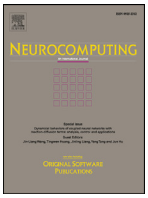
Neurocomputing 330 (2019) 280–286

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



DCT–CNN-based classification method for the Gongbi and Xieyi techniques of Chinese ink-wash paintings

Wei Jiang[a,c], Zheng Wang[a,*], Jesse S. Jin[a], Yahong Han[b], Meijun Sun[b]

[a] School of Computer Software, College of Intelligence and Computing, Tianjin University, Tianjin, China
[b] School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China
[c] School of Computer Information and Engineering, Changzhou Institute of Technology, Changzhou, China

ARTICLE INFO

Article history:
Received 16 March 2018
Revised 30 October 2018
Accepted 2 November 2018
Available online 8 November 2018

Communicated by Prof. X. Gao

Keywords:
Chinese ink-wash paintings
Discrete cosine transformation (DCT)
Convolutional neural networks (CNNs)
Image classification

ABSTRACT

Different from the western paintings, Chinese ink-wash paintings (IWPs) have own distinctive art styles. Furthermore, Chinese IWPs can be divided into two classes, Gongbi (traditional Chinese realistic painting) and Xieyi (freehand style). The extraction of Chinese IWP features with good classification results is challenging because of similar content. This paper presents a novel framework by combining a discrete cosine transformation (DCT) and convolutional neural networks (CNNs). In this framework, a CNN automatically extracts Chinese IWP features from a small subset of the DCT coefficients of an image instead of raw pixels commonly because of its good performance. We evaluate the proposed framework on a dataset including 1400 Chinese IWPs. Experimental results show that the proposed framework achieves competitive classification performance compared to existing benchmark methods.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Chinese ink-wash paintings (IWPs) are important cultural heritages in ancient China, even in old East Asia. As invaluable Chinese IWPs, people seldom saw them before. With the rapid progress in data acquisition, abundant digitized Chinese IWPs can be browsed and retrieved easier from the Internet. The computerized analysis and classification of Chinese IWPs becomes an imperative topic that needs to be researched.

Studies on the art and cultural heritages based on the computational techniques including image processing, feature extraction, pattern recognition, and image classification, have been reported for the western arts in the past decade. Brushstrokes were regarded as an important part of the techniques, which could be used to distinguish painting styles in [1]. Brush strokes were detected by both model-based and semiparametric neural networks. The work of Sablatnig et al. [2] was based on brushstrokes, which characterized the artist personal style. They introduced a hierarchically structured classification scheme to separate the classification into three types of features: color, shape of the region, and structure of brush strokes. To support automatic classification on large western painting image collections, Shen [3] proposed a multiple feature-based framework for classification. By integrating multiple vision features effectively, they improved the accuracy of identification from 52.7% to 69.7% in the Radial Basis Function (RBF) classifier on a dataset of 25 artists. Li et al. [4] proposed a brushstroke style extraction method by exploiting integration of edge detection and clustering-based segmentation to distinguish Van Gogh's paintings from those by his peers.

In recent years, some research of Chinese IWP classification and Chinese artist classification has been reported. Jiang et al. [5] proposed an approach first to distinguish traditional Chinese paintings (TCPs) from non-TCPs using the C4.5 decision tree classifier. After that, some low-level features, such as color histogram, color coherence vectors, autocorrelation (AC) texture features and newly proposed edge-size histogram (ESH), were fed into the support vector machine (SVM). Then TCPs were classified into two classes: Gongbi (traditional Chinese realistic painting) and Xieyi (freehand style). The Xieyi technique is marked by exaggerated forms and freehand brushwork. By comparison, the brushwork in Gongbi paintings is fine and visually complex. Fig. 1 illustrates two examples of the two different techniques, and differences are shown in the local area between two types of paintings. They achieved an accuracy of 85.95% in the first step based on AC features and 94.61% based on AC and ESH features in the final classification. Li and Wang [6] presented a scheme to classify Chinese paintings by using a two-dimensional multiresolution hidden Markov model. Brushstroke style features extracted from different artists were considered as the most representative. A classification rate of 80% was

[*] Corresponding author.
E-mail addresses: jiangweitju@163.com (W. Jiang), wzheng@tju.edu.cn (Z. Wang), jinsheng@tju.edu.cn (J.S. Jin), yahong@tju.edu.cn (Y. Han), sunmj@tju.edu.cn (M. Sun).

https://doi.org/10.1016/j.neucom.2018.11.003
0925-2312/© 2018 Elsevier B.V. All rights reserved.

