ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

A Comparative Study of Modern Lingnan Landscape Paintings and Western Landscape Paintings under Cross-cultural Perspective

Bing Wu[1,*]

[1] School of Fine Arts, Zhaoqing University, Zhaoqing, Guangdong, 526061, China

SUMMARY: *In this paper, we construct a dataset of modern Lingnan landscape paintings and Western landscape paintings in cross-cultural perspective from multiple aspects, and ensure the dataset usability through preprocessing operations. Aiming at the problem that the traditional CycleGAN model cannot retain more contextual detail information, three loss functions are introduced on the basis of the loss functions of the generative network and the discriminative network, and a detailed training process is also given, aiming at retaining the contextual features of the artworks. Subsequently, it is inputted into the convolutional neural network based on hybrid attention mechanism to extract the contextual features of each frame of modern Lingnan landscape painting and western landscape painting images, and then it is put into the bidirectional GRU model to learn the contextual information, and finally realizes the contextual classification of the images of modern Lingnan landscape painting and western landscape painting, and thus designs a contextual classification model based on CNN-BiGRU-At. Compared with the CycleGAN model, the improved CycleGAN model has a higher priority in image digitization conversion, and its values are classified as 0.759~0.941. In addition, the accuracy distribution of the context classification of digital images of modern Lingnan landscape paintings of the CNN-BiGRU-At model ranges from 0.5 to 0.73, and the accuracy distribution of digital images of western landscape paintings of the context classification is 0.7~0.9, which is 0.5~0.73. 0.7~0.9, clearly perceived modern Lingnan landscape paintings and western landscape paintings to create differences in mood, western landscape paintings in general to simple, abstract mood is mainly gradual to attract people to think deeply, Lingnan landscape paintings are complexity and richness, visual elements have rocks, fishing village, smoke and rain, boat trip, etc., the overall mood for the rich and concrete mainly.*

KEYWORDS: *improved CycleGAN model; loss function; CNN-BiGRU-At; Lingnan landscape painting; Western landscape painting; mood classification model*

1 Introduction

Artistic conception, as an important concept in Chinese aesthetics, runs through almost all artistic creations. The idea of "artistic conception" first emerged in the literary theories of the Wei, Jin, Northern and Southern Dynasties, with the theories of "image" and "realm" [1]. Since then, “mood” has always been the profound code word of Chinese art. The mood of Chinese painting is not only a blend of feelings and scenery, but also a fusion of the artist's thoughts, aesthetic concepts and aesthetic ideals, interests and objective scenery [2, 3]. There is a very obvious difference between the aesthetic pursuits of Eastern and Western paintings. Oriental paintings mainly based on Chinese traditional paintings emphasize the expression of emotions,

*wuhuashi2000@163.com
https://doi.org/10.65102/is2026151
Published: 30 April, 2026

while Western traditional paintings mainly based on oil paintings emphasize the imitation of nature [4]. Oriental paintings mainly based on Chinese traditional paintings pay attention to the beauty of the mood, while western traditional paintings mainly based on oil paintings pay more attention to the beauty of the human body [5]. Oriental traditional painting presents a philosophical aesthetic tendency, while Western traditional painting presents a scientific aesthetic tendency. Compared with the white aesthetics promoted by China, Western art prefers light and shadow, perspective and spatial hierarchy as the means of expression when constructing a mood. Antrop, M analyzes the uniqueness of Eastern and Western art in the portrayal of landscapes, and points out that Western landscapes mostly use linear perspective and changes in light and shadow to create a sense of depth, which results in realistic spatial layouts [6]. For example, Impressionist painters used light and shadow changes and color superimposition to build a sense of instantaneous atmosphere, and Renaissance artists used the principle of precise perspective, objects in the picture in accordance with the spatial layout of the orderly arrangement, presenting a clear effect of hierarchy [7].

Relying on the visual logic of shaping the mood, it is far from the traditional painting advocated the concept of “the idea is first in the brush” [8]. However, in the contemporary art world, it is increasingly common for Western artists to borrow the compositional style of Chinese ink painting. For example, in the painting of landscape oil paintings, artists tend to reduce the details of the detailed carving, the use of symbolic brush strokes to depict the natural scrolls, the works presented by the atmosphere more and more seem more remote, in the stage of cross-cultural art exchanges, the East and the West in the creation of art in the mood of the interactive effect is gradually appeared [9, 10]. Bao, Y et al. argue that Western and Chinese artists follow different traditions when presenting themselves in the form of paintings; Chinese artists, on the other hand, pay more attention to the contextual information in their paintings, whereas Western artists are more likely to adopt a central perspective in depicting the world and focusing on the salient elements in the picture [11]. Xuan, Q argues that the differences between Chinese and Western paintings mainly stem from different historical evolutions and cultural concepts; Chinese paintings embody the harmony and unity of man and nature, often employing symbolic representations and leaving space for subjective imagination, while Western paintings place more emphasis on realism, with detailed depictions of nature and the human world based on rational observation [12]. Wang, Y et al. considered composition, color, subject matter, and technique to construct a comprehensive aesthetic evaluation system for revealing differences between Eastern and Western paintings, and then used correlation analysis and structural equation modeling to analyze the functional relationships between evaluative factors, the power of perspective, and the overall aesthetic appeal of the paintings [13].

The deeper meaning of the mood can be interpreted and analyzed from multiple dimensions, at first this performance presents the integration of the artist's subjective emotion and the objective scene, and the mood embodied in the oil painting is not only the results of the visual presentation, but also the reflection of the emotion and the expression of ideas [14-16]. Morphy, H made a cross-cultural analysis of the phenomenon of art fusion, in the face of the wave of globalization, he pointed out that artists are tending to cross the boundaries of traditional art, and synthesize a variety of cultural means of expression, taking many modern artists as the object of study, in the field of painting, they drew on the color expression of Western oil paintings, and also incorporated the mood of the Oriental ink paintings to convey the works of the works of the East and the West, which presents the mutual penetration of the aesthetics of the scene. This kind of artistic fusion, originating from cross-cultural backgrounds, broadens the diversity of mood creation and extends a wider exploration space for modern art creation [17]. Yang, T et al. neurologically compared traditional Chinese and Western landscape paintings and found that participants showed stronger brain activation to artistic expressions

from their own culture [18]. Huo, C and Choi, D believe that the emotional expression and artistic conception creation of a work are closely related to the spatial layout. Relying on a meticulous layout, the coordination and consistency of each element in the picture in space are achieved, enhancing the visual harmony of the work. By implementing reasonable spatial planning, the artists have successfully shaped diverse emotional atmospheres, such as using open Spaces. It can reveal the emotional traits of loneliness, silence and infinity, based on a compact layout, and then present a warm and intimate emotional experience [19].

Color is not only the visual basis for the creation of painting art, but also the core way of conveying emotions and creating a mood in painting creation. With the help of the power of color, artists can convey specific emotional connotations, highlight the theme of the picture, and echo the emotions of the viewer [20-22]. Pylypchuk, O believes that the different hues can stimulate a variety of emotional experiences, warm tones tend to stimulate a feeling of warmth, enthusiasm and vitality, while cool tones tend to shape a quiet and peaceful atmosphere, the use of color and contrasting artistic expression, further expanding the emotional dimension of the work and the creation of mood [23]. According to Zhang, H, Chinese landscape painting emphasizes the expression of emotion and the feelings of awe and emotion towards nature, and its core concept is “learning from nature”; while Western landscape painting emphasizes the realism of the scenery, which is the re-creation of nature, and it can make people feel immersed in reality when they watch the paintings [24]. Duan, X compares the artistic styles, expressive methods and image presentations of Chinese and Western paintings. Chinese landscape painting emphasizes the expression of the painter's personal emotions and artistic conception, while Western landscape painting pursues a more realistic reproduction of the depicted objects [25].

