International Journal for Housing Science and Its Applications
Publish September 9, 2025. Volume 47, Issue 1 Pages 475-486

https://doi.org/10.70517/ijhsa47140

A Comparative Study of the Similarities and Differences in the Creation of Artistic Impression between Modern and Contemporary Lingnan Landscape Painting and Western Landscape Painting in Different Cultural Contexts

Bing Wu[1,*]
[1] School of Fine Arts, Zhaoqing University, Zhaoqing, Guangdong, 526061, China
Corresponding authors: (e-mail: wuhuashi2000@163.com).

Abstract Due to its unique historical evolution, the painting techniques of the Lingnan region in modern times have been influenced to a certain extent by Western culture. This paper obtained approximately 3,000 images of modern Lingnan landscape paintings and Western landscape paintings through on-site interviews and online collection. The images were rotated, cropped, and scaled to enhance the data representation of the paintings and complete the preprocessing of the research data. In terms of painting image classification, the optimal feature subset was selected based on the CGO optimization algorithm, and the cross-task feature fusion module was used to fuse painting image features. This enabled the construction of a multi-faceted artistic painting classification model, achieving the classification task of different painting types under a unified framework. In terms of painting image emotional classification, self-learning and knowledge transfer techniques based on sparse autoencoders were introduced as methods for painting image emotional semantic analysis and unsupervised feature learning. Combining the image features of modern and contemporary Lingnan landscape paintings and Western landscape paintings, we propose a painting image emotional classification system framework comprising three major modules: source domain local feature learning, target domain global feature extraction, and image emotional classification. This framework is used to construct a painting emotional classification model. The designed painting emotion classification model not only demonstrates emotion classification accuracy significantly higher than similar models (>0.730) but also achieves a classification performance standard deviation <0.010 after oversampling strategies, demonstrating excellent robustness. This provides a robust and effective technical foundation for analyzing the artistic ambiance of painting images.

Index Terms Lingnan landscape painting, Western landscape painting, emotion classification, sparse autoencoder, CGO optimization algorithm

I. Introduction

The beauty of art is not merely confined to superficial formal and visual elements; what is truly significant lies in the deeper "aesthetic resonance." In different paintings, distinct aesthetic resonances are created to convey the painting's theme. Aesthetic resonance is an indispensable element of traditional painting, representing the thoughts and emotions expressed by artists through the use of brushwork and creativity. Traditional painting places great emphasis on the creation of aesthetic resonance, seeking to achieve a spiritual realm mediated through ink and wash [1]-[3]. In landscape and scenic paintings, artists often employ the technique of "capturing the spirit through form," combining elements such as mountains, water, and clouds, and mastering the interplay of light and shadow, texture, and atmosphere to create a profound, vast, and mysterious atmosphere [4], [5].

Lingnan landscape painting is a unique painting style from southern China, characterized by strong regional features and cultural connotations. It typically takes landscapes as its theme, using brushwork and ink to depict natural scenery and humanistic landscapes. Due to Lingnan's unique geographical location, modern and contemporary Lingnan art has had frequent exchanges with Western culture, deeply and widely influenced by foreign cultural influences [6], [7]. The Lingnan School of Painting emerged in Guangdong at the beginning of this century, represented by the "Three Masters of Lingnan." It advocates absorbing the strengths of ancient and modern, Chinese and Western, especially Western painting arts, to reform traditional Chinese painting, guiding it toward modernization, nationalization, and popularization, thereby enhancing the effectiveness of aesthetic education. It is an influential art movement both domestically and internationally, characterized by a synthesis of Eastern and Western styles and the integration of ancient and modern elements [8]–[10]. Western landscape painting emerged during the Renaissance, advocating the use of various techniques to depict natural scenery, seeking more scientific methods, and attempting to more effectively approach nature, with the most important aspect being continuous observation [11], [12]. However, due to differences in historical and cultural backgrounds, the aesthetic tastes and expressive characteristics of the artistic conception in Chinese landscape painting and Western landscape painting have gradually developed their own unique features in the course of historical development [13]. Modern and contemporary Lingnan landscape painting, as an important component of Chinese landscape painting, has developed under the innovative concept of "synthesizing Eastern and Western traditions," distinguishing itself from Western landscape painting. By exploring the differences in the creation of artistic intent between modern and contemporary Lingnan landscape painting and Western landscape painting, this study contributes to the development of both traditions.

Shen[14] analyzed the art of landscape painting, whose artistic creation of aesthetic appeal is centered on the dynamic balance of "conveying the spirit through form," and has evolved over time to incorporate personal life experiences and natural landscapes into the paintings. Wang[15] pointed out that the artistic aesthetic appeal of Chinese painting integrates the thoughts and emotions of the individual creator, blending brushwork and painting scenes to achieve a more authentic and emotionally compelling aesthetic appeal. Xuan[16] analyzed the cultural philosophical differences and historical evolution between Chinese and Western painting. Chinese painting emphasizes the harmonious coexistence of humanity and nature, possesses imaginative space, and is influenced by Daoist, Confucian, and Buddhist philosophy. In contrast, Western painting draws inspiration from religious themes and humanism, grounded in rational observation. Among them, Zhou and Ma[17] pointed out that in Chinese landscape painting, Daoism emphasizes nature, Confucianism emphasizes the harmonious unity of humanity and nature, and Buddhism emphasizes inner peace and transcendence. These are all reflected in the artistic conception and spiritual realm of landscape painting. In contrast, Western traditional culture is independent, progressive, rational, and materialistic, emphasizing logic, epistemology, and methodology, with a more scientific standard, pursuing an artistic realm of realistic representation. In terms of subject matter, their spiritual orientation is inseparable from the struggle against nature, whether in aspects of human habitation, labor, and daily life, or in the conquest of the ocean, emphasizing human dominance [18], [19]. The artistic realm of landscape painting is expressed through vigorous brushwork and vague, subtle forms, conveying a spiritual realm of etherealness, tranquility, and vastness. Through the method of "using form to depict spirit," it creates an artistic realm of "scenes beyond scenes, landscapes beyond landscapes" [20].

