Intelligent Computing Technology and Automation
Z. Hou (Ed.)

doi:10.3233/ATDE231287

Painting Style Recognition and Classification Based on Convolutional Neural Network and Multi-Kernel Learning

Qiang WANG[a,1]
[a] *Hohhot Vocational College, Hohhot, 010051, China*

Abstract. The existing methods of style classification mainly focus on Chinese landscape paintings and Western oil paintings, but their texture and color information are different from those of face paintings. It is impossible to directly use current methods to distinguish the different styles of face paintings. Aiming at the above problems, a classification algorithm of portrait style based on components and features of painting style is proposed. Based on summarizing the traditional image recognition technology, this paper takes VGG-Network as an example to discuss the convolution neural network used in image recognition and the network structure of convolution neural network used to realize image style feature recognition. Then a multi-core learning based image style classification algorithm is put forward, which maps different feature components corresponding to heterogeneous data into different feature spaces using different kernel functions. The experimental results show that the proposed classification method of portrait style has stronger robustness and higher classification accuracy.

Keywords. painting style; CNN; multi-kernel; feature extraction; VGG

1. Introduction

Painters have their own styles in portrait depiction. At present, the recognition of portrait style plays an important role in portrait screening and criminal investigation. Determining the style of portrait to find the true author of famous paintings enables the workers to be engaged in the study of ancient famous paintings in the department of culture and exhibition to accurately know the author of famous pas, so as to adopt the method of conformity with the country. The custodial measures and means prescribed by the family also enable amateurs of painting to have blueprints to watch and copy, so as to learn to create excellent works and inherit the cultural tradition of national excellent calligraphy and painting. It is difficult to convince people with the evidence of rational identification. Therefore, we need to explore a scientific and effective method to quantify the authenticity of portraits. In view of this, through the preprocessing, feature extraction and classification of the portrait works, the authenticity probability of the works can be obtained. Inspired by image recognition based on convolution neural network, a method of image style transformation based on convolution neural network training for image recognition is presented and implemented. The network is used to extract and separate the content information and style information of image through feature extraction. Finally, the experiments show

[1] Corresponding Author: Qiang Wang; Hohhot Vocational College, Hohhot, 010051, China; wangqiang2310@126.com

that such scheme can quickly and effectively distinguish the painting styles of different painters.

2. Feature Extraction of Portrait Style

2.1 Method of Distinguishing the Style of Painting

Painters usually use "brightness" and "brushwork" to distinguish the style of portrait painting. Light falls on objects and forms different kinds of light and shadow. Light and shadow are expressed in different degrees by "brightness". The brightness of pale and bright parts is "high", while that of dark and dark parts is "low", and other parts are between the highest and the lowest brightness. Different painters have different perceptions and understandings of light and shadow. In a picture, the highest brightness is the white of paper. The lowest brightness is arranged by pencil lines, using the darkest color of graphite as far as possible. There are generally two kinds of "brushwork" to describe darkness: "cross-line arrangement" and "tracing method". Almost every portrait has a degree of darkness, but the brushwork used for the degree of darkness is varied. Every painter has his own style for the degree of lightness, just like a personal signature.

2.2 Feature Extraction

(1) Color feature extraction

The color characteristics of a picture can be described by a color histogram. The first is to standardize the color. The quantization of colors depends on the proportion of all pixels in different colors, through which we can see the form and proportion of various color assignments in the whole picture. For three-dimensional space such as RGB and HSV, the color histogram can be well described, especially in three channel spaces. The color histogram has better representation effect, and it can describe the color characteristics of the picture more accurately. The method used in this paper is to use histogram to represent the color details of the picture and the main space is HSV.

Usually, we can get many dimensions. Generally, the values of S, H and V are within 0, 360, [0,1] and [0,1] respectively. In the process of image recognition to obtain features, the space complexity will be very high, which will lead to too long operation time. In order to avoid such situation, the best way is to quantify it. After quantization, the operation time will be greatly reduced. People's eyes can't recognize colors very precisely, so it's not necessary to calculate them one by one. Reduce the amount of data extracted and make quantitative calculation. According to our usual recognition habits, we quantify the hue H, saturation S and brightness V at unequal intervals, which are 16, 4 and 4. Then we have the following equations:

$$H = \begin{cases} 0 \ \ if \ h \in [345, 15] \\ 1 \ \ if \ h \in [15, 25] \\ 2 \ \ if \ h \in [25, 45] \\ \ldots \\ 14 \ \ if \ h \in [316, 330] \\ 15 \ \ if \ h \in [330, 345] \end{cases}$$

$$s=\begin{cases}0 \ \text{if} \ s\in[0,0.15] \\ 1 \ \text{if} \ s\in[0.15,0.4] \\ 2 \ \text{if} \ s\in[0.4,0.75] \\ 3 \ \text{if} \ s\in[0.75,1]\end{cases}$$

$$V=\begin{cases}0 \ \text{if} \ v\in[0,0.15] \\ 1 \ \text{if} \ v\in[0.15,0.4] \\ 2 \ \text{if} \ v\in[0.4,0.75] \\ 3 \ \text{if} \ v\in[0.75,1]\end{cases}$$

