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

Symbolic Expression of Chinese Folk Art Elements in Contemporary Illustration Creation

Yanran Liang[1] *and Yumeng Yan*[1,*]

[1] College of Fine Arts, Education of Guangxi Arts University, Nanning, Guangxi, 530000, China

SUMMARY: *When illustration meets traditional Chinese folk art, how to create a new vision with both cultural roots and contemporary aesthetics is the goal of the research. In this study, we design a set of combined paths for target recognition and style migration. On the one hand, the key features of line curvature, texture pattern (LBP) and spatial frequency are extracted from the images. On the other hand, the style migration model is innovatively adapted to address the asymmetry in the amount of information between folk art and real photographs. Two asymmetric generators with different capabilities are designed and equipped with feature-level cyclic consistency loss and saliency edge loss. The average accuracy (mAP) of the target recognition model in this paper is 57.82% and 53.13% on the mural and illustration datasets, respectively, which outperforms the mainstream detection models. The asymmetric style migration model generates images with a structural similarity of SSIM = 0.7087 to the original photographs, and a content classification accuracy of 70.72%, both of which are substantially better than the comparison methods. Its social acceptance was further explored through a questionnaire survey, with more than 93% of respondents expressing acceptance or great acceptance of this form of integration. They are most looking forward to interacting with it in the form of offline exhibitions (72.08% support) and cultural products (61.04%).*

KEYWORDS: *Chinese folk art; contemporary illustration; feature extraction; stylistic migration; circular consistency loss*

1 Introduction

Contemporary illustration, as an important art field, has experienced leapfrog development and transformation, and has entered people's vision with a variety of artistic expressions [1]. Under the wave of globalization, Chinese contemporary painters have actively absorbed the essence of Western art, explored and integrated different cultural traditions, and injected new vitality and creativity into contemporary illustration [2, 3]. Therefore, the influence of Western art on Chinese contemporary painting is obvious. This injection of cultural elements from painting techniques, forms of expression to the transformation of aesthetic concepts can be seen in the shadows of Western painting genres, such as Impressionism, Expressionism, and Abstract Art [4-7]. By learning and borrowing from Western art, Chinese contemporary painters have been exploring new ways of artistic expression to make their works more modern and internationalized, while contemporary illustration creation has gone through the same stage of cultural integration [8, 9].

It is worth noting that while absorbing foreign art forms, Chinese contemporary painters

*yympeach@126.com
https://doi.org/10.65102/is2026446 Published: 30 April, 2026

also pay attention to drawing inspiration from local culture, among which folk art is one of the sources of artistic inspiration for many contemporary painters [10, 11]. Chinese folk art is a treasure of the excellent traditional Chinese culture, including paper-cutting, shadow, New Year's paintings, embroidery, ceramics, wood carving and other forms. Its main features include distinct ethnicity, simple artistic style, symbolized visual language and life-like subject matter, and these elements not only carry the collective aesthetic interests and cultural memories of the Chinese people, but also unite strong regional characteristics and cultural identity [12-15]. As important symbols of traditional culture, Chinese folk art elements, with their diverse shapes, colors and craftsmanship, have brought more diversified expression possibilities for contemporary painting, and their strong visual impact, intuitive narrative, decorative performance qualities and rich folklore have become important materials for the creation of many contemporary paintings [16-21].

Based on this, folk art elements are often integrated into contemporary illustration creation, thus giving the works a unique flavor and cultural connotation. This integration is not a simple pile or imitation, but through the understanding and expression of traditional culture, it is integrated into the context of contemporary art to form a unique national art style, reflecting a new understanding and new thinking in the context of contemporary painting [22-26].

The movement of traditional styles does not carry the charm of the Chinese folk art elements, but rather the faint resemblance or crude imitation. The root cause of the issue is asymmetry. We will discuss an approach to the solution of art style migration that relies on asymmetric cyclic consistency to solve this main contradiction in the current paper. Rather than depending on the deep learning model self-learning, the critical features derived through human art analysis are actively applied to the training architecture of the modern illustration style migration network. The extraction of the line feature, texture feature and frequency feature is calculated by curvature, LBP texture analysis and frequency analysis respectively. VGGNet-19 style migration network is also used to achieve optimal content loss and style loss. And finally, and where this paper is innovative, an art style migration model built on a generative adversarial network and asymmetric consistency loop is presented. It is given two asymmetric generators, the first one to generate illustrations with fewer details and the second one to reconstruct photos with higher information volume. At the same time, the original pixel matching is replaced with the feature-level cyclic consistency loss that is closer to human visual perception, and the saliency edge loss is introduced to strengthen the stroke portrayal of the main body of the picture, so that the generated illustrations of Chinese folk art elements not only resemble the shape, but also capture the charm.

2 Fusion of image feature extraction and deep illustration style migration network

2.1 Image feature extraction algorithm

The selection of features and the constraints of loss function have crucial importance. In this section, we will analyze the characteristics of image features and image semantic content, propose feature representations that can effectively express Chinese folk paintings, and use some algorithms to extract these features.