This paper first explains the data collection methods used for painting images and the data augmentation techniques employed to prepare the research data. It then analyzes the feature extraction, optimization, and fusion processes of artistic paintings under the CrossVit and CGO optimization algorithms, constructing a multi-faceted artistic painting classification model. Subsequently, it details the operational principles of self-learning and knowledge transfer techniques based on sparse autoencoders, determining the model's emotional semantic analysis and feature learning forms for painting images. Based on this, a painting image emotional classification system scheme is designed, and a painting emotional classification model based on a convolutional sparse autoencoder is established. Subsequently, the performance of the proposed painting emotional classification model under oversampling training strategies is evaluated through comparisons of different painting image styles and similar models. Finally, based on the historical and cultural backgrounds of the Lingnan region and the West, the similarities and unique characteristics of the two types of portraits under different cultural influences are explored.

II. Image acquisition and processing

II. A. Data collection

To obtain more detailed data on modern and contemporary Lingnan landscape paintings and Western landscape paintings, this paper employs methods such as literature research and visiting exhibitions to understand the historical background of modern and contemporary Lingnan landscape paintings and Western landscape paintings. Additionally, a large amount of image resources were collected through field research and electronic resources.

During the field research phase, the development process of modern and contemporary Lingnan landscape paintings was explored in depth, primarily through in-depth interviews with inheritors. The collection of image resources was mainly accomplished by photographing works collected by inheritors. The collection of Western landscape paintings was completed by screening electronic resources from online museums and libraries in multiple Western countries, and the production processes and materials of both were documented in detail. To preserve these data, the collected images were stored in the created image resource database.

Online collection was also an important method for obtaining data on modern and contemporary Lingnan landscape paintings. First, relevant academic papers, research reports, expert opinions, and official websites of cultural institutions were searched online to obtain information on the developmental background of modern and contemporary Lingnan landscape paintings. Then, crawler technology and online resource databases were utilized to collect and organize related image and video materials.

During the data collection process, this study collected a total of 1,589 images of modern and contemporary Lingnan landscape paintings and 1,541 images of Western landscape paintings. These activities provided valuable data sources, facilitating a better comparison and study of the cultural backgrounds and artistic atmospheres created by modern and contemporary Lingnan landscape paintings and Western landscape paintings.

II. B. Image Data Enhancement

Image data augmentation involves applying a series of transformations, occlusions, and noise additions to the original training data to expand the scale and diversity of the training dataset, thereby effectively increasing the number of samples available for model learning. These operations simulate the various factors that training data may encounter in real-world scenarios, such as image rotation, scaling, partial occlusion, and shadowing caused by changes in lighting and shooting angles. The new samples generated through data augmentation differ from the original samples, thereby serving as a form of "data mining" that provides the object detection model with richer learning information. This helps the model learn stable features from the data and improves its generalization ability. For tasks with small sample sizes and imbalanced data distributions, data augmentation can prevent the model from overfitting specific samples and enhance the model's robustness. The image data augmentation methods used in this paper are as follows:

(1) Randomly rotate part of the image between −15° and 15° to simulate small-angle rotations that occur during photography.

(2) Randomly selecting some images for horizontal or vertical flipping to expand the data volume using the symmetry of the images themselves.

(3) Randomly selecting a rectangular area from the image for cropping, then resizing it to the original image size. The size and position of the cropped area are randomly set. This method addresses the issue of incomplete target shapes, enhancing the model's ability to recognize partial inputs.

(4) Randomly scale the image size, setting the scaling ratio range to 0.8–1.2 times the original image size, to simulate image size differences caused by distance or aperture during photography.

(5) Randomly adjust the image brightness and contrast by increasing or decreasing specific values to simulate brightness differences caused by lighting conditions or device parameters, thereby expanding data features under varying lighting conditions.

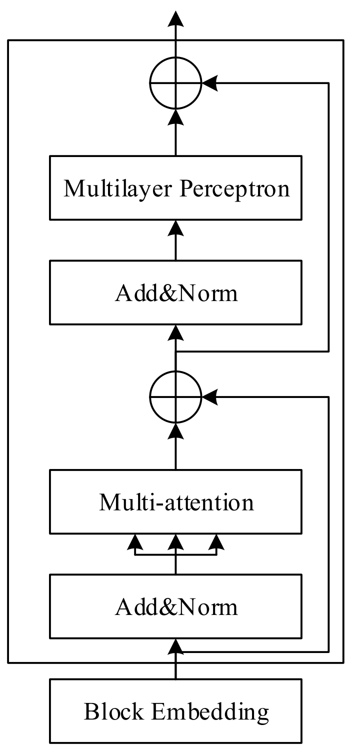

Figure 1: The encoder structure of ViT

III. Types of Artistic Paintings and Emotional Types

III. A. Multifaceted Art Painting Classification Model

III. A. 1) Feature extraction based on Transformer encoders

ViT can be divided into two stages: tokenization and Transformer encoding. In the tokenization stage, the image is divided into L blocks s of size $(h \times h)$, which are then flattened into one-dimensional vectors. Through linear mapping, each block x_i is linearly embedded into a low-dimensional space as shown in Equation (1):

$$E(x_i) = x_i W_e + b_e \tag{1}$$

In equation (1), $E(\cdot)$ denotes the embedding operation, and W_e and b_e are the weight and bias parameters, respectively. Position encoding is added to the tile embedding to fuse spatial information.