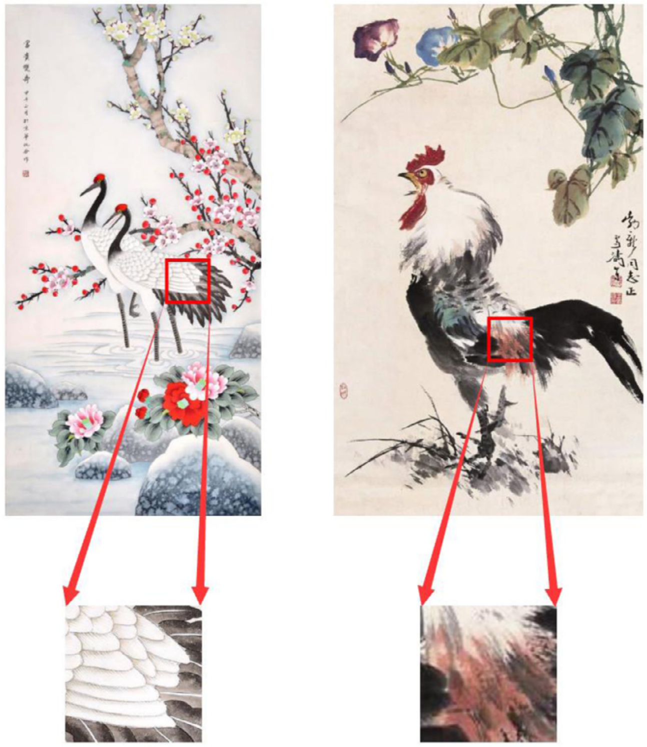


Fig. 1. Illustration of two examples of different techniques: the first row shows a Xieyi painting (left) and a Gongbi painting (right), and the second row shows the difference in the local area between the two types of paintings.

obtained on the paintings of two artists. A precision rate of 62% was obtained on a dataset of five artists. In [7], the histogram-based local feature and the global feature to characterize different aspects of art styles were fed into neural networks to complete the classification. A windowed and entropy-balanced fusion scheme was proposed to make integrated decisions to optimize the classification results. They reported the average precision rate of 96% on a dataset of two artists, 92% of four artists, and 88% of five artists. Sun et al. [8] proposed a novel stroke-based sparse hybrid CNN method for author classification of IWPs. Using a dataset of 120 IWPs from six famous artists, the proposed method achieved successful classification results in comparison to two other state-of-the-art approaches. Besides the aforementioned features applied to classify paintings, other features, such as scale-invariant transformation (SIFT), histogram of gradient (HoG), Wavelet transformation, and bag of features (BoF), were often fed into the classifier to complete the recognition and image classification before.

One of the motivations of this work is that many researchers have focused on the deep learning method because of its good performance for automatic learning good representations of the underlying distribution of raw data [9–18]. Zhu et al. [19] proposed an approach using deep neural networks, which exploited object segmentation and improved the accuracy of object detection. The region-based deep learning network methods achieved state-of-the-art object detection accuracy and improved training and testing speed on benchmark datasets for object detection [20–23]. Convolutional neural networks (CNNs) have led to a series of breakthroughs in image classification [24–29] and other fields [30–32]. In [26], a hybrid model of integrating a CNN and an SVM was proposed to automatically extract features from raw images and complete the predictions. The best recognition rate of 99.81% without rejection was achieved on the MNIST digit database. Krizhevsky et al. [24] proposed a large, deep CNN, which has 60 million parameters and 650,000 neurons and consists of five convolutional layers. The large CNN model demonstrated impressive classification performance based on some tricks on the ImageNet benchmark. Zeiler and Fergus [27] introduced a visualization technique to expose the function of intermediate layers and the operation of the classifier. Tan et al. [28] proposed a novel photograph aesthetic classifier with a deep and wide CNN for fine-granularity aesthetical quality prediction and achieved better classification accuracy than state-of-the-art methods. Yu et al. [29] introduced a flexible deep CNN model, called local-global-CNN, to improve multilabel image classification performance on Pascal VOC2007 and VOC2012 multilabel image datasets. In our work, a deep CNN is proposed to extract good features from Chinese IWPs. As known, in order to extract good features from high-dimensional raw images, a large and complex deep learning model with more parameters and neurons must be built. Furthermore, it will be time-consuming to train a large model, and experiment data must be big enough to prevent overfitting.

The second motivation of our work is to reduce the dimensionality of raw images. Discrete cosine transforms (DCTs) [33] are used to reduce raw image redundancy in this study. A few DCT coefficients are fed into a CNN for classification instead of a raw image. The experiment has demonstrated that only a subset of the DCT coefficients is necessary to preserve the most important data information. As mentioned above, brushstrokes are the most important feature representation of art styles [1,2,4]. Edges detected in an image are the locations of the brushstrokes. In our work, before using a DCT, we perform edge detection by using the Sobel operator.

Based on the analysis and investigation above, we designed a DCT–CNN-based classification method for the Gongbi and Xieyi techniques of Chinese IWPs in this study. In comparison with the existing research on Chinese IWP recognition and classification, our contributions can be highlighted below and detailed in the next section.

- Different from commonly used handcrafted feature extraction, such as color, texture, etc., brushstrokes are extracted from Chinese IWPs, and a well-designed CNN model is utilized as the feature extractor from the brushstrokes because of its ability to extract good features automatically.
- A hybrid model comprising a CNN model and an SVM, which acts as a final classifier because of its good classification performance, is proposed to achieve better results for prediction of Chinese IWPs.
- Taking full advantage of the DCT ability to extract the most important data information, a subset of the DCT coefficients is fed into the CNN instead of the whole data to reduce the model complexity and the training time.

