Research on Painting Image Classification Based on Transfer Learning and Feature Fusion

Qian Yong

Dankook University, 152 Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do 19860, Republic of Korea

Correspondence should be addressed to Qian Yong; 2020020198@stu.cdut.edu.cn

Received 12 February 2022; Revised 21 March 2022; Accepted 24 March 2022; Published 15 June 2022

In order to effectively solve the problems of high error rate, long time consuming, and low accuracy of feature extraction in current painting image classification methods, a painting image classification method based on transfer learning and feature fusion was proposed. The global characteristics of the painting picture, such as color, texture, and form, are extracted. The SIFT method is used to extract the painting's local features, and the global and local characteristics are normalized and merged. The painting images are preliminarily classified using the result of feature fusion, the deterministic and non-deterministic samples are divided, and the estimated Gaussian model parameters are transferred to the target domain via a transfer learning algorithm to alter the distribution of nondeterministic samples, completing the painting image classification. Experimental results show that the proposed method has a low error rate and low feature extraction time and a high accuracy rate of painting image classification.

1. Introduction

With the continuous improvement of digital technology, more and more paintings are stored in the form of video, image, and multimedia, and anyone can use the Internet to retrieve and view paintings, which provides great convenience for people [1, 2]. The computer can properly assess the painting’s distinctive characteristics, analyse, and quantify them, accurately determine its true worth and author, and even undertake an in-depth interpretation of its creative style [3]. With the fast expansion of social economics, relevant specialists and researchers have placed a greater emphasis on the categorization and identification of painting pictures in order to assess their aesthetic quality and worth, among other things. As a result, it is critical to investigate a categorization system for painting pictures [4].

For the above problems, there have been many excellent works. For example, reference [5] proposed a Chinese painting image classification method based on convolutional neural network. The design idea of this method is mainly to solve the problem that the traditional classification method has complicated feature extraction steps of painting images, which leads to the increase in feature extraction time. Thus, the benefits of SoftSign and ReLU activation functions are combined to create a new activation function, which is then used to the convolutional neural network training process. Convolutional neural networks are used to categorise photographs of Chinese paintings. However, it is discovered that this method’s classification accuracy is low in practise, and its impact is poor. Reference [6] presented a technique for classifying styles of painting based on information entropy. Using web crawler technology to collect image painting images from the database, the collected images are preprocessed to obtain higher-quality painting images, calculate the image information entropy, and classify the style of painting using a combination of the information entropy and support vector machine (SVM), but this method is too complex, and painting image feature extraction takes too long.

However, the above methods have the problems of high error rate of feature extraction, long extraction time, and low accuracy of painting image classification. Therefore, this paper proposes a new painting image classification method based on transfer learning and feature fusion.
2. Design of Painting Image Classification Method

Web crawler refers to the crawling of security rules and related data to the Web or database. Therefore, in order to improve the acquisition speed of painting images, this paper sets crawler rules and uses web crawler to capture painting images in the image database to improve the subsequent image feature extraction and classification rendering speed.

2.1. Image Feature Extraction and Fusion

2.1.1. Global Feature Extraction. Color features: HSV color space corresponds to three visual features of the human eyes, respectively. The hue range is 0–360°, which corresponds to the actual color [7, 8]. Saturation S is proportional to the degree of brightness of a color, and brightness is proportional to the ability of light. The more the energy is, the brighter the light is [9]. HSV is more closely related to human perception and experience of color than the RGB system and is thus more extensively utilised. Therefore, this paper adopts HSV color space to extract color features of painting images. This model is shown in Figure 1.

$$E_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij},$$
$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left( p_{ij} - E_i \right)^2},$$
$$s_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left( p_{ij} - E_i \right)^3},$$

where \( i \) represents the total amount of color channel of the painted image, and \( p_{ij} \) represents the color value of the \( i \)-th color channel of the \( j \)-th pixel [10].

Texture features: Wavelet analysis is the most widely used method in extracting image texture features [11]. This method is very stable in signal processing and has shown good results in many research fields, with broad application prospects. The principle of texture feature extraction using wavelet transform is shown in Figure 2.