Figure 1 shows the encoder structure of the ViT model, where add denotes a residual connection and Norm denotes normalization. Multi-head self-attention (MHA) consists of parallel self-attention heads. In the self-attention mechanism, the key, value, and query all come from the same source. This allows each position within the encoder to consider all positions from the previous layer of the encoder.

The self-attention function maps the query Q and the key-value pair $(K-V)$ to the output, calculated as in equation (2):

$$Attention(Q,K,V) = soft\max\left(QK^T \Big/ \sqrt{d_k}\right)V \tag{2}$$

In equation (2), d_k denotes the dimension of the key matrix. MHA extends self-attention by performing multiple attention operations in parallel. Specifically, Q, K, and V are divided into multiple small vectors, and the self-attention mechanism is applied to each small vector separately. Subsequently, the outputs of MHA are concatenated and a linear transformation is performed to obtain the final output. MHA is calculated as shown in equation (3):

$$MHA(Q,K,V) = Concat(h_1, \ldots, h_h)W^o \tag{3}$$

In equation (3), h is the attention head, W^o is the weight matrix that combines all attention head outputs, and each attention head h_i is calculated as in equation (4):

$$h_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{4}$$

In equation (4), QW_i^Q, KW_i^K, and VW_i^V are the learnable weight matrices for each h_i.

In the CrosVit architecture, the L branch uses larger tiles and more Transformer encoders. The S branch acts as a supplementary branch, operating on finer-grained tile sizes with smaller embedding dimensions. Multiple fusions are performed between the two branches using an interaction attention module, and classification predictions are made using the classification tokens from the L branch. CrossViT improves performance by obtaining fine-grained information from the S branch through its dual-branch architecture, thereby significantly reducing computational costs.

III. A. 2) Feature Selection Optimization

This paper's painting classification model optimizes the fine-grained features extracted from the S branch by improving the CGO algorithm to remove irrelevant features and enhance the model's performance across various artistic painting classification tasks.

CGO is an optimization algorithm based on chaos theory and game theory, leveraging the randomness and ergodicity of chaotic systems combined with game theory principles to explore and develop the solution space, thereby addressing complex optimization problems. While the strong randomness of chaotic systems aids in global search, it may fall short in local search (i.e., precisely identifying the optimal solution within a specific region of the solution space). The NO algorithm is an optimization algorithm inspired by the nutcracker in nature, featuring strong local search capabilities. The proposed optimization algorithm combines CGO with the NO algorithm, achieving both the breadth of global search and excellent performance in local exploration, thereby effectively selecting the optimal feature subset.

Specifically, in feature selection, the improved CGO algorithm first generates a random population as shown in Equation (5):

$$X = L + rand \times (U - L) \tag{5}$$

In equation (5), L and U are the boundaries of the solution space (search space), and $rand$ is a random number.

Next, the Boolean form of each solution X_i is calculated as shown in equation (6):

$$BX_i = \begin{cases} 1, & rand < 0.5 \\ 0, & rand \geq 0.5 \end{cases} \tag{6}$$

Select features corresponding to BX_i from the training set and train the classifier. Calculate the fitness value of X_i based on the classification error γ and the proportion of selected features as shown in Equation (7):

$$Fit_i = \rho \cdot \gamma + (1-\rho) \times \left(\frac{|BX_i|}{D} \right) \tag{7}$$

In equation (7), ρ is the equilibrium parameter, and D is the total number of features.

Determine the optimal solution X_b and attempt to update other solutions using the NO algorithm. Specifically, first, use the CGO algorithm to update X_i. Second, compare the fitness values of the updated solution X_n and X_i and select the solution with the better value. Finally, interact with the global information of the solutions within the search domain, rather than relying solely on the information provided by X_b for optimal solution updates. This increases collaboration between different solutions and improves diversity during the search process. The solution update process is represented by Equation (8):

$$X_i = \begin{cases} X_i - r_1 \times (X_b - X_j), & rand < 0.5 \\ X_i + r_2 \times (X_b - X_j), & rand \geq 0.5 \end{cases} \tag{8}$$

In equation (8), r_1 and r_2 are two random numbers. X_j is a randomly selected solution. Thus, considering the distance between X_b and X_j, a new solution that is beneficial to both X_b and X_j is generated while retaining the current information in X_i. To reduce the time complexity, this strategy is only applied when $Prob_{ex}^i < 0.5$, defined as Equation (9):

$$\Pr ob_{ex}^i = \frac{Fit_i}{\sum_{i=1}^{N} Fit_i} \tag{9}$$

Continue the update process until the termination criteria are met, then return the optimal solution X_b for that stage. Subsequently, use the binary form of X_b to remove irrelevant features from the test set.

III. A. 3) Multi-task cross-integration

The optimized output of the shared feature extraction stage, S, is used as input for the multi-task cross-fusion stage, where multiple fine-grained branches related to art painting classification are introduced. For example, the E fine-grained branch is used for type classification, and the M fine-grained branch is used for author classification, etc. The fine-grained branches share the same tile size as the S branch from the feature extraction stage, and the number of encoders for the fine-grained branches can be independently configured.

The optimized features from the output of the S branch are copied and input into the fine-grained branches. Cross-attention modules are then used for multiple fusions, and finally, cross-additive attention is used for the final feature aggregation. The classification primitives of each branch are fused with the classification primitives of the L branch for task-specific predictions. This design explicitly utilizes information and correlations between multiple tasks without incurring additional independent network costs.