The remainder of this paper is organized as follows. Section 2 presents details of the proposed framework. Experimental results and analysis are described in Section 3. In Section 4, conclusions are drawn and discussion on the future work is given.

2. DCT–CNN-based classification model

In order to take the advantages of computing efficiency of DCTs and automatic feature extraction of the CNNs, the proposed classification model is a combination of a DCT and a CNN. The whole classification procedure contains five steps. For a given image, at the initial stage of the process, an edge detection algorithm is adopted to identify brushstroke features by using the Sobel operation. After that, the morphological operation is performed to complete nearly enclosed edges [4]. Next, a few number of the coefficients are obtained to reduce information redundancy by the use of a DCT. Then the good-feature representation is extracted by feeding the limited coefficients to a CNN. Finally, the features are input to the SVM classifier to complete the classification task.

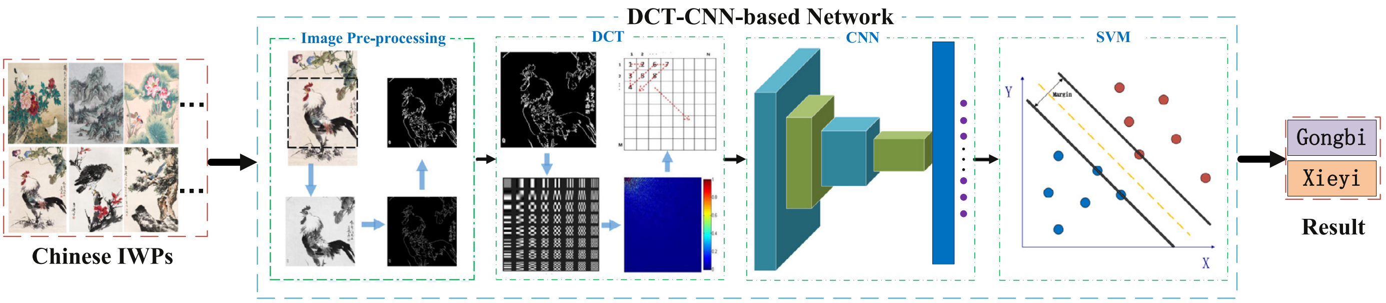


Fig. 2. Proposed classification model of our work.

The DCT–CNN-based classification model of our work is shown in Fig. 2.

2.1. Image pre-processing

As mentioned in [1,2,4], the brushstroke style lies in the detected edges. To find the representation-painting strokes, in this stage, we propose a simple pre-processing method to convert the input image to the grayscale one, because most Chinese IWPs do not have color or even tones. The Sobel edge detection algorithm is applied to the grayscale image to identify the edge pixels.

The edge line detected from the image around a brushstroke may not be like a brushstroke, because perhaps the edge is not completely sharp and broken. To resolve this problem, the morphological operations are developed to complete nearly enclosed edges. Fig. 3 illustrates an example of the procedure of image pre-processing and shows the difference in the local area between the Gongbi and Xieyi paintings after image pre-processing. The Gongbi style, as the name implies in Chinese, is to depict objects with neat and rigorous painting techniques. In contrast, the Xieyi style requires the use of extensive and concise brushstroke and ink to draw the object shape to express the author's artistic conception. As shown in Fig. 3, different from the Xieyi paintings, the Gongbi paintings have longer length and smoother edges because of their different painting styles. Moreover, an enclosed edge is not regarded as a brushstroke if its size is too small or too large. Therefore, we select an edge as a brushstroke if the number of pixels that the edge contains is between two thresholds, 200 and 900, in our model.

2.2. Two-dimensional discrete cosine transform

Because of a large computation cost and long-time processing, the dimensionality of an image with a large size should be reduced. To resolve this problem, the principal component analysis (PCA) and DCT are often used.

The PCA is a useful dimensionality reduction technique which is widely used in various fields such as face recognition and image compression. The PCA is a prevalent technique to find simple patterns behind the complex data.

The DCT is used to transform a signal or a picture from the spatial domain to the frequency domain [34]. The DCT matrix coefficients are data-independent, while the PCA is data-dependent. In this study, the DCT is chosen as the transformation method.

The DCT is defined as:

$$C(u,v)=\alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cos\frac{\pi(2x+1)u}{2N}\cos\frac{\pi(2y+1)v}{2N}. \tag{1}$$

$$St. \qquad 0\le u,v\le N-1.$$

The inverse DCT is defined as:

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\alpha(u)\alpha(v)\ C(u,v)\cos\frac{\pi(2x+1)u}{2N}\cos\frac{\pi(2y+1)v}{2N}. \tag{2}$$

$$St. \qquad 0\le x,y\le N-1.$$

$$\text{Where } \alpha(u)=\begin{cases}\sqrt{\frac{1}{N}} & u=0\\ \sqrt{\frac{2}{N}} & u\neq 0.\end{cases}$$

Only a limited amount of significant DCT low-frequency coefficients, which have higher energy value, are reserved by the $Zig-zag$ scan of the two-dimensional (2D) DCT coefficients. The function of the $Zig-zag$ method allows the recovery of these data in order of decreasing energy.