For the extraction of useful features that can capture the style of an image, this research paper uses three types of features: line, frequency, and texture features. These are explained as under:

2.1.1 Line feature extraction

Line is considered the most basic component of Chinese paintings. This means that artists use line in order to describe, abstract, and generalize natural subjects through the use of two dimensional planes in order to represent three dimensional spaces. Another role of line in Chinese painting involves modeling which is achieved through lines of different thickness. Color, on the other hand, usually acts as the supplementary factor since it does not take into account elements such as light and shade. Irrespective of whether the artist uses landscapes, flowers, birds, or human beings as his/her subjects, he/she often draws lines and then fills them in using colors. The curvature is used to represent the line fluidity F_{line}, as shown in equation (1):

$$F_{line} = \frac{(1+f_x^2)f_{yy} + (1+f_y^2)f_{xx} - 2f_x f_y f_{xy}}{(1+f_x^2+f_y^2)^{3/2}} \tag{1}$$

where x and y denote the coordinates of a pixel in the image and $f(x,y)$ is the gray value of the pixel. The $f_x, f_y, f_{xy}, f_{xx}, f_{yy}$ are the first-order, second-order, and mixed partial derivatives of $f(x,y)$, respectively, and F_{line} is the Gaussian curvature of the pixel.

2.1.2 Texture Feature Extraction

Texture refers to the pattern that is present on the surface of an object and is normally used to describe the smoothness or roughness of the object's surface in common language usage. The term encompasses the regular change of colors that appears on the object's surface. In art drawing, the texture appears inside the contour lines of an object and can be illustrated either by line drawing or background information. Texture is regarded as a visual feature in the field of computer vision that has lots of pixels and obeys some rules. These texture features fall into four groups: statistical texture features, modeling texture features, signal processing texture features and structural texture features. There are multiple methods of extracting texture including statistical methods, image structure methods and model methods. In this experiment, the LBP method is used to extract texture.

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} 2^p S(i_p - i_c) \tag{2}$$

where (x_c, y_c) represents the center element of the 3x3 field, and his pixel values are i_p, i_c represent the other pixel values in the field. $S(x)$ is the sign function, defined as follows:

$$s(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & else \end{cases} \tag{3}$$

2.1.3 Frequency feature extraction

Image frequency is a frequency in space domain that denotes the variability of grayscale information in the image. Images have different frequencies that are significant and influence the structure of the image in different ways. The lowest frequency portion of the image is the most important aspect of the image, as it forms the basis of the grayscale information and impacts the image structure less. Middle frequency part is the structural elements and edge data of the image. The high frequency portion of the image has edges and details of the image that add more information to middle frequency information of the image.

2.2 Style Migration Networks

After the extraction and analysis of the styles contained in the traditional Chinese paintings' images, this part will discuss about the deep learning approaches in transferring style. A style transfer network based on VGGNet can be used to achieve a successful separation and combination of the sematic content and styles through the definition of the two loss components, which becomes the basis for further improvement in asymmetric transfer tasks.

The suggested style transfer network will use VGGNet to extract the textures of the style image and the content of the content image. The network constructs two loss functions-content loss and style loss and derives the total loss of the objective through the sum of the above loss functions. Eventually, the loss minimization in iteration leads to the generation of style transferred images. VGGNet contains a variety of subtypes, where VGG-16 and VGG-19 are the most widely used. Figure 1 shows the model of VGG-19.

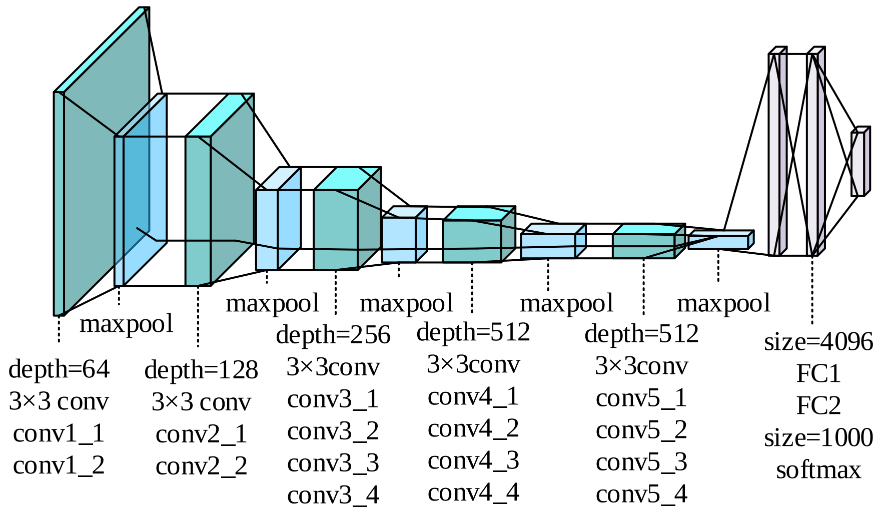


Figure 1: VGG-19 structure diagram

The VGG-19 model consists of 16 convolution layers and 3 fully connected layers. In the training phase, the VGGNet does not update any weights; instead, it is only used to calculate the losses corresponding to both content and style layers. The input image x is initially very noisy. As a result of minimizing the loss function and back propagation during training, the noisy input x becomes a final image that resembles the image p in terms of content and resembles image a in terms of style after the training process. To accurately calculate the loss function, it is essential for the network to have the same structure as VGGNet.