(2) Texture feature extraction

Texture is the scene on the surface of an object. Graphics is the intuitive tactile feature of a picture. This feature is independent of a single pixel. Texture feature is the sum of several pixels in a block, and the statistical value is obtained by calculating the sum of them. In the process of image recognition, there are three texture features: roughness, direction and contrast.

In this paper, gray level co-occurrence matrix is adopted to extract texture features. Each pixel represented by gray histogram is independent of each other, and it is self-made during the processing, so texture is needed. Thus, texture studies the data values of correlation between pixels. The number obtained by the gray level co-occurrence matrix (C0-Occurrence/Co-Matrix) is calculated twice. These data are the representation values of texture features. Gray distribution is concrete texture. In other cases, such as the change of a graph, if it is expressed by specific numerical value, it is the texture feature. That is the co-occurrence matrix. i is the starting point, until a fixed position, the distance between the two points is $\Delta=(d_x,d_y)$.

To acquire certain features better, we should choose four angles. On this basis, their symbiotic bureaus are calculated, that is to say, some directions are introduced to represent texture features. Generally speaking, the angle shown in the following figure will be used to calculate some data values from these angles, so as to provide data basis for later image recognition. Certain distance d and certain angel θ are chosen. A point (i,j) is also selected to computed the occurrence times of pixel point at (i,j). Such method of judgment is determined by their epoch. For larger epochs, if the length d is less than its size, it means that most of the pixels have the same gray level, and $p(i,j,d,\theta)$ occupies the larger proportion.

All the data are close to the diagonal line. For small epoch maps, the length of d is the same as the size of epoch, but the gradation of gray level is very large, so the distribution of gray level is discrete. If the roughness is different in different directions, θ on the principal diagonal line is changed continuously. It facilitates the analysis of directionality and the direction of symbiosis can be determined by setting different positions for symbiosis.

3. Painting Style Recognition Based on Convolutional Neural Network

3.1 Image Recognition Based on Convolutional Neural Network

We select 16 layers of VGG-Network for image classification. It should be pointed out that each convolution layer and full connection layer in the network have Re LU. We train VGG-Network with a small batch gradient descent method driven by reverse propagation. The batch size was 256 and the momentum factor was 0.9. The training process is regularized by weight attenuation, and the first two full connection layers are regularized by dropout strategy. The learning rate is set to 0.01 and decreases by factor 10 when the correct rate on the verification set no longer increases. Training will stop after 370000 iterations, and learning shorthand will drop three times. VGG networks converge faster because they have deeper layers and smaller convolution filters lead to implicit regularization and pre-initialization of some layers. In VGG-Network, the initialization of network weights is very important, since bad initialization will make network learning stagnant. We can randomly initialize the network without any pre-training. After the VGG-Network training, over 90% TOP5 accuracy can be achieved on some ILSVRC datasets over the past few years.

3.2 Separation of Content and Style

There is a nonlinear filter banks in each layer of the network and its complexity is increased with the depth of network. Therefore, the input image $\vec{x}$ is encoded in each layer by the filter. There are N_l different filter have N_l feature mapping with the size of M_l, where M_l is the product of length and width of feature mapping. We can represent the response of network on the l_{th} layer as a matrix $F^l (F^l \in R^{N_l \times M_l})$, where F_{ij}^l is the activation value of position j of the i_{th} filter.

To provide a visual representation of the encoded image information in each layer of the network, we can use white noise. The gradient descent method is used to obtain a new image which can match the original image features. We use $\vec{c}$ to denote the original image and $\vec{x}$ to denote the image to be generated; C^l and F^l denote the feature representation on layer l. We can define the variance loss function between two feature representations as follows:

$$L_{content}(\vec{c},\vec{x},l) = \frac{1}{2}\sum_{i,j}(C_{ij}^l - F_{ij}^l)^2 \tag{1}$$

The partial derivative of the activation value on the l_{th} layer of the loss function is

$$\frac{\partial L_{content}}{\partial F_{ij}^l} = \begin{cases} (F^l - C^l)_{ij}, F_{ij}^l > 0 \\ 0 \qquad\qquad\quad , F_{ij}^l \le 0 \end{cases} \tag{2}$$

Thus, the gradient of image_ $\vec{x}$ can be calculated by standard error back propagation. Then we can change the random initialization image $\vec{x}$ until it has the same feature representation as the original image $\vec{c}$ in one layer of the network.