2.3. Convolution neural network and support vector machine classifier

In this study, after the DCT, the data is fed into a CNN to extract the deep learning feature. The architecture of our five-layer ConvNet model is shown in Fig. 4.

The data with size of 64 × 64 is presented as the input. In layer C1, the first convolutional layer filters the input data with six kernels of size 5 × 5, producing six maps with size of 60 × 60. In following layer S1, with a subsampling ratio of 2, each map reduces the feature size to 30 × 30 by performing max pooling. In layer C2, the second convolutional layer filters the data with 12 kernels of size 5 × 5, producing 12 maps with size of 26 × 26. Then the 12 maps reduce the feature size to 13 × 13 in layer S2. In the last layer, a full connection layer is designed by feeding a 2028-dimensional vector from layer S2. Finally, a 1014-dimensional feature is obtained in the output layer.

Instead of using the CNN classifier, the output from the CNN last layer is fed into the final SVM classifier in this study. The SVM uses the output as a new feature vector for training and does the final classification.

The SVM is used to perform the classification in our work since it is confirmed to be effective. The basic idea of the SVM is to convert a nonlinear separable problem into a linear separable problem by searching an optimal hyper-plane. The optimal solution is to maximize the distance of each class from the hyper-plane. The SVM was proposed for a two-class classification problem originally.

Giving the training data (x_i, y_i) of m instances, it can be defined as:

Training data:$\{(x_i,y_i)\mid x_i\in R^N, y_i\in\{-1,1\}\}$, $St.$ $i=1,2,3\ldots,m$. where x_i is from the N-dimensional feature space X and y_i indicates the class, to which the corresponding x_i belongs.

The goal of the SVM is to search the optimal hyper-plane by maximizing the width of the margin between the two classes. The

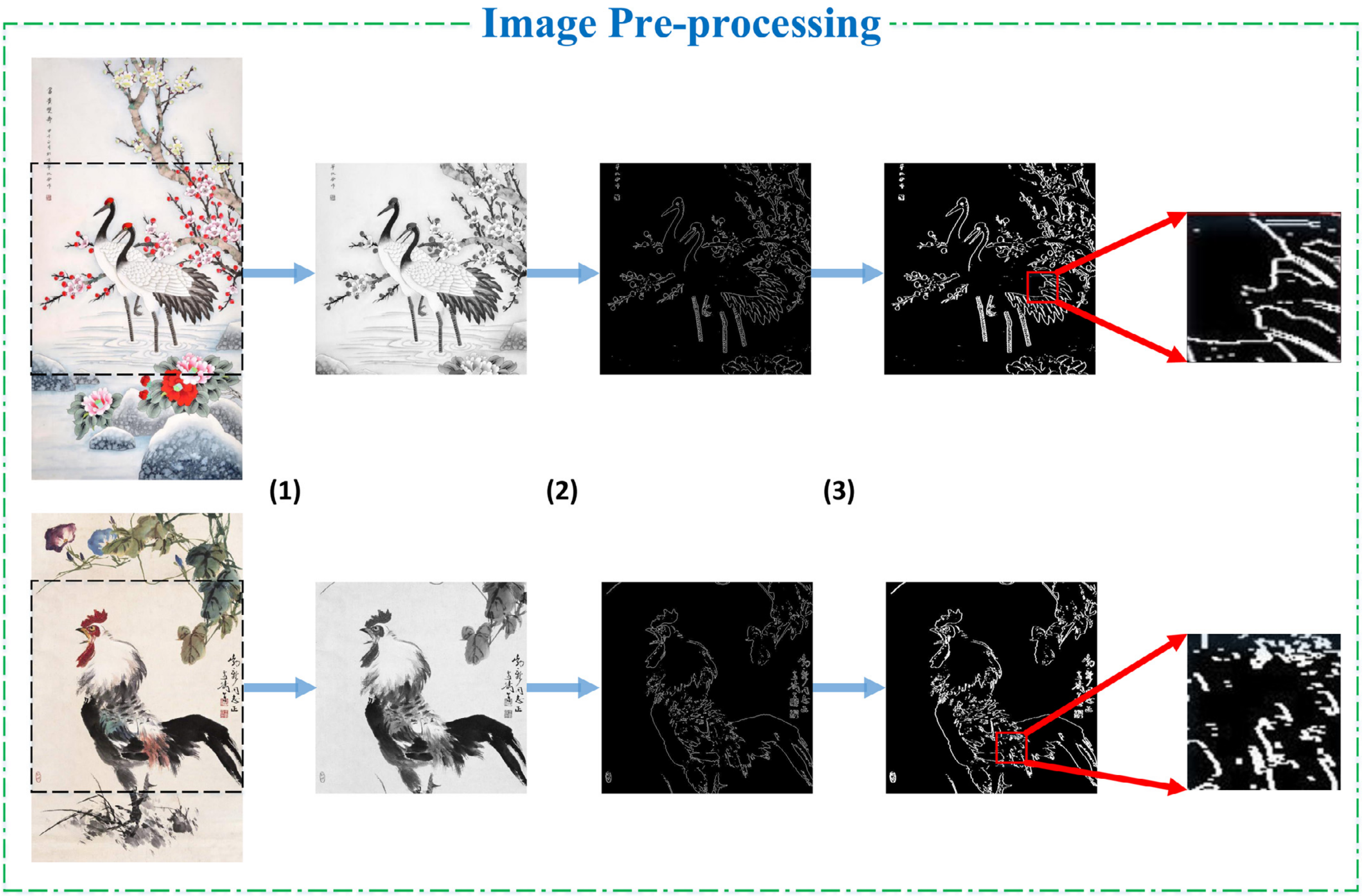


Fig. 3. Illustration of an example of the procedure of image pre-processing: (1) convert to gray scale with the same size; (2) Sobel edge detection; (3) morphological operation. The fifth column shows the difference in the local area between the Gongbi and Xieyi paintings after image pre-processing.