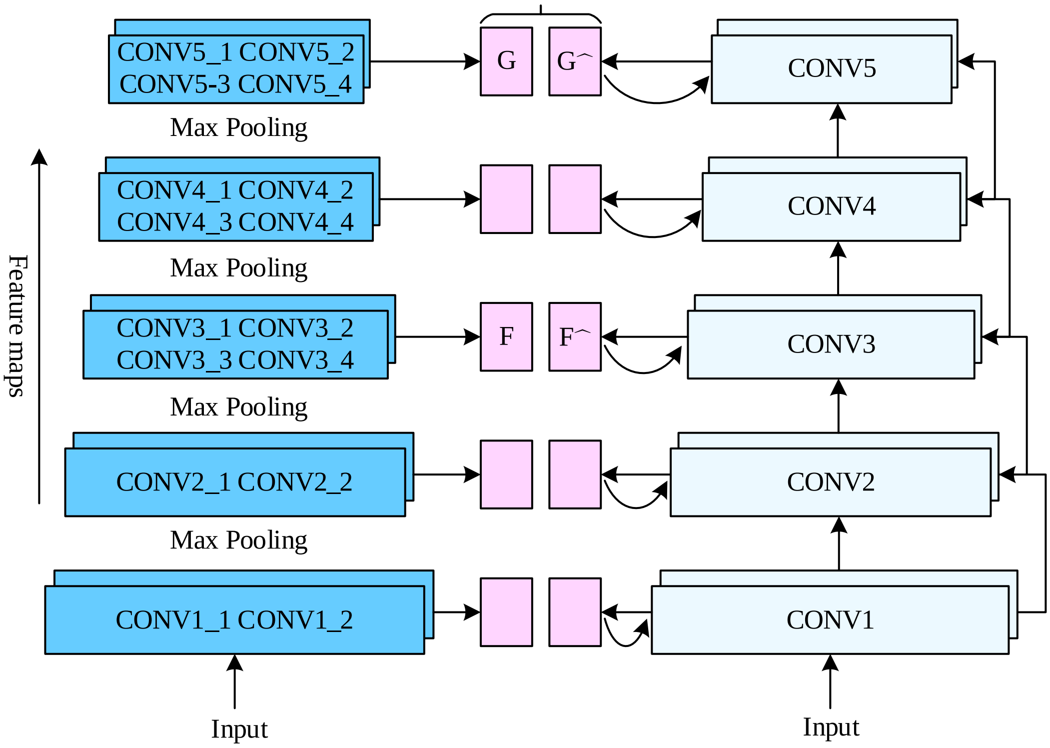


Figure 2: Style rendering network training process

where the content loss function is defined as follows:

$$L_{content}(p, x, l) = \frac{1}{2}\sum_{i,j}\left(F_{ij}^{l} - P_{ij}^{l}\right)^{2} \tag{4}$$

Both P and F denote a matrix that consists of the column vectors of all the feature maps in a particular layer of the network. P_{ij}^{l} denotes the activation value of the image to be generated x at the j position of the i-th featuremap in the l-th layer. F_{ij}^{l} denotes the activation value of the content image p at the j position of the i th featuremap in the lth layer. Where P matrix is a matrix of the to-be-trained of the column vectors of the featuremap of a layer in the network, while the F matrix is composed of the column vectors of the featuremap of the corresponding layer in VGGNet.

The concept of style is a little complex, thus we should explain the Gram matrix which is described as the matrix of the inner product of two vectors that are present in a -dimensional space between any vectors. The inner product of two vectors is multiplication of two vectors followed by addition of the results.

$$G_{ij}^{l} = \sum_{k} F_{ik} F_{jk} \tag{5}$$

G_{ij}^{l} is the inner product of the vectors consisting of the i and j feature maps in layer l of the network, and k is the corresponding element of the feature map, for a given layer, the inner product of the feature maps i and j is in fact the i rows j columns of the Gram matrix of the element values in the Gram matrix. Since the inner product of two vectors can determine the angle and direction between the vectors, it is often used in style comparison. The loss function for each style layer is defined as follows:

$$L_{style}^{l}(a,x)=\frac{1}{4N_l^2M_l^2}\sum_{i,j}\left(G_{ij}^{l}-A_{ij}^{l}\right)^2 \tag{6}$$

N_l is the number of feature maps in the layer and M_l is the size of each feature map. A_{ij}^{l} denotes the value of the style picture a in the i row and j column of the Gram matrix consisting of all the feature maps in the lth layer. The role of $\frac{1}{4N_l^2M_l^2}$ is mainly normalization, mainly to make the order of magnitude of style loss close to the order of magnitude of content loss. In practical coding, we often take multi-layer networks for style comparison, and then weight them together to get the final style loss, as shown in the following formula:

$$L_{style}(a,x)=\sum_{l}w_l L_{style}^{l}(a,x) \tag{7}$$

where w_l denotes the style weights for each layer of the network, and the final loss function is defined as shown in the following equation:

$$L_{total}(p,a,x)=\alpha L_{content}(p,x)+\beta L_{style}(a,x) \tag{8}$$

where α and β represent the weights of content and style loss respectively. If it is necessary to highlight the content image in the final generated image, a larger weight is given to α; if it is necessary to highlight the style image, a larger weight is given to β.