The classical wavelet functions include Haar, Morlet, dbn, and sym N. The following is a brief introduction to Haar wavelet functions. The definition is shown in the following formula:

$$\psi(t) = \begin{cases} 1, & 0 \leq t \leq \frac{1}{2} \\ -1, & \frac{1}{2} \leq t \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Let kernel function \( \psi(x) \) be the female wavelet, and a cluster of orthogonal basis can be obtained through translation and scaling [12].

$$\psi_{m,n}(x) = 2^{-m/2} \psi(2^{-m}x - n).$$

Then, the wavelet coefficients of the signal \( f(x) \) can be expressed by the following formula:

$$c_{m,n} = \int_{-\infty}^{\infty} f(x) \psi_{m,n}(s) dx.$$ (4)

If the scale function \( \phi(x) \) satisfies the two-scale difference equation [13], it can be expressed as follows:

$$\phi(x) = \sqrt{2} \sum_k h(k) \phi(2x - k).$$ (5)

Then, the relation between the wavelet kernel function \( \psi(x) \) and the scale function is

$$\psi(x) = \sqrt{2} \sum_k g(k) \phi(2x - k), \quad g(k) = (-1)^k h(1 - k).$$ (6)

Explicit \( \phi(x) \), \( \psi(x) \) are not required for wavelet transformation but depend on \( h(k) \), \( g(k) \). A \( J \)-order wavelet decomposition can be written as follows:

$$f_0(x) = \sum_k c_{0,k}(x) = \sum_k \left( c_{J+1,k} \psi_{J+1,k}(x) + \sum_{j=0}^{J} d_{J+1,k} \psi_{J,1,k}(x) \right).$$

(7)

If the coefficient \( c_{0,k} \) is known, the relationship between the coefficient of grade \( j + 1 \) and that of grade \( j \) is [14]

$$c_{J+1,k} = \sum_k c_{J,k} h(k - 2n), \quad d_{J+1,k} = \sum_k c_{J,k} g(k - 2n).$$

(8)

Combined with the above analysis, the wavelet coefficient \( d_{J,n} \) is used to construct the recursive algorithm, which is described as follows:

$$c_{J,k} = \sum_n c_{J+1,n} h(k - 2n) + \sum_n d_{J+1,n} g(k - 2n).$$

(9)

Let \( \bar{h}(n) \) and \( \bar{g}(n) \) be signals with impulse response and use their high and low filters to sample filter signals and combine them to achieve wavelet transform [15]. \( \bar{h}(n) \) and \( \bar{g}(n) \) are defined as follows:

$$\bar{h}(n) = h(-n), \quad \bar{g}(n) = g(-n).$$

(10)

In two-dimensional signal, the small and medium wave function can be expressed as

$$\begin{cases} \phi(x, y) = \phi(x)\phi(y), \\ \psi^1(x, y) = \phi(x)\psi(y), \\ \psi^2(x, y) = \psi(x)\phi(y), \\ \psi^3(x, y) = \psi(x)\psi(y). \end{cases}$$

(11)

The corresponding two-dimensional filtering function is
The traditional Chinese painting image is decomposed into four components: LL, LH, HL, and HH. LL was decomposed and sampled to obtain second-order wavelet decomposition, as shown in Figure 4 [18]. And, if further decomposition is required to obtain more vector information, the same operation is continued for LL2.

Shape features: Compared with color and texture features, shape features belong to deeper image features. Therefore, in order to improve the extraction accuracy of painting image shape features, this paper uses Fourier shape descriptors [19, 20] to achieve this important goal.

Suppose in XY there is a boundary in the plane, the boundary mainly consists of N coordinates for \((x_n, y_n)\) points, the starting point at this point, moving counterclockwise along the border, will form the multiple trajectories, mainly expressed in \((x_0, y_0), (x_1, y_1), \ldots, (x_{N-1}, y_{N-1})\), at this time will be XY plane and complex plane UV overlap, and redefine the boundary points, as shown in Figure 5.

\[
s(n) = x(n) + jy(n), \quad n = 1, 2, \ldots, N - 1. \tag{13}
\]

The discrete Fourier transform of \(s(n)\) [21, 22] is

\[
S(w) = \frac{1}{N} \sum_{n=0}^{N-1} s(n) \exp\left(-j \frac{2\pi nw}{N}\right), \quad w = 0, 1, \ldots, N - 1.
\tag{14}
\]