Existing resources of modern Lingnan landscape paintings and Western landscape paintings are selected to compose a dataset of modern Lingnan landscape paintings and Western landscape paintings in cross-cultural perspective, and the dataset is subjected to preprocessing operations to ensure the rigor of the research results. Aiming at the CycleGAN model's inability to retain more detailed information about the mood, it is proposed to optimize three aspects, namely, generative network, discriminative network, and loss function, so as to make the digitized images of Lingnan landscape paintings and Western landscapes conform to the requirements of the study. Convolutional neural network based on hybrid attention mechanism is utilized to extract the contextual features of each frame of modern Lingnan landscape paintings and western landscapes images, followed by the use of bi-directional GRU model, *Softmax* function to classify the contextual features. This constitutes a CNN-BiGRU-At based mood classification model. Finally, the improved CycleGAN model and the context classification model are used to carry out a comparative analysis of the context creation of modern Lingnan landscape paintings and Western landscape paintings.

2 Modern Lingnan Landscape Painting and Western Landscape Painting Context Exploration

2.1 Data set construction and preprocessing

2.1.1 Data set construction

The construction of the dataset is the basis for the successful implementation of the research work, especially when it comes to contextual feature extraction, the quality and diversity of the training data directly determines the effectiveness of the final model. Since the improved CycleGAN uses unsupervised learning, the dataset needs to be rich enough and consistent with

the target features to ensure that the model can accurately learn the contextual features of modern Lingnan landscape paintings and Western landscape paintings.

The construction of the dataset of modern Lingnan landscape paintings and Western landscape paintings in cross-cultural perspective clarifies the source of the dataset, and 666 classic modern Lingnan landscape paintings and Western landscape paintings are selected, and ensures that the selected works cover different characteristics of brushwork, composition, and ink rhythms. High-quality data of 758 modern Lingnan landscape paintings and Western landscapes were acquired through the museum's high-definition digitization resources. The next step was to screen the data. (1) Stylistic consistency. Try to choose the typical styles of ink, light-red and green landscape to avoid the mixing of painting styles from different periods. (2) High-definition resolution. It is important to ensure that the details of the training images are clear, to avoid that the model mistakenly learns the noise due to low resolution. (3) In diversity. To cover different compositional forms of lofty, deep, flat, brushwork phi-ma-chou, axe-chou, subject landscapes, figures, to improve the generalization ability of the model. The image dataset is constructed by integrating high-quality resources on the Internet, and now contains 3611 selected images. On the one hand, we use Google, Flickr, Pixabay, etc., to obtain basic materials through keyword search and intelligent classification function, and on the other hand, we select works with high artistic value from the portfolio of professional photographers and photography community.

2.1.2 Pre-processing

Preprocessing and enhancement are essential in improving the training efficiency and training effect of the model. The preprocessing of the images of modern Lingnan landscape paintings and western landscapes should firstly denoise, using median filtering, convolutional smoothing, etc. to eliminate the noise caused by scanning and shooting, and then unify the sizes, changing the sizes of the images of modern Lingnan landscape paintings and western landscapes to 512×512, so as to make the input images unified during training, thus speeding up the speed of model training, and finally perform the standard, normalized to [-1 ,1] in order to be consistent with the activation function used by the improved CycleGAN, which facilitates the training. Color information is also important for style transfer, especially the ink color change as well as light ink, thick ink and so on in modern Lingnan landscape painting and western landscape painting. When data preprocessing, the image color space may have to be converted to Lab or HSV space, so that the model can better learn the color relationship and gradient, improve the model robustness and prevent overfitting.

2.2 Improved CycleGAN modeling

Improve the stability of CycleGAN model training to increase the stability of modern Lingnan landscape painting and Western landscape painting by converting the mood characteristics of modern Lingnan landscape paintings and Western landscape paintings from traditional painting forms to digital images, which not only maintains the color characteristics, but also better preserves the details of the brushstrokes, hierarchical relationships and spatial composition, and enhances the visual authenticity, aiming to gain an in-depth understanding of modern Lingnan landscape paintings and Western landscapes with respect to these aspects of the composition, color, and mood The aim is to gain a deeper understanding of modern Lingnan landscape painting and Western landscape painting in terms of composition, color and mood.

2.2.1 Generating networks

The source of generator network construction ideas in this paper is precisely based on the jump layer structure mentioned in U-Net and the residual block principle mentioned above, which is a further improvement of the CycleGAN base model. However, such a disadvantage is due to its narrow bottleneck layer, which prevents the output image from retaining more contextual detail information. In addition, this approach is still more limited to low-resolution images due to the low network capacity.The CycleGAN architecture learns the contextual features of modern Lingnan landscape paintings versus Western landscapes through the use of residual blocks, which have been shown to work in very deep networks where they can represent low-frequency information well compared to the network capacity of DiscoGAN. However, residual blocks are used at a scale such that this results in much less information being available to pass through bottlenecks, limiting the features that can be learned by the network. The generator includes residual blocks in multiple layers of the decoder and encoder, allowing the network to learn multi-scale transformations over higher and lower spatial resolution functions.

2.2.2 Discriminative networks

The discriminator network in the CycleGAN algorithm model uses a Markovian discriminator (PatchGAN), which can only manipulate the corresponding piece individually for fast convergence, but limits the network's awareness of global spatial information, making the generator less able to cope with coherent, global changes in the graph. Because for the discriminator, it is often a higher resolution segmentation graph, which must cause the generator and discriminator to require a larger flow of information. Dilated convolution, on the other hand, is highly effective in segmentation network applications, often reaching high levels with few parameters relative to traditional convolutional networks. Compared to patch block based discriminators, dilated convolution is able to have a larger range of regions for data prediction at the same cost of parameters. Dilated convolution allows the discriminator to implicitly learn the context, which allows for a greater sense of wildness in image processing of modern Lingnan landscape paintings versus Western landscape paintings. On the task of learning contextual features of modern Lingnan landscape painting and Western landscape painting images, the multi-scale discriminator can present better results and stability.

2.2.3 Loss function

The loss function of the improved model changes from the original adversarial and cyclic consistent loss to three parts: feature matching loss, adversarial loss, and cyclic consistent loss with MS-SSIM.

(1) Feature matching loss

Because a multi-scale structure is used in the improved network model, a feature matching loss is used in the objective function in order to improve the stability of the model.

$$L_{FM}(G,D)=\frac{1}{n-1}\sum_{i=1}^{n-1}\left\|E_{x\sim pdata}f_i(x)-E_{z\sim p_z}f_i(G(z))\right\|_2^2$$
$$L_{FM}=L_{FM_x}\left(G,D_x\right)+L_{FM_y}\left(F,D_y\right) \quad (1)$$

(2) Adversarial loss

The mapping of X to Y is shown in Fig. 1, the model is built on the sample space of two domains, for example, to realize the conversion between X and Y domains, when using CycleGAN model for the mapping of X domain to Y domain, we represent this mapping with G. G is the generator in the generative adversarial network, and we use G to convert the image

in the X domain into the image in the target domain Y domain, and through the discriminator D to judge whether it is is real data, this is a generative adversarial process.