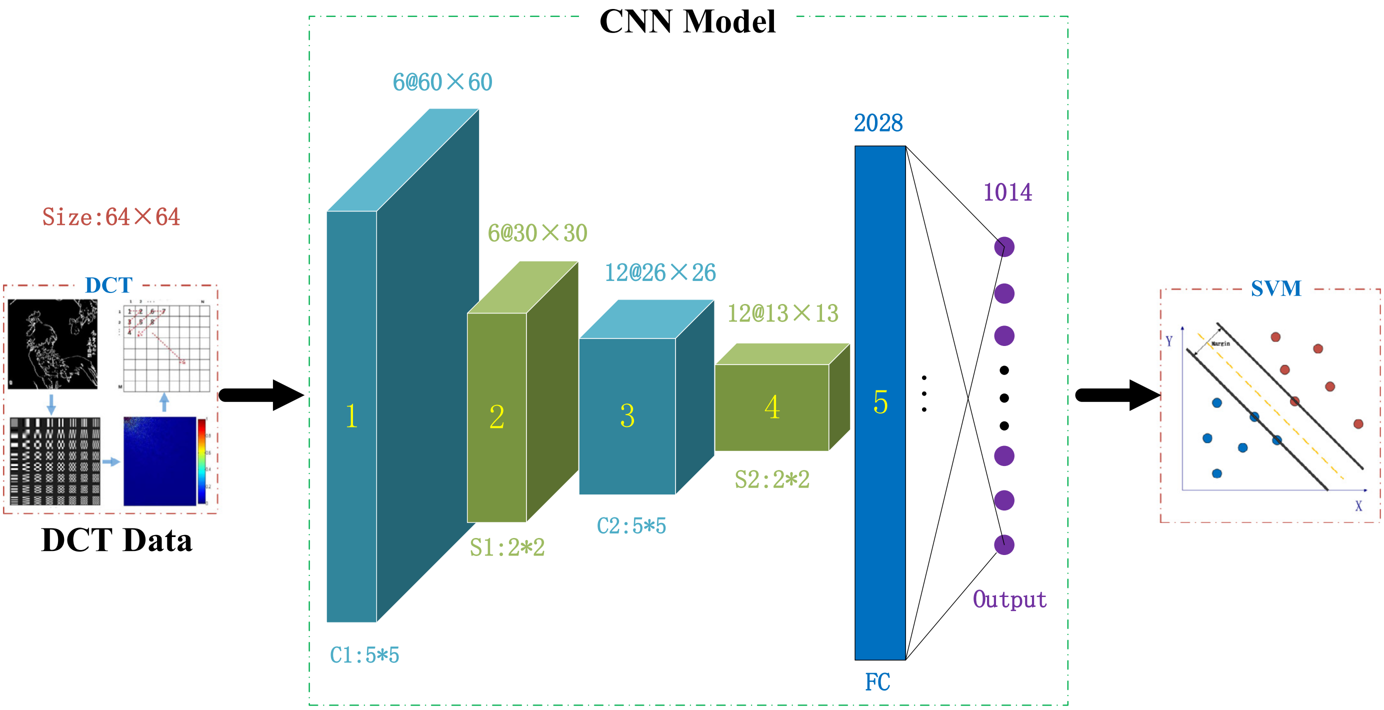


Fig. 4. Architecture of our CNN model presented in the paper.

Table 1
Comparison of the classification results of different classifiers.

Accuracy/Classifiers	ID3	C4.5	KNN	Naive Bayesian	SVM
$P(G)$	81.14	83.43	82.29	89.14	**95.43**
$P(X)$	80.57	85.71	83.43	91.43	**94.86**
$P(O)$	80.86	84.57	82.86	90.29	**95.15**

hyper-plane is:

$$f(x) = \omega^T \phi(x). \tag{3}$$

While the objective function is:

$$\min_{\omega,\xi,b} \left\{ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i \right\}. \tag{4}$$

where C is the regularization parameter, ξ_i is a slack variable. The objective function above can be reformulated to solve the optimization problem using quadratic programming:

$$\max \sum_{i=1}^{m}\lambda_i - \frac{1}{2}\sum_{i,j=1}^{m}\lambda_i\lambda_j y_i y_j K(x_i, x_j). \tag{5}$$

$$St. \qquad 0 \le \lambda_i \le C, i = 1, 2, 3 \ldots m, \sum_{i=1}^{m}\lambda_i y_i = 0.$$

where $K(x_i, x_j) = x_i \cdot x_y$ is the kernel function. The LIBSVM [35] is used to implement the SVM classification in our work.