The inverse Fourier transform of \(S(w)\) is

\[
S(k) = \sum_{w=0}^{N-1} S(w) \exp\left(-j \frac{2\pi wk}{N}\right), \quad k = 0, 1, \ldots, N - 1.
\tag{15}
\]

If only the first \(P\) coefficients of \(S(w)\) are selected here, namely \(P \leq N - 1\), another similar expression of \(S(k)\) can be derived, namely the Fourier shape descriptor.

\[
\hat{S}(k) = \sum_{w=0}^{P-1} S(w) \exp\left(-j \frac{2\pi wk}{N}\right), \quad k = 0, 1, \ldots, N - 1. \tag{16}
\]

The shape features vector of the painting image is described by the Fourier shape descriptor, and the shape feature of the painting image is extracted.

2.1.2. Local Feature Extraction Based on SIFT Algorithm. Scale-invariant feature transform (SIFT) is a classical shape feature and belongs to a local feature algorithm. Local features extracted have invariance and stability [23]. The specific process is as follows.

Extreme value detection of scale space: The scale-space kernel is defined as follows:
$f_{\text{out}} = K \ast f_{\text{in}}, \quad (17)$

where $\ast$ represents convolution operation, and $K$ represents scale-space kernel. Gaussian difference scale space can be obtained by using this parameter, as follows:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \ast I(x, y).$$

$$= L(x, y, k\sigma) - L(x, y, \sigma). \quad (18)$$

Among them,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-((x^2+y^2))/2\sigma^2}, \quad (19)$$

where $\sigma$ is the standard deviation of the Gaussian kernel, and $(x, y)$ is the spatial coordinate point of an image $I(x, y)$.

Positioning of key points: Key points are selected according to the scale and stability of candidate points. If the value of operation comparison is greater than or less than all adjacent points, the candidate point is identified as the key point [24].

Key point direction assignment: After the key point is found, one or more directions are assigned to its local image gradient. Taylor’s expansion (fitting function) corresponding to the Gaussian function is

$$D(x) = D + \frac{\partial D}{\partial x} x + \frac{1}{2} x^2 \frac{\partial^2 D}{\partial x^2}. \quad (20)$$

For alternative points $D$ and $D'$, $D' = 0$, $x$, and the corresponding $D(x)$ are calculated, and the threshold is set as follows:

$$x = \left[ \frac{\partial^2 D}{\partial x^2} \right]^{-1} \frac{\partial D}{\partial x}.$$

$$D(x) = D + \frac{1}{2} \frac{\partial^2 D}{\partial x^2} x. \quad (21)$$

The Hessian matrix $2 \times 2$ at the “feature point” with constant scale [25] is obtained, and the trace and determinant of the $2 \times 2$-dimensional Hessian matrix $\begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$ are solved.

$$\text{Tr}(H) = D_{xx} + D_{yy} = \alpha + \beta,$$

$$\text{Det}(H) = D_{xx} + D_{yy} - (D_{xy})^2 = \alpha\beta. \quad (22)$$

Discard the edge response point, set $\gamma$ as the ratio of two characteristic values, have $\alpha = \gamma\beta$, then the following relationship exists:

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\alpha + \beta)^2}{r\beta^2} = \frac{(R + 1)^2}{R}. \quad (23)$$

Feature points are judged by the threshold, more representative feature points are selected, and unstable key points are screened.

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} \leq \frac{(R + 1)^2}{R}. \quad (24)$$
Generate descriptors: From the local picture gradient information, a descriptor is constructed for each key point, and its scale is determined in the second step. The critical element is that the algorithm creates a huge number of highly distinctive characteristics in a variety of sizes and places [26]. The gradient’s modulus and direction are as follows:

\[ m(x, y) = \sqrt{(L((x+1), y) - L((x-1), y))^2 + (L((x, (y+1)) - L((x, (y-1)))^2)}, \]

\[ \theta(x, y) = \arctan \left( \frac{L(x, (y+1)) - L(x, (y-1))}{L((x+1), y) - L((x-1), y)} \right). \]  

(25)

The local feature extraction of the painting image is realized by generating descriptors to describe the local feature of the painting image.

2.1.3. Feature Fusion. In the process of painting image classification, not all feature descriptors have the same classification and discrimination ability for each category. For example, texture features can be used to classify different leaves well, but color features cannot achieve the same effect. Considering that texture features and shape features have different contributions to the classification of debris in different parts, multifeature fusion is needed to highlight the influence of different image features in different parts.