Additive attention, also known as Bahdanau attention, is another mechanism for computing attention scores in sequence-to-sequence models. Attention alignment scores are computed using a single hidden layer feedforward network as shown in Equation (10):

$$f_{Att}(h_i, s_j) = v_a^T \tanh\left(W_a [h_i; s_j]\right) \tag{10}$$

In equation (10), v_a and w_a are learnable attention parameters, s denotes the encoder's hidden state, and s denotes the decoder's hidden state. Additive attention provides flexibility in capturing complex relationships between query vectors and key vectors, making it suitable for tasks involving nonlinear or complex dependencies, where it is used as an alignment score function. In the proposed framework, after calculating the alignment scores, a soft voting strategy is employed to output classification probabilities for each branch (fine-grained branch) via FFN, and the weighted average of all FFN output probabilities is taken as the final classification result.

III. B. Painting Emotion Classification Based on Convolutional Sparse Autoencoders

III. B. 1) Self-learning and knowledge transfer

In recent years, transfer learning and domain adaptation techniques aimed at achieving cross-domain knowledge sharing have become hot topics of research. Unsupervised feature learning can be divided into two modes based on the data distribution during its unsupervised training phase and supervised training phase: semi-supervised learning mode and self-learning mode. Semi-supervised learning requires that the unlabeled data used for feature learning and the labeled data used for supervised learning have the same distribution, while self-learning does not require that the data used for feature learning and the data used for supervised learning have exactly the same distribution. Therefore, it can be regarded as an effective means of conducting transfer learning, significantly expanding the application scope of unsupervised feature learning techniques. To date, transfer learning methods based on autoencoders have been successfully applied in fields such as natural language processing and speech emotion analysis, achieving good results.

Unsupervised feature learning based on autoencoders is shown in Figure 2. As shown in Figure 2(a), unsupervised feature learning based on sparse autoencoders discovers weight parameters for feature extraction through data reconstruction. Assuming that the unlabeled samples used for unsupervised feature learning are $x_s^{(i)} \in R^{m\times 1}$ (where s denotes the source domain), the sparse autoencoder ultimately learns the weight coefficients W and b_1 corresponding to the m-dimensional input sample data, taking into account the bias vector. As shown in Figure 2(b), assuming that the samples used for supervised training are $x_t^{(i)} \in R^{m\times 1}$ (where t denotes the target domain), feature extraction for labeled samples involves using the weight coefficients W and b_1 to probe the samples to obtain feature responses as in Equation (11):

$$a_t^{(i)} = \sigma\left(Wx_t^{(i)} + b_1\right) \tag{11}$$

In the formula:

$a_t^{(i)}$ ——feature response of the sample;

$\sigma(\cdot)$ ——activation function that maps the response value to the range $[0,1]$.

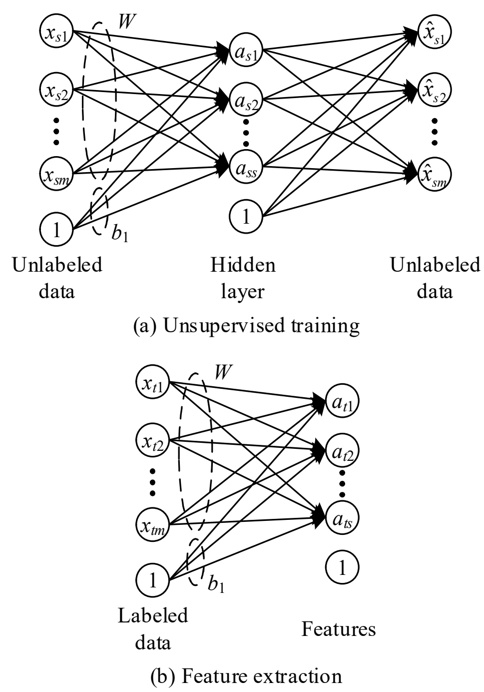

Figure 2: Unsupervised feature learning based on autoencoders

After obtaining the feature responses of labeled samples, these responses can be fed into supervised training models such as SVM, LR, and Softmax for training. Unsupervised training does not require labeled data, so the network parameters in the unsupervised feature learning stage are unrelated to the labeled values of the training data. The weights obtained from training are merely the key parameters for data self-reconstruction. If the data distributions of $x_t^{(i)}$ and $x_s^{(i)}$ are different, unsupervised feature learning adopts a self-learning mode. In this case, unsupervised feature learning is equivalent to transfer learning and domain adaptation, or learning knowledge from a source domain and applying it to other target domains, where the data in the source and target domains can have different distributions.

Since modern and contemporary Lingnan landscape paintings and Western landscape paintings typically lack concrete meanings, but share certain similarities in color and texture application, establishing a mapping relationship between painting image features and emotional semantics is feasible. Existing research on emotional semantic analysis of painting images primarily extracts low-level visual features such as color and texture. This paper attempts to apply unsupervised feature learning techniques to emotional semantic analysis of painting images, conducting feature learning based on sparse autoencoders and performing painting image classification at the emotional level. However, while there are many sample databases in the field of painting image emotional semantic analysis, the content within these databases is limited, making it difficult to systematically provide the large amount of unlabeled samples required for unsupervised feature learning, which poses challenges for image classification based on unsupervised feature learning. Therefore, this chapter employs the transfer learning method described in this section when applying unsupervised feature learning to abstract painting emotional classification, learning features from unlabeled data outside the fields of modern and contemporary Lingnan landscape painting and Western painting.