2.3 Artistic Style Migration for Generative Adversarial Networks Based on Asymmetric Cyclic Consistency

Aiming at the problem of asymmetric information richness in the image domain that exists in illustration style migration, this section proposes an asymmetric cyclic consistency structure based on CycleGAN. By introducing generators with different representational capabilities, cyclic consistency loss at the feature level, and saliency edge loss, the model is better adapted to the asymmetric transformation task from real natural images to Chinese folk illustration styles. The network model framework of the whole algorithm is shown in Fig. 3.

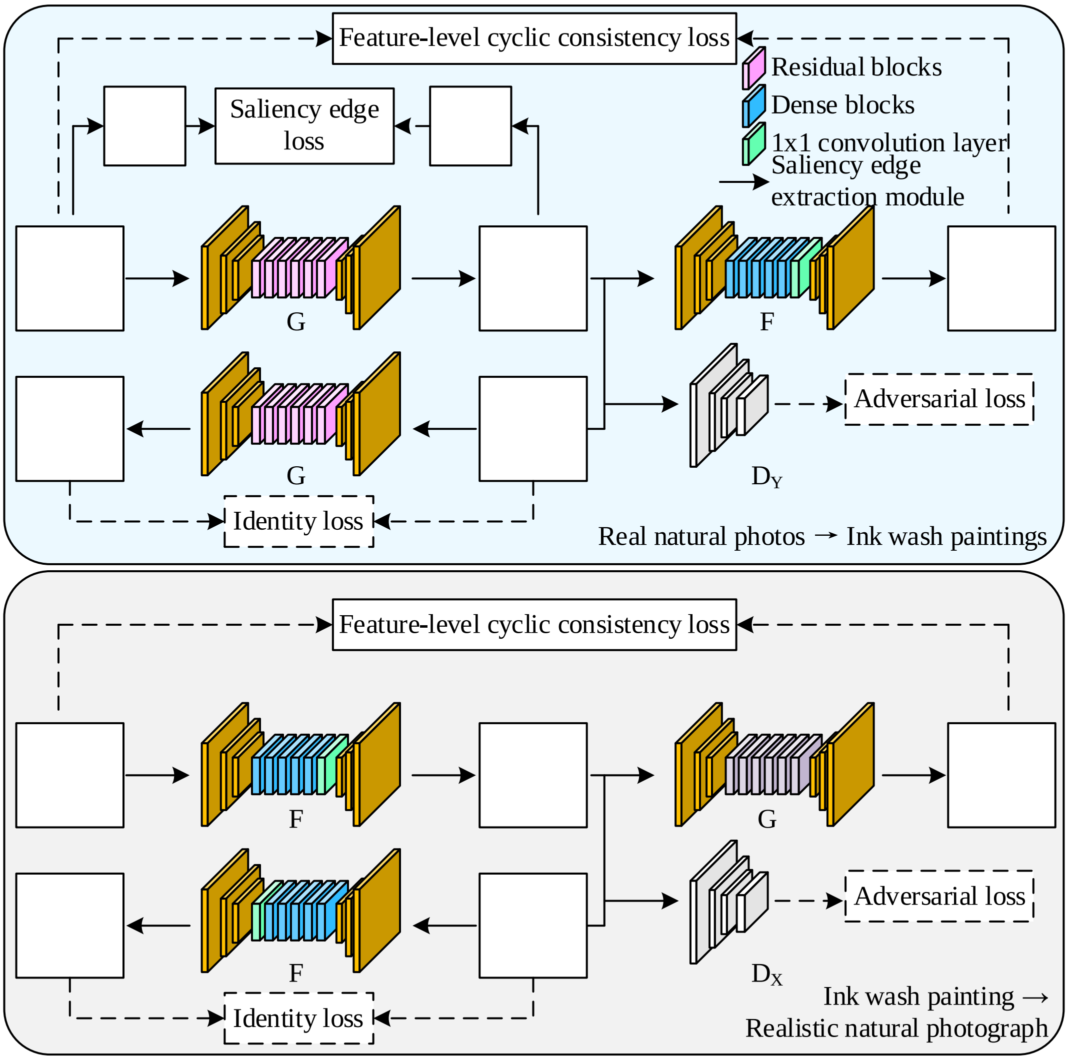


Figure 3: Illustration style transfer framework of asymmetric cyclic consistency

2.3.1 Quantitative analysis of information richness asymmetry in the illustration style migration task

In order to evaluate quantitatively the extent of information richness asymmetry within the illustration style migration task that includes Chinese folk art components, this experiment calculates the average entropy value of images belonging to different image domains. The entropy value measures the information content of a monochromatic image. Entropy of an image domain can be described as:

$$mEntropy = \frac{1}{N}\sum_{n=1}^{N} Entropy\left(img_n\right) \tag{9}$$

where N denotes the total number of images in an image domain, img_n denotes the image instances in the image domain, and $Entropy\left(*\right)$ denotes the calculation of image entropy value on the input image, and the detailed definition of the formula is shown in Equation (10).

$$Entropy = -\frac{1}{c}\sum_{c}\sum_{i=0}^{255}\sum_{j=0}^{255} P_{(c,i,j)} \times \log P_{(c,i,j)} \tag{10}$$

where c denotes the channel index of the HSI color space channel, i and j denote the length and width positions of the image, and $P_{(c,i,j)}$ denotes the probability of the number of occurrences of the pixel at the corresponding position over the total number of pixels.