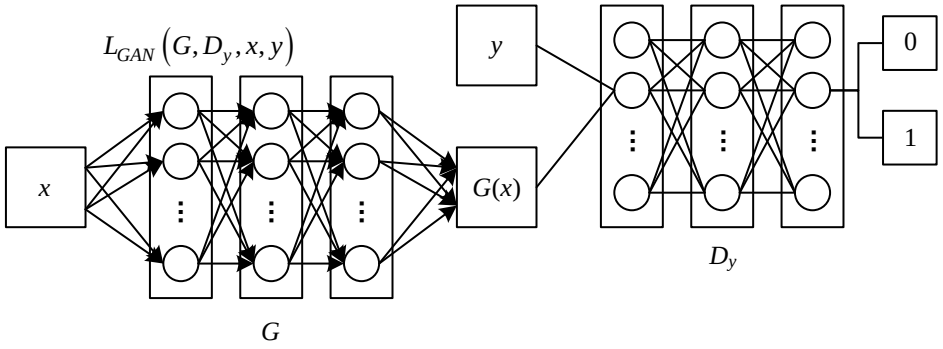


Figure 1: The mapping from X to Y

Corresponding to one of the generative adversarial processes above, a generative adversarial loss is created with the expression:

$$L_{GAN}(G, D_y, x, y) = E_{y \sim p_{data}(y)}\left[D_y(y)\right] - E_{x \sim p_{data}(x)}\left[D_y(G(x))\right] + \lambda E\left[\left(\left\|\nabla_y D_y(y)\right\|_2 - 1\right)^2\right] \tag{2}$$

The Y to X mapping is shown in Figure 2. On the contrary, when CycleGAN training process needs to convert the image in the Y domain into the image in the X domain, let the Y to X mapping be F. F is the generator in the generative adversarial network, and we use F to convert the image in the Y domain into the image in the X domain of the target domain, and then judge it by the discriminator to determine whether it is the real data or not, and this is another generative adversarial process.

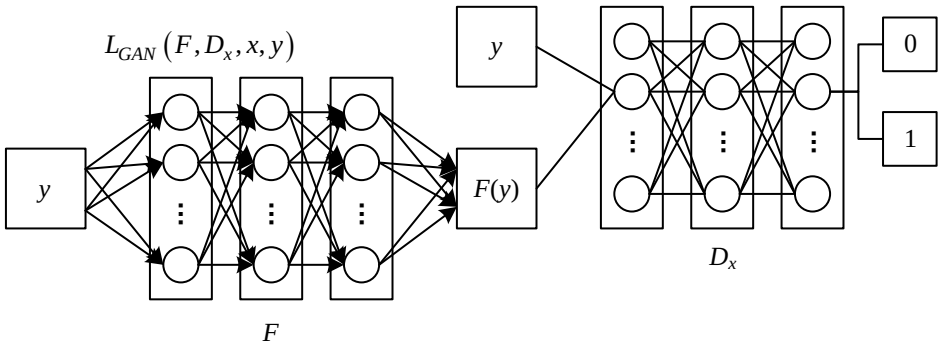


Figure 2: Y to X mapping

Corresponding to the other generative adversarial process above, in building a generative adversarial loss with the expression:

$$L_{GAN}(F, D_x, x, y) = E_{x \sim p_{data}(x)}\left[D_x(x)\right] - E_{y \sim p_{data}(y)}\left[D_x(F(y))\right] + \lambda E\left[\left(\left\|\nabla_x D_x(x)\right\|_2 - 1\right)^2\right] \tag{3}$$

We train two pairs of generative discriminators, so the total adversarial loss can be expressed

as:

$$L\left(G,F,D_x,D_y\right)=L_{GAN}\left(G,D_y,x,y\right)+L_{GAN}\left(F,D_x,y,x\right) \tag{4}$$

(3) Cyclic Consistency Loss with MS-SSIM

The cyclic consistency loss with MS-SSIM is shown in Fig. 3. Without adding the cyclic consistency loss, the GAN may experience the common crash problem of the model when learning the mapping relationship of $X \to Y$. With cyclic consistency loss, the data in the source domain can all be mapped to different places in the target domain. In the X domain by mapping to the Y domain, it also needs to go through the $Y \to X$ mapping relation to be able to map back well, so adding the cyclic consistency loss can be considered as a solution to the model collapse problem. Based on the combination of pairwise learning and GAN, we can train 2 mappings at the same time, for the data on a certain domain through 2 transformations, and finally can generate a distribution close to the initial data, this is the idea of cyclic consistency, because we are trying to transform each other on two domains, so the process is pairwise.

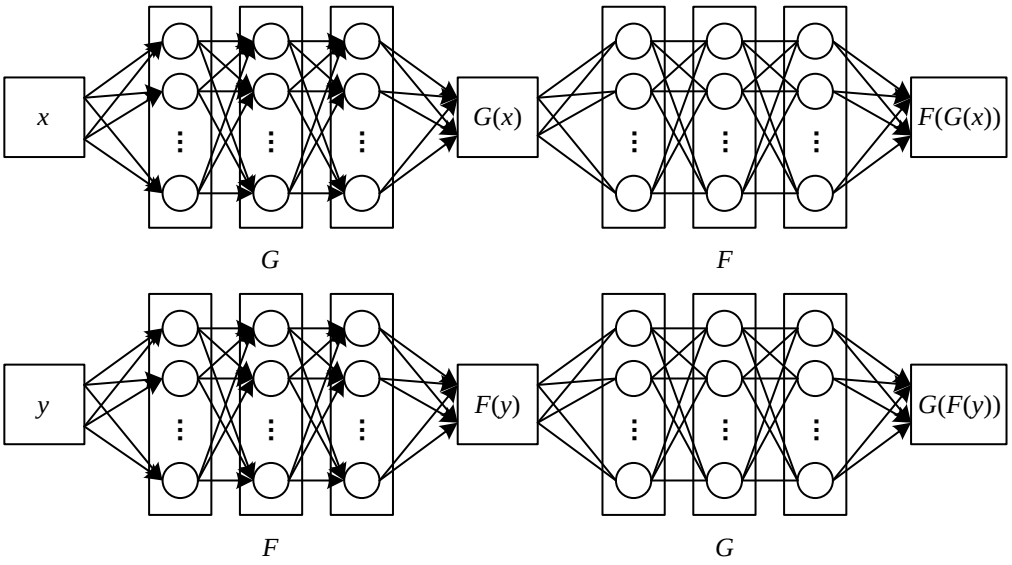


Figure 3: Loop consistent loss with MS-SSIM

In order to offset the defect that the ordinary L2 distance cannot measure the structural similarity of the pictures, we need to improve in the cyclic consistent loss to do the improvement of the structural similarity of the last changed picture with the pre-change one, SSIM loss matches the luminance (1), contrast (c), and structural (s) information of the generated image and the input image and it proves to be very helpful in improving the quality of the image.

Multi-scale SSIM loss is considered as follows:

$$MS-SSIM\left(x,y\right)=\left[l_M\left(x,y\right)\right]^{\alpha M}\prod_{j=1}^{M}\left[c_j\left(x,y\right)\right]^{\beta j}\left[s_j\left(x,y\right)\right]^{\gamma j} \tag{5}$$

Image brightness comparison subsection:

$$l(x,y)=\frac{2u_xu_y+c_1}{u_x^2+u_y^2+c_1} \tag{6}$$

Image contrast comparison section:

$$c(x,y)=\frac{2\sigma_x\sigma_y+c_2}{\sigma_x^2+\sigma_y^2+c_2} \tag{7}$$

Image structure comparison section:

$$l(x,y)=\frac{\sigma_{xy}+c_3}{\sigma_x\sigma_y+c_3} \tag{8}$$

The improved model adds MS-SSIM loss to CycleGAN cyclic consistency to force the similarity between the recovered image and the original image and the cyclic consistency with MS-SSIM is shown in Fig. 4.