III. B. 2) Emotional classification system for painted images

This chapter uses a convolutional autoencoder model for abstract painting image emotion classification, but unlike previous image classification methods, this chapter utilizes unsupervised feature learning based on transfer learning concepts using unlabeled samples from outside the target domain. The overall framework of the abstract painting image emotion classification system used in this chapter is shown in Figure 3 and can be divided into three main components:

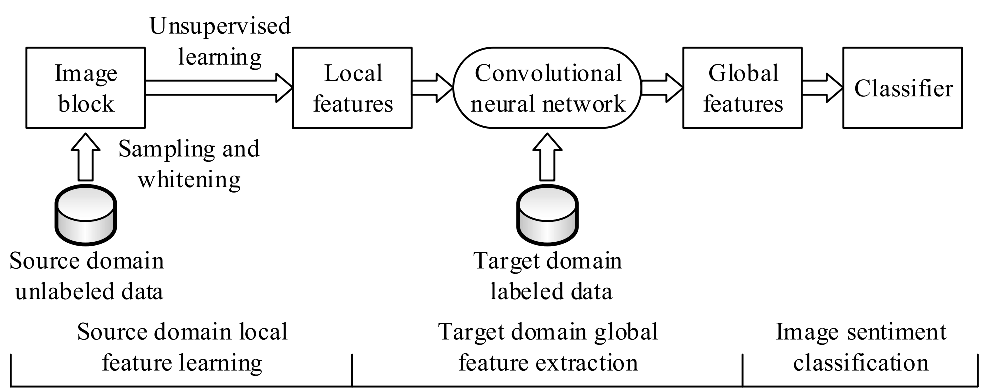

Figure 3: Abstract painting image emotion classification system

(1) Local feature learning in the source domain

In this paper, we utilize the large unlabeled image database STL-10 for abstract image sentiment classification. We collect image sub-blocks from the database and select appropriate regularization constants for whitening processing. Then, we perform unsupervised training based on a single-layer autoencoder with sparse constraints to obtain local feature weights corresponding to the image sub-blocks.

(2) Global feature extraction in the target domain

After obtaining the local feature weight coefficients, a convolutional neural network is used to extract the global feature responses of abstract paintings in the target domain using parallel two-dimensional convolution. Under the premise of whitening, the feature responses are aggregated using average pooling and converted into vector form.

(3) Image Emotion Classification

Since this paper performs binary classification for emotion classification of abstract paintings, the LR model is used as the classifier during the supervised training phase. The global feature responses extracted from the abstract painting images are fed into the LR classifier, and supervised training and testing are conducted based on sample label information. The performance of abstract painting emotion classification based on sparse autoencoders and transfer learning is evaluated by comparing it with traditional methods.

Visualizing the feature weights learned by the sparse autoencoder allows for an intuitive assessment of learning effectiveness from the perspective of “edge-based” analysis. However, during actual training, network weight coefficients are typically initialized using a random method. Unsupervised feature learning experiments conducted under different conditions yield feature weights without discernible patterns, making direct comparisons challenging. To conduct a thorough performance comparison, this paper collects different numbers of training samples under cross-domain and non-cross-domain conditions for experimentation. To enable an intuitive comparison of the feature weights obtained from multiple experiments, a method is proposed to rank the feature weights obtained from unsupervised feature learning based on the strength of marginality.

Assume that the input data dimension of the sparse autoencoder is $m = n \times n \times 3$, and the number of units in the hidden layer is s. Then, the size of the weight matrix W is $s \times m$. The s row vectors of W are all data with a dimension of $m = n \times n \times 3$, representing the relationship between the s hidden layer responses and the input data as shown in equation (12):

$$W = \begin{bmatrix} w_1 \\ \vdots \\ w_j \\ \vdots \\ w_s \end{bmatrix} = \begin{bmatrix} w_{11} & \cdots & w_{1k} & \cdots & w_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{j1} & \cdots & w_{jk} & \cdots & w_{jm} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{s1} & \cdots & w_{sk} & \cdots & w_{sm} \end{bmatrix}_{s \times m} \tag{12}$$

In the equation, w_{jk} — connection coefficients between each hidden layer feature and each element of the input data, w_j — coefficient vectors corresponding to each hidden layer feature.

As a parameter for evaluating the relative clarity of an image, the average gradient reflects the grayscale changes in the image in multiple directions. Assuming that a grayscale image I has a size of $e \times f$, its average gradient is defined by equation (13):

$$AG(I) = \frac{1}{(e-1) \times (f-1)} \sum_{p=1}^{e-1} \sum_{q=1}^{f-1} \sqrt{\frac{(\Delta_x I(p,q))^2 + (\Delta_y I(p,q))^2}{2}} \tag{13}$$

In the formula, $\Delta_x I$ — image gradient in the horizontal direction, $\Delta_y I$ — image gradient in the vertical direction, (p,q) — image pixel coordinates.

If the weight coefficients corresponding to each feature in the weight matrix (i.e., the $n \times n \times 3$ elements in each row of W) are treated as image data, the strength of their edges can be evaluated using the average gradient parameters, and they can then be sorted and displayed accordingly. If the weight w_j corresponding to the j th feature corresponds to the weight $\mathbf{w}_j^c$ of the color component $c \in \{R, G, B\}$, it can be converted into an $n \times n$ coefficient matrix and the gradient data in the horizontal and vertical directions can be calculated as in equations (14)-(15):

$$\Delta_x w_j^c(p,q) = w_j^c(p,q) - w_j^c(p+1,q) \tag{14}$$

$$\Delta_y w_j^c(p,q) = w_j^c(p,q) - w_j^c(p,q+1) \tag{15}$$

In the formula, (p,q) ——The element coordinates in the weight component w_j^c converted to matrix form.