To acquire the rep to obtain the style representation of the input image, we define a feature space in the network to obtain the texture information of the image. This feature space can be established on the filter response of any layer in the network. It is constructed by the relationship between the responses of different filters. This characteristic relation is represented by a Gram matrix G^l ($G^l \in R^{N_l \times N_l}$) , where N_l denotes the product of the product of the length and width of the feature mapping; G_{ij}^l denote the inner product of the i_{th} mapping and the j_{th} mapping after vectorization , that is

$$\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2}((F^l)^T(G^l - S^l)_{ji,} F_{ij}^l > 0 \\ 0 \qquad\qquad\qquad\qquad\quad , F_{ij}^l \le 0 \end{cases} \tag{3}$$

The gradient of image $\vec{x}$ can be easily calculated by standard error back propagation.

3.3 Multi-core Learning of Portrait Style

By mapping the different components or feature components of the portrait data into the corresponding kernel function, these data can be better expressed in the new component space or feature space, and the recognition rate of portrait style can be higher. Related researches show that using multiple kernels instead of single kernels can enhance the interpretability of decision function and it also improves the performance of decision function.

A simple method is to express the kernel function $K(x, x_i)$ as a convex combination of M bases:

$$K(x, x_i) = \sum_{m=1}^{M} d_m K_m(x, x_i), \; d_m \ge 0, \sum_{m=1}^{M} d_m = 1 \tag{4}$$

where d_m is the weight of each base kernel $K_m(x, x_i)$.

In this paper, we use the simple multiple kernel learning (SimpleMILL) method to solve the problem of multi-core by seeking the decision of equation

$$f(x) + b = \sum_{m=1}^{M} f_m(x) + b \tag{5}$$

where each function $f_m(x)$ corresponds to a K_m and belongs to different RKHS H_m . Then the problem is converted to quadratic programming problem:

$$\min_{\{f_m\}, b, \xi, d} \frac{1}{2} \sum_{m=1}^{M} \frac{1}{d_m} \| f_m \|_{H_m}^2 + C \sum_{i=1}^{l} \xi_i \tag{6}$$

$$\begin{aligned} s.t.\quad & \forall i, y_i \sum_{m=1}^{M} f_m(x_i) + x_i b \geq 1 - \xi_i \\ & \forall i, \xi_i \geq 0, \\ & \forall m, \sum_{m=1}^{M} d_m = 1, d_m \geq 0, \end{aligned} \tag{7}$$

The above problem can be converted to the following similar model:

$$J(d) = \begin{cases} \min\limits_{\{f_m\},b,\xi,d} \dfrac{1}{2} \sum_{m=1}^{M} \dfrac{1}{d_m} \| f_m \|_{H_m}^2 + C \sum_{i=1}^{l} \xi_i, \xi_i \geq 0 \\ s.t.\ \ y_i \sum_{m=1}^{M} f_m(x_i) + x_i b \geq 1 - \xi_i, \xi_i \geq 0, \forall i \end{cases} \tag{8}$$

The whole framework of the algorithm is shown in figure 1. The image is divided into five parts, and four features are extracted from each part. The features of five parts are input into multi-core learning to classify and get the class label to determine the style of the image.

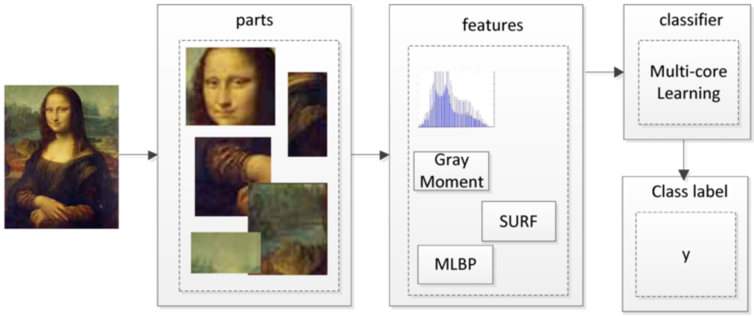

Figure 1. Framework of the proposed algorithm.

