Pattern Recognition,Vol.72(2017),p.72-81.

[6] Ren H , Wang W , Liu C . Recognizing Online Handwritten Chinese Characters Using RNNs with New Computing Architectures[J]. Pattern Recognition,Vol.93(2019),p.179-192.

[7] WangXiao,ZhangXiaFen,HanDeZhi.Calligraphy Style Identification Based on Visual Features[J]. Modern computer,Vol.21(2016)No9,p.39-46.

[8] Jiulong Zhang,Lu Wang,Xinran Wen. Combination of GIST and PHOG Features for Calligraphy Styles Classification[P]. Multimedia Systems and Signal Processing,2019,p.21-24.

[9] Mao Tianjiao,Wu Jiangqin,Gao Pengcheng,Xia Yang,Lin Yuan. Calligraphy Word Style Recognition by KNN Based Feature Library Filtering[P]. Proceedings of 3rd International Conference on Multimedia Technology(ICMT-13),2013.

[10] Gao Pengcheng,Gu Gang,Wu Jiangqin,Wei Baogang. Chinese calligraphic style representation for recognition[J]. International Journal on Document Analysis and Recognition (IJDAR),2017,20(1).

[11] Jiulong Zhang,Mingtao Guo,Jianping Fan. A novel CNN structure for fine-grained classification of Chinese calligraphy styles[J]. International Journal on Document Analysis and Recognition (IJDAR),2019,22(2).

[12] Lecun Y , Bottou L . Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[13] Yuan Lin,Jiangqin Wu,Pengcheng Gao,Yang Xia,Tianjiao Mao. LSH-based large scale chinese calligraphic character recognition[P]. Digital libraries,2013.

[14] YangZhenZhen,KuangNan,FanLu,KangBing.Review of Image Classification Algorithms Based on Convolutional Neural Networks[J].Signal processing,Vol.34(2018)No12,p.1474-1489.

[15] LinJingDong,WuXinYi,CaiYi,YinHongPeng.Structure Optimization of Convolutional Neural Networks: A Survey[J].Acta Automatica Sinica,Vol.46(2020)No1,p.24-37.

[16] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014.

[17] Martinez A M , Kak A C . PCA versus LDA[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 23(2):228-233.

[18] Miyao H , Maruyama M . Virtual Example Synthesis Based on PCA for Off-Line Handwritten Character Recognition[C]// Document Analysis Systems VII, 7th International Workshop, DAS 2006, Nelson, New Zealand, February 13-15, 2006, Proceedings. DBLP, 2006.

[19] Zhang D , Chen S , Zhou Z H . A new face recognition method based on SVD perturbation for single example image per person[J]. Applied Mathematics and Computation, 2005, 163(2):895-907.

[20] YangBoXiong,YangYuQi.Research on dimensionality reduction after deep learning image feature extraction using PCA[J].Computer system application,Vol.28(2019)No1,p.279-283.

[21] Classifying Fonts and Calligraphy Styles Using Complex Wavelet Transform[J]. Signal Image & Video Processing, 2015, 9(1):225-234.

[22] ZhouQinShu,ChenJinJie,JiPengFei.SVM-based multi-feature handwritten Chinese character recognition technology[J]. Electronic technology,Vol.29(2016)No8,p.136-139.