Then, the average gradient of each feature weight matrix in each color component can be defined as in formula (16):

$$AG(w_j^c) = \frac{1}{(n-1)^2} \sum_{p=1}^{n-1} \sum_{q=1}^{n-1} \sqrt{\frac{(\Delta_x w_j^c(p,q))^2 + (\Delta_y w_j^c(p,q))^2}{2}} \tag{16}$$

Take the minimum value of the average gradient of the three color component data $c \in \{R, G, B\}$ as the indicator to rank the unsupervised feature learning weights as shown in Equation (17):

$$mAG(w_j) = \min_{c \in \{R,G,B\}} AG(w_j^c) \tag{17}$$

Additionally, to enable a thorough comparison, this paper also references relevant research on modern and contemporary Lingnan landscape paintings and Western landscape paintings to construct a supervised deep CNN for painting emotion classification. Figure 4 shows the network structure of the painting emotion classification system based on the supervised CNN. Due to space constraints, the pooling layers are omitted here. First, all image samples are resized to 256×256 without considering the aspect ratio, and sub-blocks are extracted from the central part of the images through cropping. This process normalizes all samples to a size of 227×227. The standardized image samples are then fed into two convolutional layers. The first convolutional layer contains 96 convolutional kernels, with a local receptive field size of $11 \times 11 \times 3$ and a stride of 4. The second convolutional layer contains 256 convolutional kernels, with a local receptive field size of $5 \times 5 \times 96$ and a stride of 2. A pooling layer is appended after each convolutional layer (due to space constraints, the pooling process is not shown in Figure 4). The network is followed by a fully connected layer with dimensions $512 \times 512 \times 24$. Since all experiments conducted in this chapter for abstract painting emotion classification are binary classification experiments, a logistic regression model is used as the final layer of the deep network to establish a connection between feature data and emotion classification results.

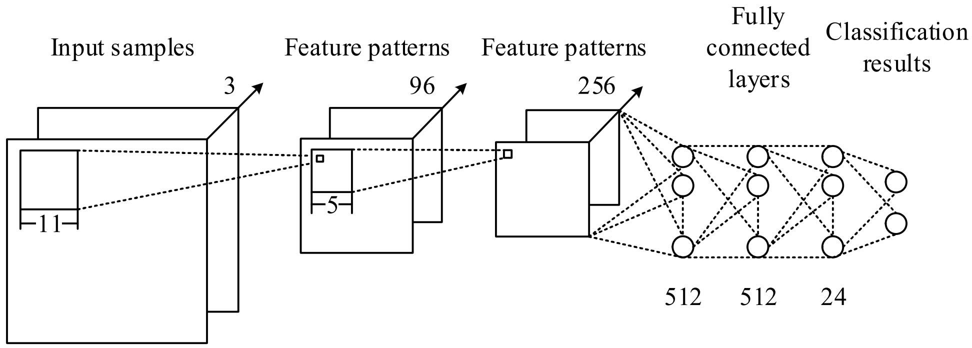

Figure 4: A painting sentiment classification network based on supervised CNN

IV. Training and performance evaluation of the painting emotion classification model

IV. A. The Impact of Oversampling Training Methods on Model Performance

This section compares the effects of various data augmentation methods—including (B2) cropping + flipping, and adjusting the (B3) brightness, (B4) hue, (B5) saturation, and (B6) contrast of images—on the classification performance of models in the task of classifying the emotions in paintings. The experimental results of the five oversampling methods are shown in Table 1, with the model results without oversampling training serving as the baseline (B1). It can be seen that the (B2) cropping + flipping, (B3) brightness, and (B5) saturation methods can effectively improve the classification performance of the convolutional sparse autoencoder-based model, with improvement effects all exceeding the (B1) baseline (≥0.725) and standard deviations remaining within an appropriate range (≤0.030). However, the (B4) hue (0.718) and (B6) contrast (0.683) methods have a negative impact on model performance, as altering hue changes the image's color, and changes in contrast affect the model's attention distribution.

Table 1: Performance of several oversampling methods on the painting image dataset

Oversampling mode	The painting image dataset
(B1) Baseline	0.725+0.035
(B2) Cutting + Flipping	0.732+0.016
(B3) Brightness	0.737+0.022
(B4) Hue	0.718+0.034
(B5) Saturation	0.734+0.008
(B6) Contrast	0.683+0.039

To verify whether the proposed painting emotion classification model also follows the oversampling fine-tuning principle obtained in the previous section for the task of classifying the emotions of painting images other than modern and contemporary Lingnan landscape paintings and Western landscape paintings, this paper selects the public dataset P for experimentation. The P dataset consists of various types of painting images, totaling 5,896 images. Using the (B2) cropping + flipping oversampling method as the baseline (B1), we randomly increased or decreased the hue (B3), brightness (B4), saturation (B5), and contrast (B6) of the source images by 20–60%. The experimental results are shown in Table 2. Unlike the oversampling methods used for modern and contemporary Lingnan landscape paintings and Western landscape paintings, the data augmentation methods for other types of painting images improved model performance (≥0.890) except for hue (B4) (0.873). This indicates that the proposed model exhibits similarities in emotional classification tasks across various types of painting images, but the same training strategy cannot be directly applied across all cases.

Table 2: Performance of several oversampling methods on the P image dataset

Oversampling mode	The P image dataset
(B1) Baseline	0.890+0.025
(B3) Brightness	0.899+0.008
(B4) Hue	0.873+0.022
(B5) Saturation	0.892+0.019
(B6) Contrast	0.895+0.021

IV. B. Model Performance Evaluation

To validate the effectiveness of this data augmentation method, the P dataset was split into training and testing sets using 5-fold cross-validation. Two commonly used baseline models were selected for comparison: (M1) the PCNN baseline model and (M2) a neural network-based painting emotion classification model. The experimental results of the two models and the (M3) model proposed in this paper are shown in Table 3, where the suffix "-os" denotes oversampling. Regardless of whether oversampling training strategies were employed, the performance of the (M3) proposed model on both the training and test sets significantly outperformed the comparison models, with overall performance consistently above 0.730. After applying oversampling training strategies, the classification performance of the (M3) proposed model reached 0.750 or higher, with a standard deviation <0.010, demonstrating strong stability. After oversampling training, the (M2) neural network-based painting emotion classification model also outperforms the (M1) PCNN baseline model (>0.710). Experiments indicate that certain oversampling methods are effective in painting emotion classification models, but the oversampling strategy must be adjusted specifically for different types of painting image classification tasks.