To explain more intuitively the discrepancy between image entropies of two image domains within an image translation task, in this study, the information entropy ratio of the two domains is also calculated, which is given below:

$$EntropyRatio = \frac{\max(mEntropy(X), mEntropy(Y))}{\min(mEntropy(X), mEntropy(Y))} \tag{11}$$

where X and Y denote different image domains in an image conversion task.

The entropy ratios in illustration style migration and asymmetric image translation tasks surpass those in the case of symmetric image translation. As a result, it can be inferred that the illustration style migration task displays asymmetry in terms of domain knowledge and can hence be classified as an asymmetric image translation task.

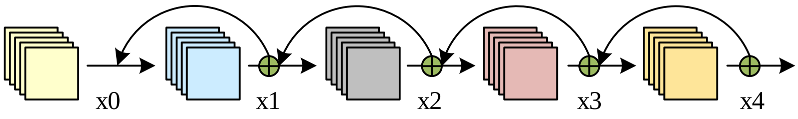
2.3.2 Asymmetric cyclic consistency structure

The illustrated migration model in this chapter is based on the theory of asymmetric cycle consistency and consists of two generators, two discriminators, and a saliency edge extraction (SEE). Generator G relies on a residual block design, whereas generator F adopts a dense block design. The two discriminators, D_X and D_Y are used for evaluating the authenticity of natural images and illustrations, respectively. The SEE algorithm is designed for extracting saliency edges from the input images.

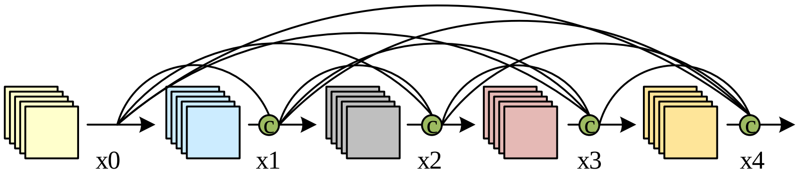
(1) Generator

In light of the symmetric cycle-consistency mechanism of CycleGAN, this research further improves the design of the generator models G and F , aiming at tackling the imbalance in information density in illustration style transfer, which integrates features of Chinese artworks. Specifically, the generator model F, serving as the main transformer, is designed using Dense Blocks due to its stronger ability to represent features compared to the other generator model G.

The densely connected nature that is associated with Dense Blocks supports the propagation of features and also allows for feature reuse, which makes it possible for the transformation function to adequately describe the generator whenever a Dense Block replaces a Residual Block. The designs for the residual block and the dense block are depicted in Fig. 4. Whereas residual blocks have a short-circuit nature to their design, the densely connected nature in Dense Blocks allows for concatenation of features through multiple layers in the channel feature dimension.



a) Residual blocks based on short-circuit connection mechanism (where + denotes feature-level addition operation)



b) Dense blocks based on dense connection mechanism (where C denotes channel-level connection operation)

Figure 4: The network architecture of residual blocks and dense blocks

A transformation module based on the residual block is adopted in the generator G, while a transformation module based on the dense block is designed for generator F, which implies a necessity for improving the capability of representation. Such a structure helps to generate more realistic styled illustrations with the whole network.

(2) Discriminator

The main task for the discriminator is to differentiate between the synthesized image and the real image. In this paper, the 70×70 PatchGAN is chosen as the discriminator. It includes several convolution layers and generates an $n \times n$ probability matrix, where the final authenticity score is calculated by averaging all elements in the probability matrix.

(3) Significance Edge Extraction Module

For extracting salient features of the main strokes of illustrated paintings, saliency edges are adopted to represent these main strokes. Therefore, it is proposed to use the saliency edge extraction module, which contains both the saliency detector and edge detector modules. The approaches to the saliency detection have been shown to be capable of modelling the human visual and cognitive processes, as well as discovering salient areas in the picture that are not identical to any other parts.

The neural network that has been pre-trained is PFAN which is applied to determine masks of salient objects of the image in regional level. Following that, a very effective edge detector network, HED, is utilized to find edges in the image. Edge maps based on saliency edge extraction network are used in the calculation of saliency edge loss in the training process. The architecture of the saliency edge extraction network is illustrated in Figure 5.