Given a target image, its global feature vector \( C_q = [C_q(1), C_q(2), \ldots, C_q(m)] \) and local feature vector \( M_q = [M_q(1), M_q(2), \ldots, M_q(m)] \) are extracted where \( m \) and \( n \) represent the dimensions of the two feature vectors, respectively.

Due to the different dimensions of global feature and local feature algorithms, there will be differences when using similarity measurement algorithms. In order to avoid this trouble, \( C_q \) and \( M_q \) need to be normalized, and the formula is as follows:

\[ D_{C_i}'(i) = \frac{D_{C_i}(i)}{\max(D_{C_i}) - \min(D_{C_i})}, \quad i = 1, 2, \ldots, N, \]

\[ D_{M_i}'(i) = \frac{D_{M_i}(i)}{\max(D_{M_i}) - \min(D_{M_i})}, \quad i = 1, 2, \ldots, M, \]  

(26)

where \( D_{C_i}(i) \) and \( D_{M_i}(i) \) represent the distance between the original global feature and local feature and the corresponding feature of the sample image. \( D_{C_i}'(i) \) and \( D_{M_i}'(i) \) represent the normalized results of \( D_{C_i}(i) \) and \( D_{M_i}(i) \), and \( N \) represents the number of images stored in the image database.

Let \( Q_q \) represents the features after fusion of the two features, \( w_1 \) is the weight of global features, and \( w_2 \) is the weight of local features, and then, the features after fusion can be expressed by the following formula:

\[ Q_q = w_1 D_{C_i}'(i) + w_2 D_{M_i}'(i). \]  

(27)

2.2. Painting Image Classification Based on Transfer Learning.

Transfer learning is a kind of machine learning algorithm whose primary objective is to derive new domain knowledge from previously acquired domain information. By using current information in the source domain to acquire unknown knowledge in the target domain, the acquisition of target domain knowledge is no longer a zero-sum game and may significantly boost the target domain’s learning impact. The transfer learning algorithm’s fundamental flow chart is shown in Figure 6.

Based on the result of feature fusion, the painting images are preliminarily classified, the deterministic samples and the nondeterministic samples are divided, and the parameters of the Gaussian model estimated are transferred to the target domain by the transfer learning algorithm to change the distribution of nondeterministic samples, so as to obtain accurate painting image classification results.

Assuming that the painting image sample to be classified is called the target domain \( D_T \), the specific process of dividing \( D_T \) is as follows:

Step 1 K-means clustering algorithm is used to cluster sample \( D_T \) in the target domain to obtain different painting image sample category labels;

Step 2 Assume that the purpose of painting image classification is to divide paintings into \( k \) categories, the label vectors of two different classification results are \( \lambda_a \) and \( \lambda_b \), and the results of painting image classification are \( [C_{a1}, C_{a2}, \ldots, C_{ak}] \) and \( [C_{b1}, C_{b2}, \ldots, C_{bk}] \). Then, a symmetric matrix \( S \) with a size of \( k \times k \) can be obtained, and the value in it represents the number of overlaps between classes. When the value is the largest, the category labels of the two classes are the same, and the matching results of all classes can be obtained after repeated for many times.

Step 3 After the successful matching of category labels, the clustering consistency value \( CI \) is introduced for the effective discriminative sample \( D_A \) and nondeterministic sample \( D_B \), which is mainly calculated by the following formula:

\[ CI(x) = \frac{1}{H} \max \left\{ \sum_{t=1}^{H} \delta(\pi_t(x), L) \right\}, \]  

(28)

where \( H \) represents the number of clustering, \( \pi_t(x) \) represents the classification result of painting image sample \( x \) in the \( t \)-th classification, and \( L \) represents each category identification.

Step 4 Set a threshold value of \( Q \), and take samples of \( CI \geq Q \) as deterministic samples \( D_A \), and samples of \( CI < Q \) as nondeterministic samples \( D_B \).
The new Gaussian mixture model parameters $\alpha_{\text{new}}, \mu_{\text{new}},$ and $\sigma_{\text{new}}$ were used to calculate the membership function $Z_{i,j}$ of the target domain $D_T$, thus completing the task of migrating the parameters of sample points in the source domain to the target domain.