Table 3: Average classification performance and standard deviation on the P dataset

Model	Training set	Test set
M1	0.702	0.726
M2	0.698+0.028	0.704+0.026
M2-os	0.714+0.017	0.731+0.013
M3	0.735+0.012	0.745+0.011
M3-os	0.754+0.008	0.759+0.009

V. Similarities and Differences in the Aesthetic Realm of Painting in Different Cultural Contexts

V. A. Overview of Cultural Background

This section briefly outlines the historical and cultural background of modern and contemporary art in the Lingnan region and the West as follows:

(1) Modern and contemporary landscape paintings in Lingnan. Generally speaking, the modern and contemporary history of Lingnan refers to the period from the start of the Opium War in 1840 to the liberation of the Lingnan region in 1949 and the establishment of local governments under central authority. Over the course of a century, the Lingnan region underwent multiple major transformations in social structure, with the lives of its people marked by instability. This was accompanied by intensified ethnic conflicts, sharpened class contradictions, and continuous struggles and movements. Additionally, as a primary window for interaction between several dynasties and Western culture, the Lingnan region also absorbed a significant influx of Western cultural influences during the modern and contemporary periods, gradually integrating them into its own traditions.

(2) Western landscape painting. During the ancient Greek period, scholars such as Plato and Aristotle laid the foundation for Western art to use the objective world as a reference point. During the Middle Ages, the prevalence of Christianity and the concrete narrative requirements of religious propagation further reinforced the realistic techniques of Western art. By the 16th-century scientific revolution, scientists and the general public's worship of objectivity and rationality further emphasized the realistic tendencies of Western painting.

V. B. Similarities and Differences in Painting Imagery Based on Cultural Backgrounds

After using a multi-faceted artistic painting classification model to preliminarily categorize the collected painting image data, we employed the proposed painting emotion classification model based on a convolutional sparse autoencoder to statistically analyze and organize the emotional expressions in modern and contemporary Lingnan landscape paintings and Western landscape paintings. By considering different cultural contexts, we examined the similarities and differences in emotional expression and the creation of artistic ambiance between the two types of paintings.

V. B. 1) Similarities: Scenic representation, blending emotion into scenery

Among the 1,589 modern and contemporary Lingnan landscape paintings and 1,541 Western landscape-style paintings, the most frequently appearing landscape motifs are listed in Table 4, including trees, streams, mountains and rivers, land, sky, architecture, and birds and flowers. The figures in parentheses indicate their respective proportions of the total. Overall, both modern and contemporary Lingnan landscape paintings and Western landscape-type paintings prefer to use two landscape motifs—trees and streams—which account for 40.00% or more of the total. Both types of paintings use natural landscapes as their medium, supplemented by different spatial treatments and color application methods, to express the inner emotions of contemporary artists.

Table 4: Statistics of landscape imagery in portraits

Landscape type	Modern and contemporary Lingnan landscape paintings	Western painting
Trees	1021 (64.25%)	1277 (82.88%)
River	689 (43.36%)	1012 (65.70%)
Mountains	1341 (84.39%)	70 (4.53%)
Ground	247 (15.54%)	888 (57.61%)
Sky	162 (10.20%)	926 (60.19%)
Architecture	59 (3.71%)	683 (44.34%)
Birds and flowers	389 (24.48%)	1087 (70.55%)

V. B. 2) Unique feature: Appreciation or reproduction of artistic conception

A statistical analysis of modern and contemporary Lingnan landscape paintings and Western landscape paintings reveals the spatial treatment and color usage employed in these works, as shown in Table 5. Spatial treatment methods are divided into three categories: (S1) scattered perspective, (S2) focal perspective, and (S3) abstract deformation. Color usage methods are divided into two categories: (C1) narrative-serving and (C2) narrative-independent.

Table 5: The situation of spatial treatment and color application

		Modern and contemporary Lingnan landscape paintings	Western painting
Spatial processing	S1	1336 (84.08%)	70 (4.53%)
	S2	68 (4.28%)	564 (36.57%)
	S3	185 (11.64%)	908 (58.90%)
Use of color	C1	1458 (91.76%)	229 (14.89%)
	C2	131 (8.24%)	1312 (85.11%)

In terms of spatial treatment, modern and contemporary Lingnan landscape paintings remain consistent with traditional Chinese ink paintings, primarily employing (S1) scattered perspective (84.08%). At the same time, influenced by Western foreign cultures, many painters have begun to experiment with (S3) abstract deformation methods, accounting for 11.64%. In terms of color usage, they remain rooted in traditional Chinese cultural heritage, primarily using ink wash with subtle colors as an accent. 91.76% of the paintings employ (C1) narrative-serving color usage, where color serves the narrative of the painting. Artists use negative space techniques to guide viewers in perceiving and experiencing the emotional ambiance hidden within the brushstrokes and ink marks.

Western landscape paintings adhere to the principle of "depicting form through reason" in both spatial treatment and color application, primarily employing (S2) focal perspective (36.57%), (S3) abstract deformation (58.90%), and (C2) independent narrative (85.11%). Through spatial techniques like focal perspective and abstract deformation, Western landscape paintings aim to perfectly align with human visual perception of landscapes, establishing a human-centered system for observing objects. Additionally, the use of color in Western landscape paintings is independent of the narrative of the artwork. Whether it be the inherent color of classicalism or the decomposition of light and color in impressionism, the goal is to scientifically and vividly recreate natural scenes, directly showcasing the beauty of nature to the audience and expressing corresponding emotions.