3. Experimental result and discussions

3.1. Dataset

In this study, we used a dataset including 1400 Chinese IWPs to evaluate the proposed model. There are 700 Gongbi paintings and 700 Xieyi paintings. In order to verify the superiority of the model we designed, we selected some similar content paintings in the two different types of Chinese IWPs.

In the experiment, the dataset is divided into two parts: three quarters for training and one quarter for testing.

3.2. Classification results and discussions

To evaluate the proposed classification model, we designed an experiment to extract several types of the features mentioned above including SIFT, PHOG, HSV, Wavelet transformation, BOF, and ESH against the feature extracted by the proposed method. Furthermore, we carried out another comparative experiment by adopting several classifiers including ID3, C4.5, KNN, Naive Bayesian, and SVM to achieve better classification performance.

In our paper, the classification accuracy to measure the performance is described as follows:

$$\begin{aligned} P(G) &= C_g/T_g \\ P(X) &= C_x/T_x \\ P(O) &= (P(G) + P(X))/2 \end{aligned} \tag{6}$$

where C_g and C_x indicate the number of correctly classified Gongbi paintings and Xieyi paintings, respectively. T_g and T_x indicate the total number of Gongbi paintings and Xieyi paintings, respectively, and $P(O)$ indicates the final classification accuracy.

These classifiers mentioned above are frequently utilized to perform the prediction in different application fields, and each of the classifiers has its own characteristics. To evaluate the performance of each classifier, the feature extracted by the proposed algorithm is fed into these classifiers. Table 1 illustrates the comparison of the classification results of different classifiers. It can be seen from the Table 1 that the best performance is achieved by using the SVM. Because of its superior classification performance in this experiment, the SVM is adopted as the final classifier in this study.

It is well known that delicate feature extraction is one of the most important factors in the success of a classification system. Different types of feature extraction methods, such as SIFT, PHOG, HOG, WTF, etc., which are compared with the proposed method in our work were confirmed to have good distinguishable representations in special application fields. However, traditional hand-designed feature extraction could not perform well in this study because Chinese IWPs have own special styles different from other images, even from western arts. Actually, as mentioned above in Section 1, brushstrokes have been regarded as one of the most important techniques to distinguish image styles. In order to extract the intrinsic feature representation, a well-designed CNN model is utilized for automatic learning good features from the image edges, which act as brushstrokes. From Table 2, it is reported that the classification performance achieved by the proposed method in our work outperforms the results of different features by using the SVM classifier obviously.

Furthermore, to benchmark the proposed framework, we carried out comparative experiments between the three representative existing techniques, including Li and Wang [6], Jiang et al. [5], and Sun et al. [8]. Another additional method named *RawImageCNN* is also used for comparison with the proposed framework in this study. The *RawImageCNN* method directly extracts features from a raw image using the CNN model, which is described in Fig. 4 in Section 2.3, and then classifies Chinese IWPs with a softmax classifier. The comparison results of the different representative methods published on our dataset are listed in Table 3. As can be seen from the table, the proposed DCT–CNN-based method outperforms the selected benchmarks. The experimental results indicate that the proposed algorithm can extract the feature representation from Chinese IWPs well compared to other benchmark methods. For example, in Jiang et al. [5], several meticulous designed handcraft features and the SVM were used to achieve good classification performance for Chinese IWPs. However, with the advent of deep neural networks, CNN has been proven to acquire discriminative feature representation capabilities and better recognition performance. In the *RawImageCNN* method, the CNN model is adopted to automatically extract discriminative features from Chinese IWPs, and it produces better result than the algorithm in Jiang et al. [5]. Nevertheless, the *RawImageCNN* method is to learn features directly from a raw image, and in our work, the DCT–CNN framework is proposed to extract features from the brushbroke of Chines IWPs instead of a raw image and achieves the best classification performance.

The procedure of deep learning is implemented in Matlab R2017a on a desktop with Intel Core i5-6400 2.7GHz and 8GB RAM.

In the training process of the CNN model, the normalization operation is applied to the input data, and truncated normal distribution strategy is applied to weight initialization. Through an adequate experiment for the model training, the value of the batch-size is set to 50 and the initial learning rate is set to 0.01, which is decreased with the proper decay rate of 0.95 during the training process. The parameters of the proposed model are trained by a back-propagation algorithm with an appropriate epoch value to obtain the best classification accuracy.

Another experiment is designed to make a comparison among the different data sizes in terms of training time and accuracy. The data sizes are decided by using raw pixels or different subsets of the DCT coefficients. By feeding data with different sizes into the CNN, a competitive accuracy is achieved and the training speed is much higher than that of the system by using raw pixels. From

Table 2
Comparison of the classification results of different features by using the SVM.

Accuracy/Features	SIFT	PHOG	HSV	WTF	BOF	ESH	OUR
$P(G)$	79.43	84.57	77.71	82.86	83.43	86.86	**95.43**
$P(X)$	81.14	82.86	79.43	80.57	85.14	85.14	**94.86**
$P(O)$	80.29	83.72	78.57	81.72	84.29	86.00	**95.15**

Table 3
Comparison of the classification results of different methods.