4. Analysis of Experimental Results

The hardware environment of the simulation platform is: CPU Intel with strong quad-core X3220, 2.4 GHz, 16G memory PC; software development tools are: Windows 7 operating system, MATLAB R2017a. In the experiment, 100 of Morandi's works and 50 of his counterfeits are selected. After nerve recognition training, 80 works and 50 of his counterfeits were selected to test the availability of the system. For each image, 49 data of color, texture and shape are extracted. Taking color and texture layout as a whole, eight data of overall style of oil painting are calculated on the basis of histogram. Based on gray level co-occurrence moments, 35 data of overall style of oil painting are calculated. The extracted artist's detailed stroke information and stroke features are six local features, and the geometric features of the region are two-dimensional. The authenticity is shown in Figure 2 and the counterfeit is shown in figure 3, which are selected to verify the authenticity of the system.

Figure 2. Genuine product of Morandi's work.

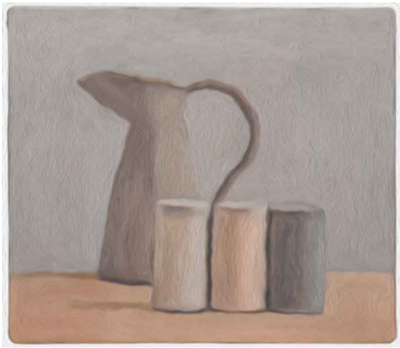

Figure 3. Forgery product of Morandi's work.

We compare four classifiers, KNN, Adaboost, SVM and CNN-SVM, to classify the facial portrait styles. Firstly, four features are extracted from the five parts of the portrait. For the four classifiers based on KNN, Adaboost, SVM and CNN-SVM, the input vectors are drawn into column vectors according to the four features of the five components. For the portrait style classification method based on multi-core learning model proposed in this paper, the input vectors are composed in the following ways. Two kinds: one is to draw column vectors according to the order of four features of five components, and the other is to draw column vectors according to the order of five components of four features. The parameters of the classification method of portrait style based on KNN, Adaboost, SVM and CNN -SVM are set as above. On the two sets of data in VIPSL database, the experimental results of the classification method of portrait style based on the above five classifiers are shown in table 1. In this paper, we propose a multi-core learning-based image style classification method with the highest accuracy.

Table 1. Accuracy of Classification of Portrait Style of different classifiers

Group	KNN	Adaboost	SVM	CNN-SVM	Multi-core Learning
1	65	60	84	94	95
92	64	56	79	87	98

In multi-core learning model, different kernel functions map heterogeneous data composed of different features into different feature spaces, which makes heterogeneous data better represented in the combination space composed of different feature spaces. The classification method of portrait style based on CNN-SVM reduces the classification error by using different component classifiers. By eliminating redundant individual component classifiers which have poor prediction ability, the generalization ability and prediction ability of the integrated component classifier are enhanced to a certain extent, but they can not be fundamentally improved. To solve the problem, the data with historical defects are processed. Therefore, the classification accuracy of the portrait style classification method based on CNN-SVM is slightly lower than that proposed in this chapter. The method of portrait style based on Adapoost has the phenomenon of over-fitting, and the result of classification accuracy is the lowest

5. Conclusion

Portrait style recognition is widely used in the field of famous paintings screening and criminal investigation. This paper proposes a recognition algorithm of portrait style based on multi-core learning. Firstly, according to the method of artistic critics to identify portrait style from the processing of portrait components, five parts are extracted from the portrait. Then, according to the brightness and shade of the portrait, the artist extracts five parts from the portrait. The method of recognizing portrait style by brush drawing method like the author extracts gray histogram feature, gray moment feature, fast robust feature and multi-scale local binary pattern feature from each component, and presents and implements a convolutional neural network algorithm for image style feature analysis based on training for image recognition. Finally, different components and features are fused by multi-core learning for the recognition of portrait style. The experiments show that the proposed algorithm performs well and achieves high recognition rate.

References

[1] Sun J, Xiao Z, Xie Y. Automatic multi-fault recognition in TFDS based on convolutional neural network. Neurocomputing, 2017, 222:127-136

[2] Avila F, Mora, Marco, Fredes, Claudio. A method to estimate Grape Phenolic Maturity based on seed images. Computers & Electronics in Agriculture, 2014, 101(1):76-83

[3] Chen Y, Jiang, Hanlu, Li, Chunyang, et al. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. IEEE Transactions on Geoscience & Remote Sensing, 2016, 54(10):6232-6251

[4] Sun J, Xu Guangluan, Ren Wenjuan, et al. Radar emitter classification based on unidimensional convolutional neural network. Iet Radar Sonar Navigation, 2018, 12(8):862-867.