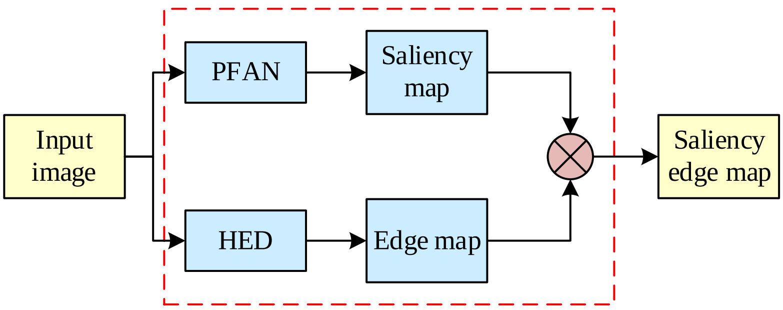


Figure 5: Significant Edge Extraction Module

2.3.3 Loss function

On the basis of the original CycleGAN loss, this study introduces feature-level cycle consistency loss and saliency edge loss.

(1) Adversarial loss

It should be noted that the main aim of the adversarial loss function is to maximize the performance of both the generator and the discriminator in order to improve their abilities in generating and discriminating, respectively. Adversarial loss is formulated as follows:

$$loss_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y\sim Y} \log D_Y(y) + \mathbb{E}_{x\sim X} \log(1 - D_Y(G(x))) \tag{12}$$

where X and Y represent the data distributions in the domain of real natural photographs and the domain of Dunhuang fresco images, respectively. An identical adversarial loss $loss_{GAN}(F, D_X, X, Y)$ is applied to the style migration from the illustration image domain to the direction of the real natural image domain.

(2) Identity loss

Identity Loss Function is used to avoid any insignificant modifications by the model, as well as ensuring that the output image maintains the same color distribution as the input image. The formula for the identity loss function is described as follows:

$$loss_{Identity}(G, F, D_X, D_Y) = \mathbb{E}_{x\sim X}\left[\left\|F(x) - x\right\|_1\right] + \mathbb{E}_{y\sim Y}\left[\left\|G(y) - y\right\|_1\right] \tag{13}$$

(3) Cyclic consistency loss based on feature level

The cycle consistency loss acts to impose constraints on the output images that force them into the desired domain and also allows unpaired data to be used in training. For this project, a feature-level cycle consistency loss is adopted. In particular, the L1 distance on the pre-trained features in the higher layers of the VGG network is adopted as a substitute for the standard pixel-level L1 loss.

$$\begin{aligned} loss_{Cycle}\left(G, F, X, Y\right) = \mathbb{E}_{x\sim X}&\left[\left\|VGG_{conv4\_4}(F(G(x))) - VGG_{conv4\_4}(x)\right\|_1\right] \\ +\mathbb{E}_{y\sim Y}&\left[\left\|VGG_{conv4\_4}(G(F(y))) - VGG_{conv4\_4}(y)\right\|_1\right] \end{aligned} \tag{14}$$

where $VGG_{conv4\_4}(*)$ represents the output features of the conv4_4 layer in the pre-trained VGG network, which can effectively characterize the content structure information of the image.

(4) Significance edge loss

In addition, a saliency edge loss term is proposed for modeling the style of illustration images that have strong body strokes. For both the real natural image and the generated illustration image, a saliency edge loss can be calculated using the saliency body edge map that is extracted from the two images. The mathematical formulation of the loss function is as follows:

$$\begin{aligned} loss_{Salient\_Edge}(G, X) = \mathbb{E}_{x\sim X}&\left[-\frac{1}{N}\sum_{n=1}^{N} \alpha E(x)_n \log\left(E(G(x))_n\right)\right. \\ &\left. +(1-\alpha)(1 - E(x)_n) \log\left(E(G(x))_n\right)\right] \end{aligned} \tag{15}$$

where N is the total number of pixels in the input image corresponding to the saliency subject edge map, $E(\cdot)$ is the edge extraction model HED network, and α is a balanced weight factor.

(5) Total target loss

Finally, the total target loss of the model is calculated in equation (16).

$$
\begin{aligned}
&L(G,F,D_X,D_Y,X,Y)=\lambda_1 loss_{GAN}(G,D_Y,X,Y) \\
&+\lambda_2 loss_{GAN}(G,D_X,X,Y)+\lambda_3 loss_{Cycle}(G,F,X,Y) \\
&+\lambda_4 loss_{Identity}(G,F,X,Y)+\lambda_5 loss_{Salient\_Edge}(G,X)
\end{aligned}
\tag{16}
$$

Optimize the following objective function

$$G^* = \arg\min_{G,F,D_X,D_Y} L(G,F,D_X,D_Y,X,Y) \tag{17}$$

3 Experimental validation of fusion methods for feature recognition and style migration

This chapter is divided into two major experimental segments: first, evaluating the performance of target recognition based on the image feature extraction algorithm of Section 2.1 on a specially constructed dataset; and then turning to the assessment of the generative capability of the asymmetric style migration model to comprehensively test the effect of the style migration model in terms of multiple dimensions such as hidden spatial features, quality of the generated images, and style genre differentiation.