VI. Conclusion

This paper proposes a multi-faceted artistic painting classification model as a tool for classifying painting image types. It designs a painting emotional type model based on a convolutional sparse autoencoder to explore and organize the emotional connotations of painting images. Among these, the painting emotional classification model demonstrates improved accuracy in oversampling training strategy experiments, with three expansion methods—cropping + flipping, brightness, and saturation—all achieving accuracy higher than the baseline (0.725). Compared to commonly used models, the proposed model not only demonstrates superior overall performance (>0.730) but also exhibits strong adaptability to oversampling training strategies, achieving performance improvements of 0.750 or higher with a standard deviation <0.010.

Combining the historical and cultural backgrounds of modern and contemporary Lingnan and the West, this paper analyzes the similarities and differences in the artistic conception of modern and contemporary Lingnan landscape paintings and Western landscape paintings under the support of the proposed art painting classification model and painting emotion classification model. Both adopt scene representation and the integration of emotion into the scene. The difference lies in the fact that modern and contemporary Lingnan landscape paintings use traditional Chinese ink painting philosophy as their overall framework, employing scattered perspective (84.08%) and color to serve narrative (91.76%) methods to create traditional aesthetic beauty.

Funding

This research was supported by the Guangdong Provincial Philosophy and Social Sciences Planning 2024 Annual Lingnan Culture Project: "A Study on the Mutual Relationship between Modern Lingnan Landscape Painting and Overseas Landscape Painting" (Approval Number: GD24LN17).

References

[1] Wang, H. (2022). Conceptual expression in modern artistic creation—The relationship and law between artistic conception and thinking. Frontiers in Art Research, 4(13), 98-103.

[2] Bi, W., & He, Z. (2019, December). Research on the Influence of Artistic Conception of Chinese Painting on Creative Image Design. In 2019 International Conference on Economic Management and Cultural Industry (ICEMCI 2019) (pp. 445-447). Atlantis Press.

[3] Baihua, Z., & De Meyer, J. (2017). The birth of artistic conception in China. Art in Translation, 9(3), 367-396.

[4] Xing, L., Razak, H. A., & Noh, L. M. B. M. (2023). The aesthetic development of traditional Chinese landscape painting in contemporary landscape painting. Art and Performance Letters, 4(7), 53-63.

[5] Diep, V. T. (2017). The landscape of the Void: truth and magic in Chinese landscape painting. Journal of Visual Art Practice, 16(1), 77-86.

[6] Xinhong, W., Dingju, Z., & Yingxi, M. (2023). Logical coupling and paradigm shift of regional cultures in terms of “the belt and road initiative”——a case study of lingnan culture. International Journal of Engineering Business Management, 15, 18479790231213556.

[7] Liu, W. (2021). Analysis on the Collision and Fusion of Eastern and Western Paintings in the Context of Globalization. Thought, 7(8).

[8] Liu, Y. (2019). Art for the New Nation: The Rise of the Lingnan School in the 1920s and the 1930s. HKU Theses Online (HKUTO).

[9] Ho, R., Vollmer, S. H., & Zheng, X. (2024, July). Subverting Syntax: Experimental narratives from the Post-Lingnan School of Painting. In Proceedings of EVA London 2024 (pp. 85-93). BCS Learning & Development.

[10] An, Q. (2024). Implementation of intelligent painting systems in art education as a way of developing student self-efficacy and involvement: Post Lingnan Painting Spirit. Education and Information Technologies, 29(12), 15173-15189.

[11] Yang, T., Silveira, S., Formuli, A., Paolini, M., Pöppel, E., Sander, T., & Bao, Y. (2019). Aesthetic experiences across cultures: Neural correlates when viewing traditional Eastern or Western landscape paintings. Frontiers in psychology, 10, 798.

[12] Chen, C. (2023). Research on the influence of western renaissance thought on modern painting Art. Journal of Education, Humanities and Social Sciences, 15, 202-206.

[13] Duan, X. (2023). A study of Chinese and Western landscape painting from the comparative perspective. Frontiers in Art Research, 5(13), 1-7.

[14] Shen, B. (2025). The Artistic Conception Description of Landscape Painting Sketching. Frontiers in Art Research, 7(1).

[15] Wang, J. (2025). The artistic conception expressed through brush and ink in Chinese painting. Cambridge Arts and Humanities Research, 2025(3), 27-31.

[16] Xuan, Q. (2025). A Study of Imparities between Chinese Painting and Western Painting from Perspectives of Arts and Cultures. Journal of Literature and Arts Research, 2(1), 141-146.

[17] Zhou, H., & Ma, Q. (2025). The Influence of Traditional Chinese Philosophy on the Artistic Characteristics of Landscape Painting Creation. Cultura: International Journal of Philosophy of Culture and Axiology, 22(3).

[18] Zhu, Z. (2024). A Comparison of Perspective in Chinese and Western Traditional Painting and the Philosophical Thought behind It. Art and Design Review, 12(4), 219-237.

[19] Chen, L., & Jing, M. Q. (2021). Contrast the difference of thinking between Chinese and Western painting. Int. J. Front. Sociol, 3(21), 4-7.

[20] Jiao, H. (2025). A Multi-Dimensional Interpretation of the View of Nature and Humanistic Spirit in Chinese Landscape Paintings from the Perspective of Art Criticism. Cultura: International Journal of Philosophy of Culture and Axiology, 22(2), 524-544.