Accuracy	Li and Wang [6]	Jiang et al. [5]	Sun et al. [8]	RawImageCNN	OUR
$P(G)$	92.57	94.29	95.43	93.71	**95.43**
$P(X)$	91.43	90.86	94.29	92	**94.86**
$P(O)$	92	92.58	94.86	92.86	**95.15**

Table 4
Comparison of the training time and accuracy for different data sizes.

Data size	Training set	Batchsize	Epochs	Training time/min	Accuracy
512 × 512	1050	50	10	73.7	94.29
128 × 128	1050	50	10	4.14	93.14
64 × 64	1050	50	10	**1.49**	**95.15**

Table 4, it can be seen that it takes only 1.49 min to train the CNN in our work. The training speed by using the data with size of 64 × 64 is nearly 50 times higher than that by using the original data with size of 512 × 512. The DCT–CNN-based classification model can achieve an accuracy of 95.15% by using the data with size of 64 × 64. As stated from Table 4, it was confirmed through this experiment that choosing a suitable data size can reduce the training time and obtain better classification performance.

As for the SVM classifier, kernel function selection is one of the key factors in the success of a recognition model. Moreover, it is crucial to choose the kernel function parameter and the penalty factor c when performing the classification. Inappropriate kernel function and parameter selection often lead to bad classification performance.

Especially in our experiments, we tested several types of kernel functions, such as RBF, linear kernel, Poly, etc., to find an appropriate kernel function. To the end, the RBF kernel function is adopted in our model, and it is confirmed to provide good classification results. Furthermore, we performed a 10-fold cross validation to find the optimal parameters c in the range of [0.001,10] and σ in the range of [0.01,5]. According to the best classification performance on the dataset, the optimal parameter c is set to 0.56 and σ is set to 2.37 finally in our model.

4. Conclusions and future work

This paper presents a novel framework by combining a DCT and a CNN for Chinese IWP classification. We take the advantages of computing efficiency of DCTs and automatic feature extraction of CNNs. Instead of raw image data, the selected DCT coefficients are fed into the CNN to extract the good features automatically. The proposed framework reduces the computational complexity and provides competitive performance. To evaluate the proposed framework, experiments on a dataset containing 1400 Chinese IWPs are performed to carry out the classification. Experimental results show that the proposed framework can work effectively and achieve better classification accuracy compared to several baseline approaches. In the coming future, one direction of research is to use different deep learning models to automatically extract features. Another direction is to use this framework for different applications to improve the classification performance.

Acknowledgment

The authors wish to acknowledge the support from the National Natural Science Foundation, China, under the grants [61572351], [61772360] , and [61876125]. The support of National Social Science Foundation of Qinghai [15XMZ057] and Application Foundation of Science and Technology Office of Qinghai [2015-ZJ-718].

References

[1] T. Melzer, P. Kammerer, E. Zolda, Stroke detection of brush strokes in portrait miniatures using a semi-parametric and a model based approach, in: Proceedings of the 14th International Conference on Pattern Recognition, 1998, pp. 474–476.
[2] R. Sablatnig, P. Kammerer, E. Zolda, Hierarchical classification of paintings using face- and brush stroke models, in: Proceedings of the International Conference on Pattern Recognition, 1998, pp. 172–174.
[3] J.L. Shen, Stochastic modeling western paintings for effective classification, Pattern Recognit. 42 (2) (2009) 293–301.
[4] J. Li, L. Yao, E. Hendriks, J.Z. Wang, Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated brushstroke extraction, IEEE Trans. Pattern Anal. Mach. Intell. 34 (6) (2012) 1159–1176.
[5] S.Q. Jiang, Q.M. Huang, Q.X. Ye, W. Gao, An effective method to detect and categorize digitized traditional chinese paintings, Pattern Recognit. Lett. 27 (7) (2006) 734–746.
[6] J. Li, J.Z. Wang, Studying digital imagery of ancient paintings by mixtures of stochastic models, IEEE Trans. Image Process. 13 (3) (2004) 340–353.
[7] J.C. Sheng, J.M. Jiang, Recognition of Chinese artists via windowed and entropy balanced fusion in classification of their authored ink and wash paintings (IWPs), Pattern Recognit. 47 (2) (2014) 612–622.
[8] M. Sun, D. Zhang, J. Ren, Z. Wang, Brushstroke based sparse hybrid convolutional neural networks for author classification of chinese ink-wash paintings, in: Proceedings of the IEEE International Conference on Image Processing, 2015, pp. 626–630.
[9] B. Shi, X. Bai, C. Yao, Script identification in the wild via discriminative convolutional neural network, Pattern Recognit. 52 (2016) 448–458.
[10] W. Yin, H. Schütze, Convolutional neural network for paraphrase identification, in: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 901–911.
[11] H. He, K. Gimpel, J. Lin, Multi-perspective sentence similarity modeling with convolutional neural networks, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1576–1586.
[12] Y. Zuo, J. Zeng, M. Gong, L. Jiao, Tag-aware recommender systems based on deep neural networks, Neurocomputing 204 (2016) 51–60.
[13] X. Jiang, Y. Pang, X. Li, J. Pan, Speed up deep neural network based pedestrian detection by sharing features across multi-scale models, Neurocomputing 185 (2015) 163–170. C
[14] B. Chandra, R.K. Sharma, Fast learning in deep neural networks, Neurocomputing 171 (2016) 1205–1215. C