3.1 Research on target recognition capability based on image feature extraction algorithm

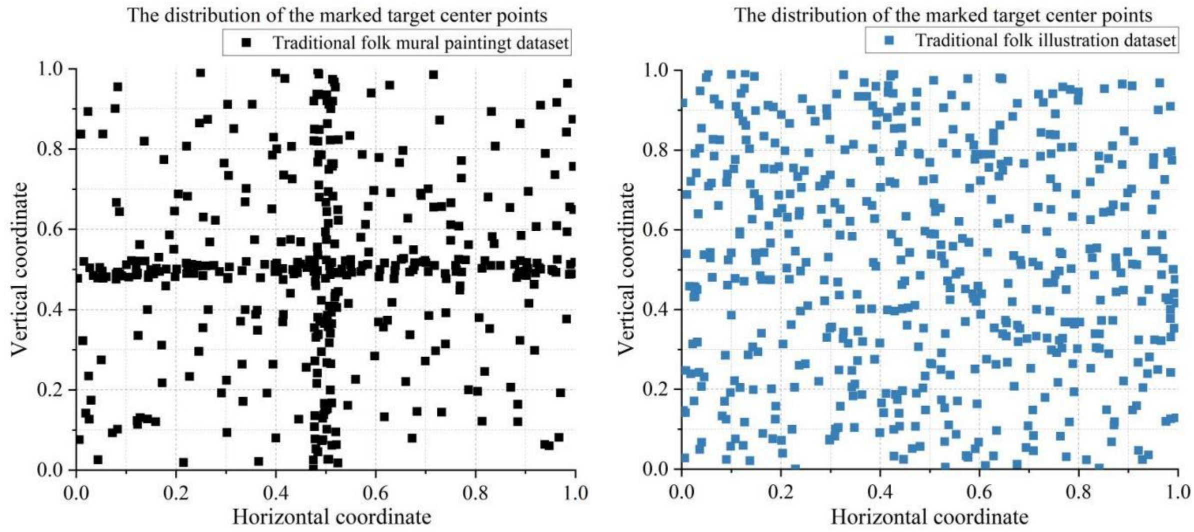
3.1.1 Introduction and Analysis of Data Sets

In order to investigate the ability of the image feature extraction algorithm based on curvature computation, LBP texture and frequency analysis proposed in this paper for target recognition and extraction of Chinese folk art elements, two datasets are designed for the study.

One is the folk mural target detection dataset, which contains 1,352 images and 3,728 target detection labels, and is classified into 56 categories. The number distribution of each type of target is more balanced, which is suitable for multi-scale target detection.

The second is the Universal Folk Illustration Target Detection dataset, which comes from the digitized resources of classic folk art illustrations from different regions and categories in China, and contains 6271 images and 16738 target detection labels. The core features of this dataset are instance-level, multi-category, and high-density. It is more in line with this paper for capturing the physical art elements of illustrations.

In order to study the relative position of the detected elements in the original image in each dataset, this paper plots the distribution of the center point of the labeling box. The original image takes the center as the origin, and the width and height of the image are proportionally divided into lengths of 1. The points in the distribution map indicate the relative coordinates of the center point of the annotation box on the whole image. The distribution of the centroids of the labeling frames on the mural and illustration datasets is shown in Figure 6.



(a)Tradition folk mural painting dataset (b)Tradition folk mural illustration dataset

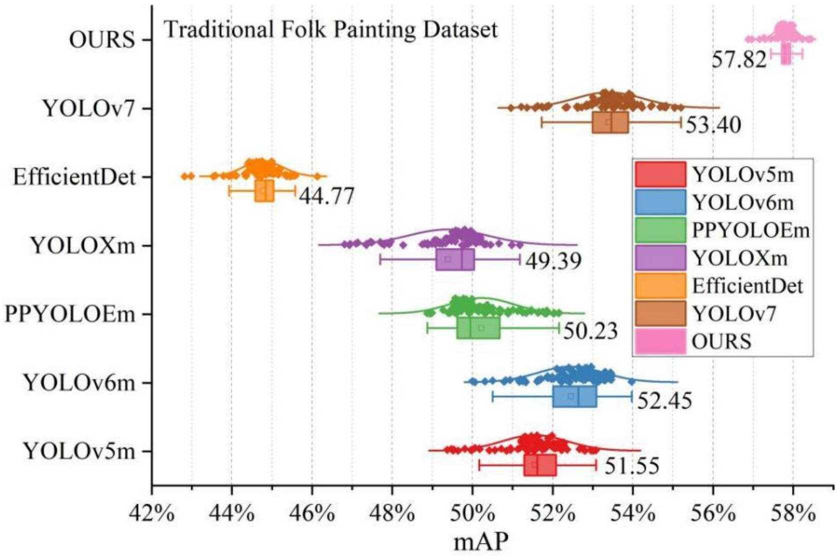
Figure 6: The distribution of the annotation center points

In the folk mural painting dataset, black dots are densely distributed near the vertical center axis of the image, forming a distinct vertical aggregation band. This is due to the fact that art elements in murals are usually carefully arranged in the central region of the image, and the composition is concerned with symmetry and subject prominence. This centered distribution makes the target relatively independent with little background interference, which is conducive for the model to capture local features, but it may also allow the model to focus too much on the center of the picture and not be sensitive enough to the elements in the edge regions and the global compositional relationships.