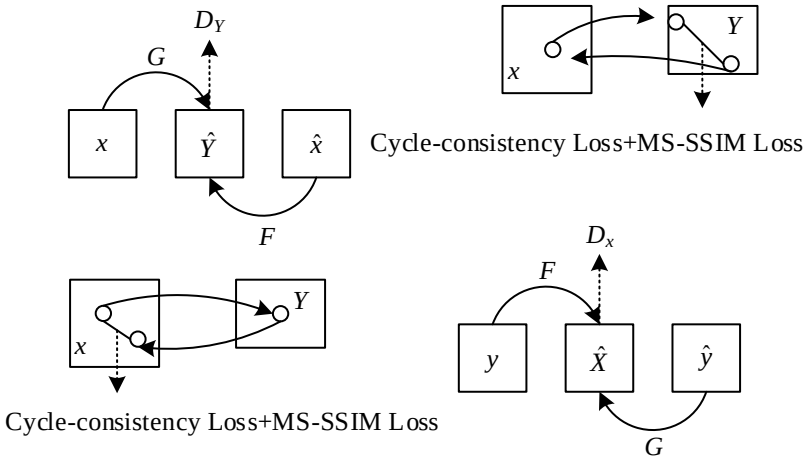


Figure 4: Loops with MS-SSIM are consistent

For the two cyclic reconstruction losses, we consider the structural similarity and L1 losses. $X'=F(G(x))$ and $Y'=G(F(y))$ for cyclic consistent reconstruction of the input image, respectively.

$$L_{SS}=\left(1-MS-SSIM\left(X',X\right)\right)+\left(1-MS-SSIM\left(Y',Y\right)\right) \tag{9}$$

$$L_{L1}=\left\|X'-X\right\|_1+\left\|Y-Y'\right\|_1 \tag{10}$$

So the cyclic consistent loss with MS-SSIM is:

$$L_{cyc+ss}=\lambda_{ss}L_{ss}+\lambda_{L1}L1 \tag{11}$$

where $\lambda_{ss}+\lambda_{L1}=1$, we set $\lambda_{ss}=0.7,\lambda_{\mathrm{L1}}=0.3$ experimentally better here.

2.2.4 Training process

(1) The training dataset of modern Lingnan landscape paintings and western landscape paintings is taken as the input of the network model.

(2) The network model is with two generator network parts and two discriminator network

parts, firstly, we train the two discriminator network parts, and the two discriminator networks discriminate whether it is the real image data or the style-converted image, and optimize the weights of the discriminator networks by the adversarial loss.

(3) The generator weights are optimized according to the feature matching loss, the adversarial loss, and the cyclic consistent loss with MS-SSIM proposed in this paper.

(4) The network model keeps repeating steps (2) and (3) and proceeds to the maximum number of iterations, and the network model training ends.

(5) The improved CycleGAN model can more accurately learn the details, ink rhythm changes and spatial hierarchy of modern Lingnan landscape paintings and western landscape paintings, avoiding only changing the color and ignoring the stroke characteristics. The model training process is outlined as follows: input the image verification set of modern Lingnan landscape paintings and Western landscape paintings, and output the image digitization conversion results.

2.3 Contextual categorization model

2.3.1 Mixed Attention (At)

It can be seen that in the field of natural language processing, the attention mechanism plays a powerful role. In recent years many scholars have also applied this idea to the field of computer vision, and the applied research on the attention mechanism for images has become richer and richer. The hybrid attention module is shown in Fig. 5, which contains two sub-modules, i.e., the channel attention module and the spatial attention module, which act on the input feature map and output the weighted feature map. It should be noted that for a given input image, the two attention modules can play a complementary role, but also need to consider how to combine the channel and spatial attention modules, and in the serial arrangement, the “channel first and then spatial” scheme has a higher accuracy.

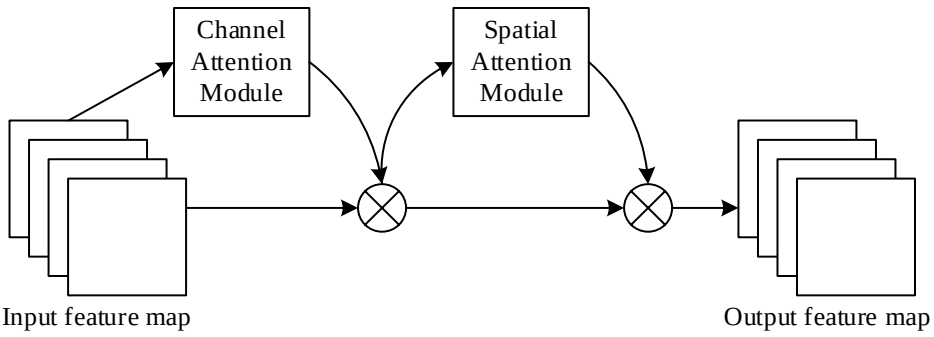


Figure 5: Hybrid attention module

The channel attention module learns the importance of the feature map in terms of the channel dimensions, each channel can be seen as a special feature detector, so the purpose of channel attention is to allow the network to focus on more effective information and to reduce the noise interference and other redundancies, i.e., to pay more attention to the important features in the image.

It can be seen that in the channel attention module, the batch normalization (BN) is used as a scaling factor. As shown in equation (12), this scale factor calculates the variance of the channels and indicates the importance of different channels. Namely:

$$B_{out} = BN\left(B_{in}\right) = \gamma \frac{B_{in} - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \int}} + \beta \tag{12}$$

where $\mu_{\mathcal{B}}$ and $\sigma_{\mathcal{B}}$ are the mean and standard deviation of a small batch of $\mathcal{B}$ and γ and β are used to perform scaling and panning operations on the normalized data. Equation (13) represents the output of the module. Where M_c denotes the output features, γ is the scaling factor for each channel and the weights can be expressed as $W_y = \gamma_i / \sum_{j=0} \gamma_j$. Namely:

$$M_c = sigmoid\left(W_\gamma\left(BN\left(F_1\right)\right)\right) \tag{13}$$

$$M_s = sigmoid\left(W_\lambda\left(BN_s\left(F_2\right)\right)\right) \tag{14}$$

$$Loss = \sum_{(x,y)} l(f(x,W),y) + p\sum g(\gamma) + p\sum g(\lambda) \tag{15}$$

Similarly the BN scale factor can be applied to the spatial dimension to detect the importance of different pixel points in the picture, i.e., pixel normalization (PN). The corresponding spatial attention sub-module is shown in equation (14). Where the output is denoted as M_s, and λ is the scale factor with weight $W_\lambda = \lambda_i / \sum_{j=0} \lambda_j$. The Sigmiod function is used for the activation function of both sub-modules. In order to suppress the less prominent weights, a regularization term is added to the loss function as shown in Eq. (15), where x denotes the input, y is the output, W denotes the network weight, $l(\cdot)$ is the loss function, $g(\cdot)$ is the l_1-paradigm penalization function and p is the balancing term.

2.3.2 Bidirectional gated recirculation units (BiGRU)

In the gated recurrent unit (GRU) on the other hand, there are only two gate control modules, i.e., the update gate and the reset gate, which are used to control the updating and selection of the previous moment's information. Unlike LSTM, gated recurrent unit (GRU) does not have a specialized memory module and achieves higher computational efficiency by simplifying the model structure.

In GRU, each unit module first multiplies the input values x_t and h_{t-1} with the weights, and then computes two gate values between 0 and 1 by the Sigmoid function. The multiplication of h_{t-1} and weights is then combined with the reset gate r_t to obtain the new memory value. The calculation process is shown in Eqs. (16)-(19):

$$r = \sigma\left(W_r x_t + U_r h_{t-1}\right) \tag{16}$$

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1}\right) \tag{17}$$

$$\tilde{h}_t = \tanh\left(r_t * U h_{t-1} + W x_t\right) \tag{18}$$

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \tilde{h}_t \tag{19}$$

where σ is the Sigmoid activation function, * denotes the multiplication of the corresponding elements, and $W_r, W_z, U_r, U_z \in R^{d\times d}$ the weight matrix, which needs to be learned during the training process. In order to process the information in the sequence, this paper adopts a bidirectional gated recurrent unit (BiGRU), because traditional GRUs can only obtain unidirectional information.The structure of BiGRU is shown in Fig. 6, which contains a forward-propagating GRU and a back-propagating GRU.The forward and backward input sequences are obtained as the corresponding hidden-layer representations, respectively. Compared to a single GRU, BiGRU has higher classification accuracy, faster response time and lower complexity.