[15] J. Hu, J. Zhang, C. Zhang, J. Wang, A new deep neural network based on a stack of single-hidden-layer feedforward neural networks with randomly fixed hidden neurons, Neurocomputing 171 (2016) 63–72. C

[16] Z. Ren, Y. Deng, Q. Dai, Local visual feature fusion via maximum margin multimodal deep neural network, Neurocomputing 175 (2016) 427–432.

[17] J. Xu, X. Luo, G. Wang, H. Gilmore, A. Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, Neurocomputing 191 (2016) 214–223.

[18] Y. Xu, Y. Han, R. Hong, Q. Tian, Sequential video vlad: training the aggregation locally and temporally, IEEE Trans. Image Process. 27 (10) (2018) 4933–4944.

[19] Y. Zhu, R. Urtasun, R. Salakhutdinov, S. Fidler, segDeepM: Exploiting Segmentation and Context in Deep Neural Networks for Object Detection, 2015,. 84(84):4703-4711.

[20] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014) 580–587.

[21] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2014) 1904–1916.

[22] R. Girshick, Fast R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

[23] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards realtime object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 1137–1149.

[24] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105.

[25] Y. Lecun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (2008) 541–551.

[26] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. Lecun, Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks, 2013,. Eprint Arxiv.

[27] M.D. Zeiler, R. Fergus, Visualizing and Understanding Convolutional Networks, Springer International Publishing, 2013.

[28] Y. Tan, P. Tang, Y. Zhou, W. Luo, Y. Kang, G. Li, Photograph aesthetical evaluation and classification with deep convolutional neural networks, Neurocomputing 228 (2017) 165–175.

[29] Q. Yu, J. Wang, S. Zhang, Y. Gong, J. Zhao, Combining local and global hypotheses in deep neural network for multi-label image classification, Neurocomputing 235 (2017) 38–45.

[30] D. Zeng, K. Liu, S. Lai, G. Zhou, J. Zhao, Relation classification via convolutional deep neural network, in: Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers (COLING), 2014, pp. 2335–2344.

[31] Z. Xie, Z. Zeng, G. Zhou, W. Wang, Topic enhanced deep structured semantic models for knowledge base question answering, Sci. China (Inf. Sci.) 60 (11) (2017) 110103.

[32] M. Sun, Z. Zhou, Q. Hu, Z. Wang, J. Jiang, SG-FCN: a motion and memory-based deep learning model for video saliency detection, IEEE Trans. Cybern. 99 (2018) 1–12.

[33] N. Ahmed, T. Natarajan, K.R. Rao, Discrete cosine transform, IEEE Trans. Comput. c-23 (1) (2010) 90–93.

[34] M.J. Er, W. Chen, S. Wu, High-speed face recognition based on discrete cosine transform and rbf neural networks, IEEE Trans. Neural Netw. 16 (3) (2005) 679–691.

[35] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 389–396.

Wei Jiang received the M.S. degree from Tianjin University, Tianjin, China, in 2005. He is currently a Ph.D. candidate at the High-dimensional Information Processing Laboratory, Tianjin University, China. He is also a teacher of school of computer information and engineering, Changzhou institute of technology, Changzhou, China. His research interests include image processing, computer vision and machine learning.

Zheng Wang (wzheng@tju.edu.cn) received the Ph.D. degree in Computer Science from Tianjin University (TJU), Tianjin, China, in 2009. He is now an associate professor in School of Computer Software, TJU. He once was a visiting scholar of INRIA institute, France, from 2007 to 2008. His current research interests include video analysis, hyperspectral imaging, and computer graphics.

Jesse S.Jin graduated with a B.Eng from Shanghai Jiao Tong University and a Ph.D. from University of Otago, New Zealand. He is a recruit professor in Tianjin University under the China Talent Program. He held the Chair Professor of IT and other academic positions in many universities in Australia. He also chaired the Academic Board of the College of Design and Commerce, and was an independent board member of Raffles University. He has published 343 articles and 13 books. His research interests include image processing,computer vision, multimedia, medical imaging, etc.

Yahong Han received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2012. He is currently a Professor with the School of Computer Science and Technology, Tianjin University, Tianjin, China. From 2014 to 2015, he was a Visiting Scholar with the Prof. B. Yus Group, University of California at Berkeley, Berkeley, CA, USA. His current research interests include multimedia analysis, retrieval, and machine learning.

Meijun Sun (sunmeijun@tju.edu.cn) received the Ph.D. degree in Computer Science from Tianjin University (TJU), Tianjin, China, in 2009. She is now an associate professor in School of Computer Science and Technology, TJU. She once was a visiting scholar of INRIA institute, France, from 2007 to 2008. Her current research interests include computer graphics, hyperspectral imaging, and image processing.