Step 4: substitute the new parameter into the Gaussian distribution formula to obtain the probability value of each sample in $D_B$, so as to obtain the classification result of $D_B$ images, namely the classification result of painting images.

The specific implementation steps of painting image classification are shown in Figure 7.

1. Determine the number of deterministic and non-deterministic samples based on feature fusion results and cluster consistency values
2. Determine the number of deterministic and non-deterministic samples based on feature fusion results and cluster consistency values
3. Determine the number of deterministic and non-deterministic samples based on feature fusion results and cluster consistency values
4. The source domain is picked at random from the remaining bands of the data, and the GMM parameters in the source domain are estimated
5. The calculated parameters are transferred to the target domain to modify the distribution of the nondeterministic sample
6. After calculating the final clustering result, Dunn’s index was used to generate the clustering validity value, which was then compared to the initial clustering validity value
7. If the clustering effectiveness decreases, repeat the above three steps and then reselect the source region for migration until a good painting image classification effect is achieved.

3. Experimental Design

In order to verify the validity of the painting image classification method based on transfer learning and feature fusion designed in this paper, relevant experimental tests are carried out, and the experimental environmental parameters are shown in Table 1.

In the network, crawler technology is used to crawl 10000 painting images, and all data are denoised. The sample images after denoising are taken as experimental sample images, part of which are shown in Figure 8.

3.1. A Comparison of Feature Extraction Error Rates. Figure 9 compares the feature extraction error rate of painting pictures using the convolutional neural network classification technique, the information entropy-based classification approach, and the classification method based on transfer learning and feature fusion.

Analysis of the data in Figure 9 shows that the error rate of feature extraction of painting image based on convolutional neural network is between 13% and 20%, that of feature extraction of painting image based on information entropy is between -17% and 2%, and that of feature extraction of painting image based on transfer learning and feature fusion is between 2% and 4%. It shows that the feature extraction accuracy of the proposed method is higher, which can lay a foundation for the subsequent feature classification of painting images.

3.2. Feature Extraction Time Comparison. The feature extraction time of painting images based on convolutional neural network classification method, information entropy-based classification method, and classification method based on transfer learning and feature fusion is compared, and the results are shown in Figure 10.

As shown in Figure 10, as the number of tests increases, the time it takes to extract painting attributes from various approaches varies. The classification approach based on convolutional neural networks has a maximum and lowest time of feature extraction of 4.0s and 2.0s, respectively. The greatest and lowest times for feature extraction based on information entropy are, respectively, 3.9s and 1.7s. The maximum time and minimum time of feature extraction of painting image based on the classification method of transfer learning and feature fusion are 0.7s and 0.5s, respectively. Compared with the experimental comparison method, feature extraction of painting image in this paper is shorter and more efficient.

![Diagram](image_url)
3.3. Classification and Comparison of Painting Images.

The accuracy of painting image classification based on convolutional neural network, information entropy, and transfer learning and feature fusion is compared, and the comparison results are shown in Table 2.

By analyzing the comparison results of the accuracy of painting image classification in Table 2, it can be seen that the average accuracy of painting image classification based on convolutional neural network is 84.35%, and that of painting image classification based on information entropy is 77.76%. The average accuracy rate of painting image classification based on transfer learning and feature fusion is 96.72%, indicating that the classification result of this method is more accurate.

Figure 8: Images of some experimental samples.

Figure 9: Comparison of error rates of feature extraction.
4. Conclusion

In the development of human civilization, the painting is the precious spiritual wealth. Through exquisite painting, it can fully display the rich spirit of the era. The present stage of painting gradually moves towards the direction of the digital and intelligent, so a lot of drawing is carried out in the network. The classification of the painting according to the content of the painting style is particularly important. Therefore, this paper introduces transfer learning and feature fusion methods to design a new painting image classification method. Experimental results show that the error rate of feature extraction of painting images in this method ranges from 2% to 4%, the maximum time of feature extraction is 0.7s, the minimum is 0.5s, and the average accuracy rate of painting image classification is 96.72%, indicating that this method can not only achieve accurate and rapid feature extraction of painting images but also achieve the correct classification of painting images. It can lay a solid foundation for the development of painting image classification.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