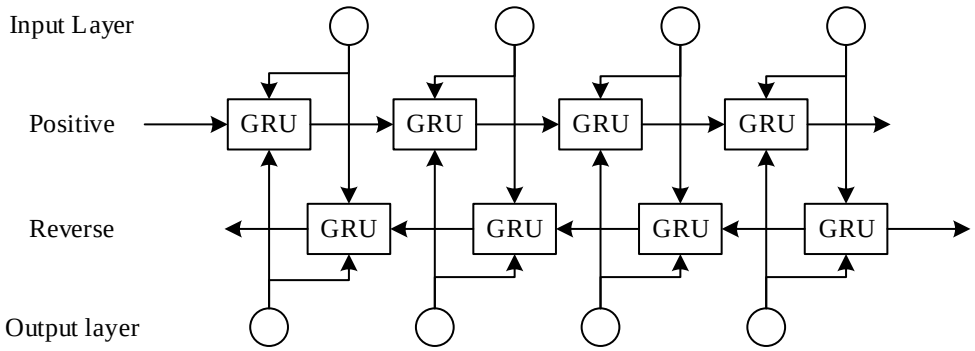


Figure 6: Bidirectional GRU structure model diagram

2.3.3 Convolutional Neural Networks (CNN)

CNN is a deep learning neural network specialized for processing data with grid-like structure such as images, audio and text etc. a CNN mainly consists of a convolutional layer, a pooling layer and a fully-connected layer. In the convolutional layer, the network performs a convolutional operation on the input data, uses a set of learnable filters, performs a sliding window calculation on the input data, and generates a new feature mapping. The role of the convolutional layer is to find patterns by extracting local features of the input data. The pooling layer is mainly used for downsampling and feature compression, which reduces the number of network parameters by downscaling the feature maps, and at the same time can improve the robustness and computational speed of the model. The fully connected layer is a common hierarchy in conventional neural networks and is used to flatten the feature maps output from the convolutional and pooling layers and then feed them into a series of fully connected neurons to generate the final output.

CNN extracts different feature levels of the input data through multiple layers of convolution and pooling layers, and then inputs these feature mappings into a fully connected layer, which ultimately generates an output that classifies or predicts the input data. Due to its excellent feature extraction and processing capabilities, CNNs are widely used in computer vision, natural language processing and speech recognition. In the field of image processing, photographs are usually represented as a three-dimensional tensor using three channels, representing image data in RGB color mode. Therefore, the input data of CNN is a 3rd order tensor. Suppose x is used to represent the input tensor, and c_i is used to represent the feature vector map of the i th layer in the CNN, where $c_0 = x$. Then the process of generating C_i can be formulated in Equation (20).

$$c_i = f\left(c_{i-1} \otimes W_i + b_i\right) \tag{20}$$

In equation (20), f -activation function.

c_{i-1} -feature vector map of the $i-1$th layer in the CNN.

W_i - Weight matrix of the ith layer.

b_i - Bias matrix of the ith layer.

$\otimes$ -Convolution operation.

With the convolution kernel, the CNN is able to perform convolution operations on the input matrix to change the width, height, and depth of the input data. As an important network representation layer, the pooling layer is usually after the convolutional layer, and the pooling operation allows to change the height and width of the component matrices of the input tensor and extract some feature information. Average pooling and maximum pooling are common pooling operations where a 2x2 pooling window is set. Average pooling takes the average of the features at the corresponding positions in the pooling window, while maximum pooling keeps the maximum value in the pooling window as the result.

2.3.4 Activation functions

Neural network activation functions are important components of neural networks and usually operate on the output value of each neuron of a neural network to perform a nonlinear transformation of the neuron output. Their main role is to introduce nonlinearity in the hidden layer of the neural network, allowing the neural network to fit nonlinear functions and improve the expressive and predictive capabilities of the model. Mainstream activation functions are basically composed of nonlinear functions, and in addition to these common activation functions located after each layer, there is another function that is listed separately because of its function and location specificity, Softmax whose formula is:

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{C} e^{x_j}} \tag{21}$$

where x_i denotes the ith real vector of inputs, the number of classes is denoted by C, and $Softmax(x_i)$ denotes the result of normalizing the inputs to a probability distribution. It can be seen that $Softmax$ is a vector that can "compress" a k-dimensional vector containing any real number into another k-dimensional real vector, so that the range of each element is between (0,1). And a function whose sum of all elements is 1. In the deep neural network structure, this is shown as the multi-channel feature maps processed by the convolutional and pooling layers are expanded into one-dimensional vectors in the fully-connected layer, and then a dot product operation is carried out with the corresponding weights in the fully-connected layer, and at the end, a pre-set bias is added to form the input data of the $Softmax$ function of the input data, and finally outputs a normalization operation on the probability distribution of each feature quantity. $Softmax$ Unlike general activation functions that can only perform non-exclusive binary categorization functions, it is able to perform categorization operations on probability distributions of multi-featured quantities and select the categorization with the highest probability as the only target for prediction. It also follows that the position of the $Softmax$ function is fixed to be placed after all the fully connected layers. And if the number of classifications of $Softmax$ is changed to 2, it is clear to see that the $Softmax$ function degenerates into the aforementioned $Softmax$ function, and its output data is no longer exclusive and normalized, and its basic composition of nonlinear functions is dominated by

exponential functions and division, which is similar to that of most mainstream activation functions.

2.3.5 Mathematical modeling

Based on the hybrid attention theory, bidirectional gated recurrent unit, convolutional neural, and activation function above, a contextual classification model (CNN-BiGRU-At) for images of modern Lingnan landscape paintings and western landscapes is designed, which firstly extracts the contextual features of each frame for the input digitized image sequences of modern Lingnan landscape paintings and western landscapes by using convolutional neural network based on hybrid attention mechanism. The contextual features of the modern Lingnan landscape painting and Western landscape painting images are then fed into the bi-directional GRU model to learn the contextual information. The outputs for different moments are aggregated using the attention mechanism, and finally the *Softmax* function is used to classify the contexts of modern Lingnan landscape paintings and Western landscape paintings.

3 Analysis of empirical studies

3.1 Digital conversion analysis

3.1.1 Experimental environment

The experimental platform for this study uses Intel Xeon series E5-4620V4 with 8 cores at 2.6GHz, 32G DDR4 RAM, 1TB SSD and NVIDIA GTX series 2070Ti GPU. The operating system is Ubuntu version 16.04, and the experiments are mainly based on the pytorch learning framework, which contains a variety of development and project research in the complete the ecosystem is an open source deep learning framework with version 1.8.0, CUDA version 7.0, and the programming language used is Python.

3.1.2 Training loss analysis

The loss function comparison of the two methods is used to verify the priority of the improved CycleGAN model in digitizing the conversion of modern Lingnan landscape paintings and western landscape paintings, and the training loss analysis is shown in Fig. 7, in which the number of iterative training is 500 times. Based on the data performance in the figure, the improved method converges faster than the algorithm using CycleGAN directly. By improving the network model, image features can be learned faster. As a result, the model is easier to learn the basic features of modern Lingnan landscape paintings and Western landscape paintings, and the algorithm has less loss in a certain area than before.

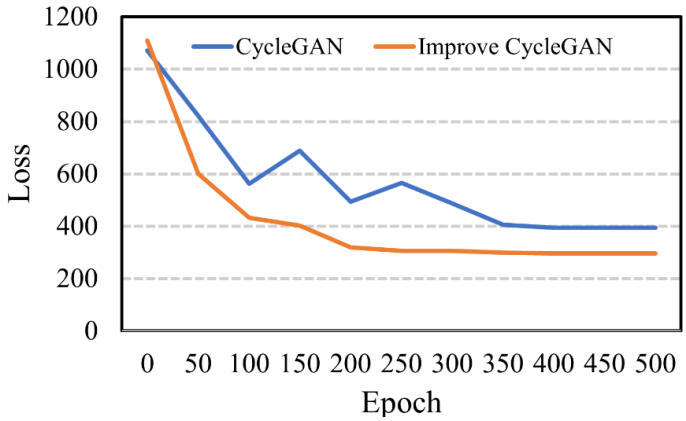


Figure 7: Training Loss analysis

3.1.3 Objective evaluation analysis

Since the digital conversion of modern Lingnan landscape paintings and western landscape paintings has not yet formed a relatively standardized and objective evaluation system, two relatively objective evaluation methods are summarized mainly by referring to the related literature, i.e., GAN-training/GAN-testing and manual evaluation testing. In order to compute GAN-train, the GAN-generated modern Lingnan landscape paintings and western landscape paintings are used to train the classification network, and their performance is evaluated in the test set consisting of real modern Lingnan landscape paintings and western landscape paintings images. Objectively, the difference between the generated digitized images of recent modern Lingnan landscape paintings and western landscapes and the distribution of real recent modern Lingnan landscape paintings and western landscapes images was measured. The classification network is trained to have the features to distinguish different categories of recent modern Lingnan landscape paintings and western landscapes images, and if the generated recent modern Lingnan landscape paintings and western landscapes images can be correctly categorized, it can be judged that the generated images are similar to the real images. In other words, GAN-training is similar to recall measurement, the higher the performance of GAN-training, the more diverse the generated samples are. However, if the accuracy of GAN-training is not high, the quality of the samples will decrease. The second measure, GAN-testing, is the accuracy of the network trained on real recent modern Lingnan landscape paintings and western landscapes images, which is evaluated on the generated recent modern Lingnan landscape paintings and western landscapes images, and the objective evaluation analysis is shown in Figure 8. The data performance in the figure shows that the numerical distribution value range of CycleGAN model is 0.656~0.837, while the numerical distribution value range of the improved CycleGAN model is 0.759~0.941, which demonstrates the priority of this paper's model in digitized conversion of modern Lingnan landscape paintings and western landscape paintings.

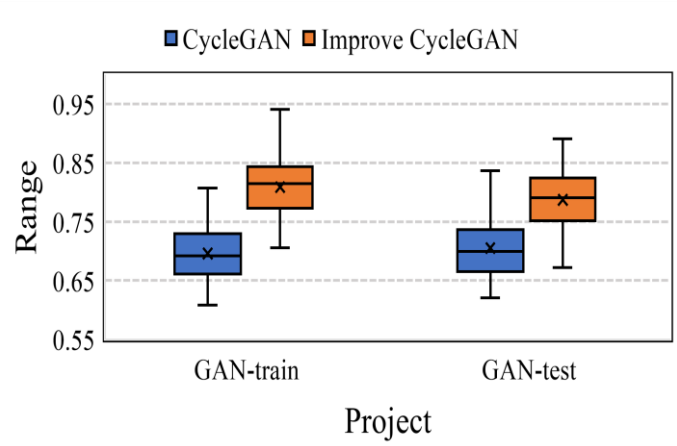


Figure 8: Objective evaluation and analysis

3.1.4 Subjective evaluation

As a subjective evaluation, a visual experience test session was also developed to compare CycleGAN and the improved model proposed in this paper, respectively. Ten images of modern Lingnan landscape paintings and Western landscape paintings were randomly selected from the dataset and digitized using the two models respectively, and the two resulting image sets were named Set I and Set II. When subjects participated in the test, the same digitized images were randomly selected from Set I and Set II, and the subjects were asked to rate which set was better in terms of visual experience.20 subjects aged 15 to 50 participated in the experiment, and all subjects were exposed to a total of 20 of the two sets of images, each corresponding to an unknown model. If one of the groups was judged to be more visually appealing than the other, a point was added without manipulating the other group, and the number of points obtained by dividing the total number of points obtained by each group by the number of points divided by the score was used as a scoring coefficient, and the results of the subjective evaluation are shown in Fig. 9. The magnitude of the values in the figure shows that the use of the model proposed in this paper is favored by young people and has a higher visual recognition with an overall rating coefficient of 83.

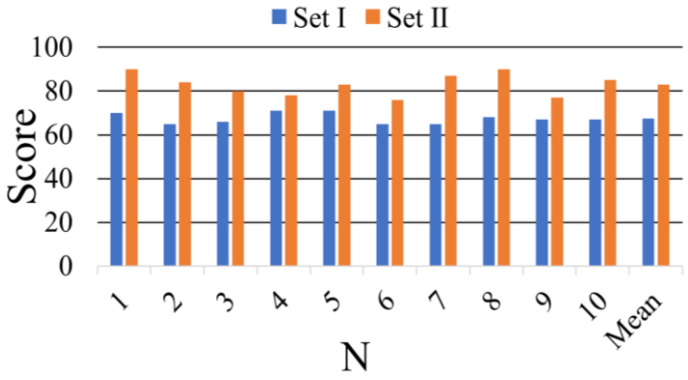


Figure 9: Subjective evaluation result

In order to test the time and space complexity of the improved method, the time cost and space volume of CycleGAN and modified CycleGAN were recorded in the experiments for 100 and 500 epochs, respectively, and the running time and video memory usage on the video card are shown in Table 1. The execution time of the modified CycleGAN barely increases, but the feature information added to the generator results in a 10.57% increase in video memory usage.

Overall, it is cost-effective to trade negligible space resources for improved model performance. The improved CycleGAN model converts the mood characteristics of modern Lingnan landscape painting and Western landscape painting from traditional painting forms to digital images, which not only maintains the color characteristics, but also better preserves the details of brush strokes, hierarchical relationships and spatial composition, enhances the visual realism, and provides an important theoretical basis for the classification of mood in the following section.

Table 1: The running time on the graphics card and the usage of video memory

Method Model	100 epoch	500 epoch	Video memory usage
CycleGAN	128s	1h46min	6836MB
Improvement CycleGAN	136s	1h49min	7599MB

3.2 Image Context Analysis

Through the above analysis, it is known that the improved CycleGAN model has higher priority in the digital conversion of modern Lingnan landscape painting and western landscape painting, which maintains the color features to the maximum extent, and also better preserves the details of the brushstrokes, hierarchical relationships and spatial composition, and enhances the visual authenticity. On this basis, it is inputted into the convolutional neural network based on hybrid attention mechanism, which in turn obtains and extracts the contextual features of each frame of modern Lingnan landscape painting and Western landscape painting images, and then learns the contextual information in the input bi-directional GRU model, and ultimately realizes the contextual classification of modern Lingnan landscape painting and Western landscape painting images. This subsection first demonstrates the feasibility of the context classification model (CNN-BiGRU-At) by constructing a confusion matrix and ablation experiments, and then utilizes the model to carry out a comparative analysis of the context creation of modern Lingnan landscape paintings and Western landscape paintings in a cross-cultural perspective.

3.2.1 Confusion matrix

In image context analysis, TP is the number of positive samples correctly predicted by the model, FP is the number of negative samples incorrectly predicted as positive by the model, FN is the number of positive samples incorrectly predicted as negative by the model, and TN is the number of negative samples correctly predicted by the model. Confusion matrix is a commonly used tool in image context analysis, which is a two-dimensional matrix used to visualize the predictions of a classifier. In the confusion matrix, the rows represent the actual mood categories and the columns represent the mood categories predicted by the classifier. The four cells of the confusion matrix are True Positive Example (TP), False Positive Example (FP), False Negative Example FN), and True Negative Example (TN). Accuracy: Accuracy is the ratio of the number of samples correctly categorized by the classifier to the total number of samples, which measures the proportion of correct categorization by the model. It is calculated by the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

F1-Score: the F1 score is the reconciled average of precision and recall, which combines the performance of these two metrics. Precision rate is the ratio of the number of samples correctly predicted as positive cases by the classifier to the number of samples predicted as

positive cases, and recall rate is the ratio of the number of samples correctly predicted as positive cases by the classifier to the number of samples that are actually positive cases. Recall is the ratio of the number of samples correctly predicted as positive cases by the classifier to the number of actual positive case samples. Their specific formulas are as follows:

$$\mathrm{Re}\,call = \frac{TP}{FP + TN} \tag{23}$$

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FN} \tag{24}$$

$$F1 = 2 * \mathrm{Re}\,call * \frac{\mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} \tag{25}$$

The digitized images of modern Lingnan landscape paintings and western landscape paintings are imported into, the mood classification model (CNN-BiGRU-At) to obtain the confusion matrix, and the results of mood classification are shown in Table 2. The data performance in the table shows that the accuracy, precision, recall, and harmonic mean of this paper's contextual classification model (CNN-BiGRU-At) for color features, stroke details, hierarchical relationship, and spatial composition in the digitized images of modern Lingnan landscape paintings are 0.8360, 0.839, 0.824, and 0.8314, respectively, whereas the accuracy, precision, recall, and harmonic mean of this paper's contextual classification model (CNN-BiGRU-At) in this paper has the accuracy, precision, recall, and harmonic mean of 0.8750, 0.882, 0.866, and 0.8739, respectively, for the color features, stroke details, hierarchical relationship, and spatial composition in the digitized images of western landscape paintings, which are all above 0.80, i.e., it indicates that the contextual classification model has good application efficacy. The model of western landscape painting digitized image mood classification index data is higher than modern Lingnan landscape painting, which also reflects that western landscape painting mood creation is more streamlined than Lingnan landscape painting, resulting in higher index data, while modern Lingnan landscape painting is richer in color features, brushstroke details, hierarchical relationships, and spatial composition, which results in lower model classification index data.

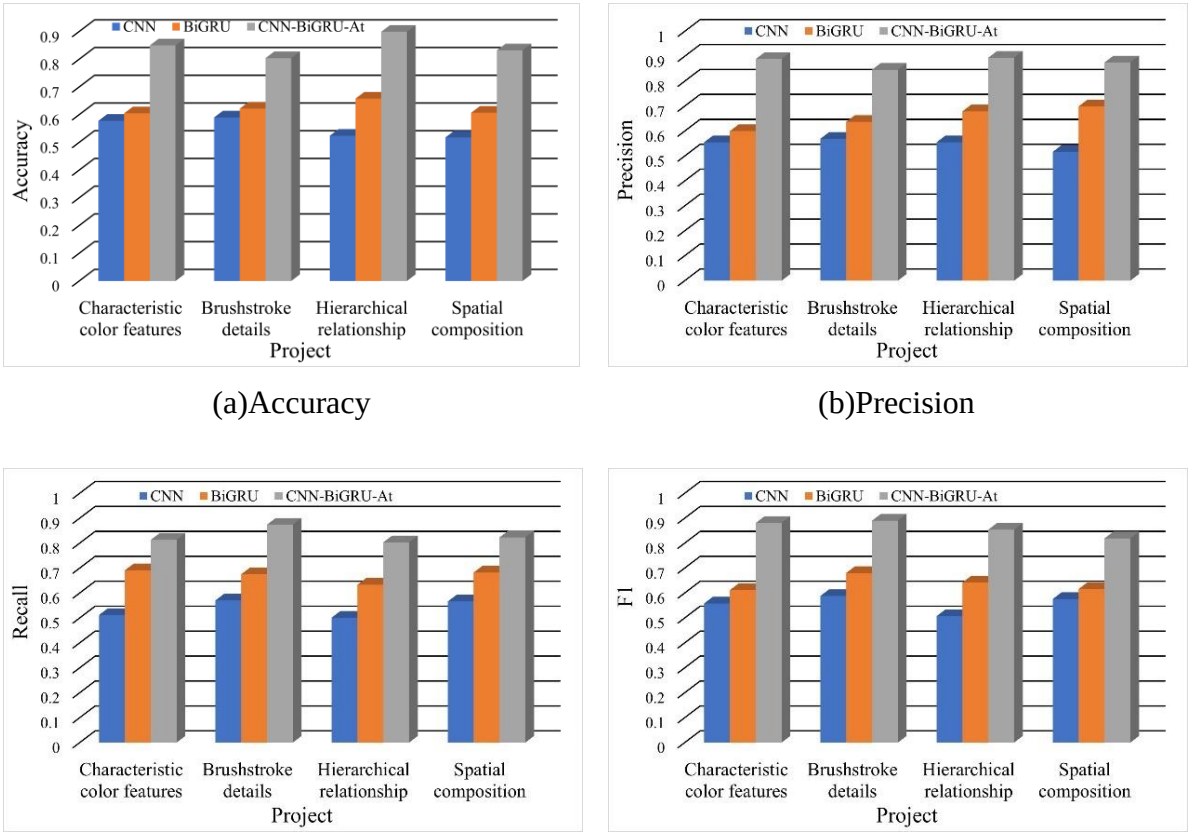
Table 2: Classification results of artistic conception

Model	Characteristic color features	Brushstroke details	Hierarchical relationship	Spatial composition	A	R	P	F1
Lingnan landscape paintings	0.772	0.798	0.961	0.813	0.836	0.824	0.839	0.8314
Western landscape painting	0.793	0.85	0.994	0.863	0.875	0.866	0.882	0.8739

3.2.2 Ablation experiments

The above confusion matrix verifies the overall efficacy of the model, and in order to make the model more convincing, it is also necessary to carry out ablation experiments to analyze it, and the results of the ablation experiments are shown in Fig. 10, in which (a) ~ (d) are the accuracy, precision, recall, and reconciliation mean, respectively. Through the data performance in the

figure, it can be seen that the introduction of hybrid attention (At) and bi-directional gated recurrent unit (BiGRU) on the basis of CNN network improves the accuracy, precision, recall, and reconciliation average of the modern Lingnan landscape painting and western landscape painting mood classification model, and its values are kept above 0.8, which fully verifies the effectiveness of each component of its model.



(a)Accuracy (b)Precision

(c)Recall (d)F1

Figure 10: Results of the ablation experiment

3.2.3 Example analysis

After verifying the validity of the model, the example analysis is formally carried out, using the context classification model (CNN-BiGRU-At) to conduct a comparative analysis of the context creation of modern Lingnan landscape paintings and western landscape digital images, respectively, and the comparative analysis of the context features is shown in Fig. 11, in which A1, A2, A3, A3 show the color characteristics, stroke details, hierarchical relationships, and spatial composition. Through the data size in the figure, it can be seen that the accuracy distribution of the model's contextual classification of digital images of modern Lingnan landscape paintings is 0.5~0.73, while the accuracy distribution of the contextual classification of digital images of western landscape paintings is 0.7~0.9, and the difference in contextual creation between modern Lingnan landscape paintings and western landscape paintings embodies the color characteristics, brushwork details, hierarchical relationships, and spatial composition, due to the fact that the western landscape paintings have the following features in terms of color characteristics, brushwork details, hierarchical relationships, and spatial composition. Since the western landscape paintings are less complex and rich in color characteristics, brushstroke details, hierarchical relationship, and spatial composition, the model classification accuracy is higher, i.e., the western landscape paintings are generally

dominated by simple and abstract mood. On the other hand, the digital images of modern Lingnan landscape paintings have higher complexity and richness in terms of color features, brushstroke details, hierarchical relationships, and spatial composition, which leads to lower model classification accuracy, such as Lingnan landscape paintings of Zhongshan mountains, rivers, waterfalls, lakes, rocks, fishing villages, thatched cottages, pavilions, ancient bridges, morning mists, sunsets, smoky rains, and boat trips.

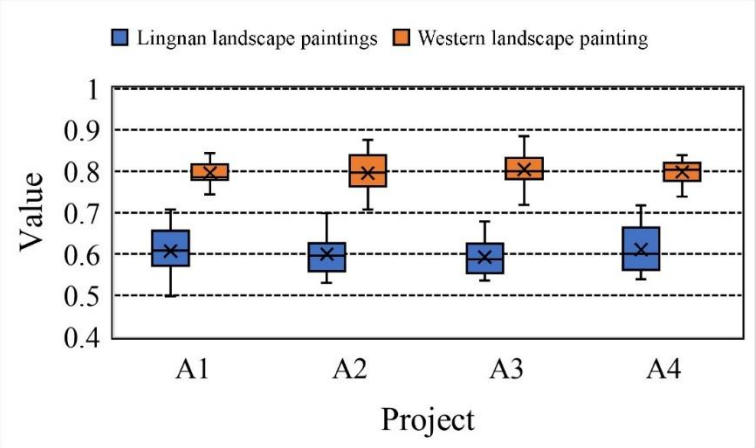


Figure 11: Comparative Analysis of Artistic Conception Characteristics

4 Conclusion

In this paper, under the cross-cultural perspective, it is proposed to use an improved CycleGAN model to digitally transform modern Lingnan landscape paintings and Western landscape paintings. On this basis, a CNN-BiGRU-At based mood classification model is constructed with a view to realizing the purpose of comparing the mood creation between modern Lingnan landscape paintings and Western landscape paintings.

(1) In the digital conversion analysis, it is found that the numerical distribution value range of CycleGAN model is 0.656~0.837, while the numerical distribution value range of the improved CycleGAN model in this paper is 0.759~0.941, which reflects the priority of this paper's model in the digitized conversion of the modern Lingnan landscape paintings and the western landscapes, and at the same time it can maximize the preservation of color features, brushstroke details, hierarchical relationships and spatial compositional features, which lays theoretical support for the analysis of mood.

(2) It is found that the accuracy, precision, recall, and harmonic mean of the contextual classification model (CNN-BiGRU-At) for the color features, stroke details, hierarchical relationship, and spatial composition in the digitized images of Western landscape paintings are 0.8750, 0.882, 0.866, and 0.8739, respectively, while the accuracy, precision, recall, and harmonic mean of the color features, stroke details, hierarchical relationship, and spatial composition in the digitized images of modern Lingnan landscape paintings are 0.8750, 0.882, 0.866, and 0.8739, respectively. , hierarchical relationship, and spatial composition, the accuracy, precision, recall, and reconciliation averages are 0.8360, 0.839, 0.824, and 0.8314, respectively, demonstrating the difference in mood creation between the modern Lingnan landscape paintings and the Western landscapes, where the Western landscapes are generally based on simple and abstract moods, while the Lingnan landscape paintings are generally based on rich and concretized moods.

Acknowledgements

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).

About the Author

Bing Wu, born in 1976 in Lu'an City, Anhui Province, China, obtained his doctoral degree from the Xi'an Academy of Fine Arts in China. He is currently employed at Zhaoqing University in Guangdong Province, China, where his primary research focuses on the creation and theoretical aspects of Chinese landscape painting.

References

[1] Hongmei, J. (2023). The ideological origins and aesthetic construction of yijing (artistic conception). International Communication of Chinese Culture, 10(2), 151-169.

[2] Duggan, T. J. (2007). Ways of knowing: Exploring artistic representation of concepts. Gifted Child Today, 30(4), 56-63.

[3] Gasparyan, S., Sargsyan, M., & Melik-Karamyan, A. (2016). Artistic Concept in the Cognitive Perspective. Armenian Folia Anglistika, 12(2 (16)), 7-14.

[4] Wu, Z. (2025). The free aesthetics of Eastern philosophy, the artistic creation and the influence of Eastern philosophy on contemporary visual art. Trans/Form/Ação, 48(5), e025056.

[5] Zhuo, Y., & Hou, G. (2024). The development of the aesthetic spirit connotation of modern Chinese art. Trans/Form/Ação, 47(6), e02400299.

[6] Antrop, M. (2018). A brief history of landscape research. In The Routledge companion to landscape studies (pp. 1-15). Routledge.

[7] Shamir, L., Macura, T., Orlov, N., Eckley, D. M., & Goldberg, I. G. (2010). Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. ACM Transactions on Applied Perception (TAP), 7(2), 1-17.

[8] Tinio, P. P. (2013). From artistic creation to aesthetic reception: The mirror model of art. Psychology of Aesthetics, Creativity, and the Arts, 7(3), 265.

[9] He, M., Barkeshli, M., & Toroghi, R. M. (2024). CULTURAL DIPLOMACY IN ARTS EDUCATION AND OIL PAINTING: A SYSTEMATIC REVIEW OF EAST-WEST FUSION ARTISTIC EXPRESSIONS. Arts Educa, 41.

[10] Motoyoshi, I. (2022). Climate, illumination, and the style of Western and Eastern Paintings. Art & Perception, 10(3), 244-256.

[11] Bao, Y., Yang, T., Lin, X., Fang, Y., Wang, Y., Pöppel, E., & Lei, Q. (2016). Aesthetic

preferences for Eastern and Western traditional visual art: Identity matters. Frontiers in psychology, 7, 1596.

[12] 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.

[13] Wang, Y., Jiang, Y., Ning, X., & Gao, L. (2024). Bridging cultural perspectives: Developing a sustainable framework for the comparative aesthetic evaluation of Eastern and Western art. Sustainability, 16(13), 5674.

[14] 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.

[15] Libin, Y., Noh, L. M. M., & Abd Razak, H. (2024). Artistic Conception In Meticulous Flower-and-Bird Painting. International Journal of Art and Design, 8(2), 42-55.

[16] Vasileva, Z. (2019). From figurative painting to painting of substance–The concept of an artist. Open Journal for Studies in Arts, 2(2).

[17] Morphy, H. (2020). Becoming art: exploring cross-cultural categories. Routledge.

[18] 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.

[19] Huo, C., & Choi, D. (2024). Exploring Emotional Representation and Interpretation in AI-Generated Art. Asia-pacific Journal of Convergent Research Interchange, 10(6), 533-546.

[20] Augello, A., Infantino, I., Pilato, G., Rizzo, R., & Vella, F. (2013). Binding representational spaces of colors and emotions for creativity. Biologically Inspired Cognitive Architectures, 5, 64-71.

[21] Toxirovna, A. N. (2020). Colors and their artistic image creation features. International Journal on Integrated Education, 3(10), 251-254.

[22] Wang, H. (2021). Expression of Aesthetic Image of Artistic Concept in Artistic Creation. Frontiers in Art Research, 3(8).

[23] Pylypchuk, O. (2024). The impact of structural analysis of works of fine art on enhancing the creativity of artists and interior designers. Arte, Individuo y Sociedad, 36(4).

[24] Zhang, H. (2024). A comparative study on the painting forms of western oil painting and Chinese ink painting. American Journal of Arts and Human Science, 3(2), 39-45.

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