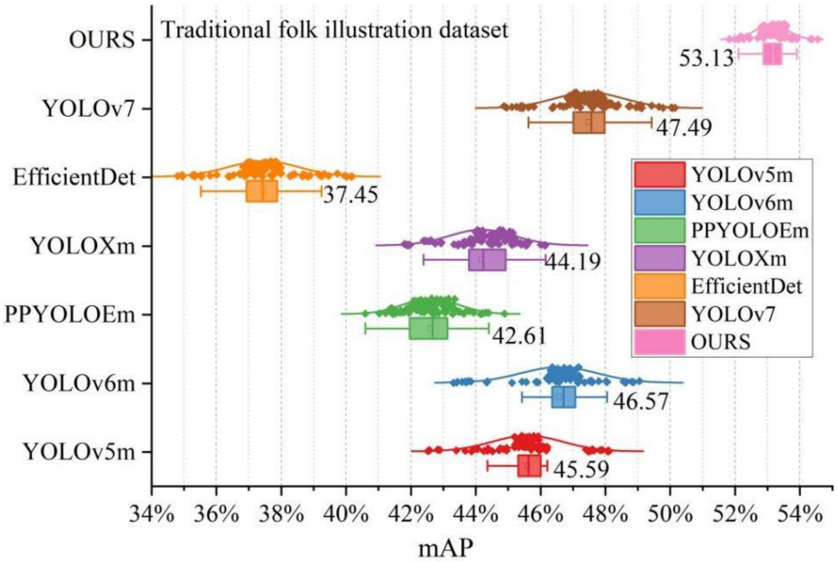
In the generic folk illustration dataset, on the other hand, the distribution of the centroids of the labeling frames is more dispersed, and the labeling frames cover all effective regions of the image. This is due to the fact that folk illustrations have more flexible and free compositions, rich elements and often appear in the form of combination and nesting, and the overall density of the image is high. This distribution puts higher demands on the model, which must have the ability to capture details and understand complex spatial relationships at the same time.

3.1.2 Comparative Experiments on Chinese Folk Element Detection

For evaluating the performance of the current state-of-the-art algorithms used for extracting features of images used in this research, comparative experiments were carried out using two folk painting datasets for several object-detection frameworks. In particular, the frameworks chosen for comparisons include YOLOv5m, YOLOv6m, PPYOLOE-M, YOLOXm, EfficientDet, and YOLOv7. The experimental setup was made identical for all networks, and the two datasets were trained for 100 epochs in order to measure their performance through mean average precision (mAP). Comparative results for these seven models are shown in Fig. 7 for the datasets of folk murals and illustrations, respectively.



(a)Tradition folk mural painting dataset



(b)Tradition folk mural illustration dataset

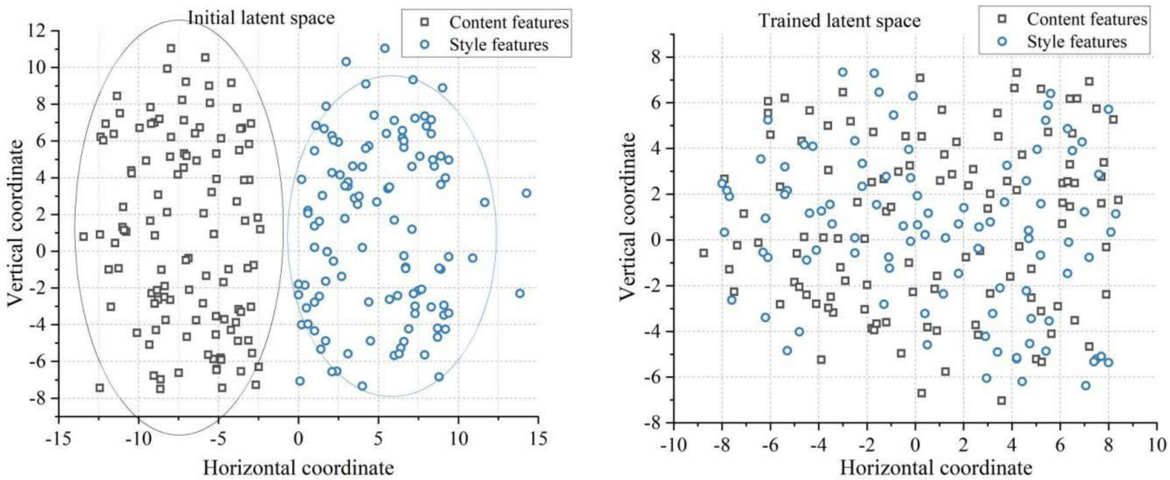
Figure 7: The experimental results of the 7 models

The method in this paper is clearly ahead in both detection tasks. On the mural dataset, the mAP of the image feature extraction-based method reaches 57.82% on average, and the results are relatively concentrated in 100 training sessions, and the model performs stably, which is 4.42 percentage points higher than that of the next best performer, YOLOv7; on the more challenging illustration dataset, the advantage of this paper's model extends to 5.64 percentage points, with the former having an accuracy of 53.13%, and the latter having a mAP = 47.49%. The leading edge of this paper's method is greater on the illustration dataset, which is more difficult for target detection, indicating that by integrating the feature extraction strategy of curvature, texture and frequency analysis, the model becomes capable of understanding the unique line and texture organization of folk art, and accurately identifying the complex folk art elements. Moreover, the distribution of the experimental results of 100 training sessions in the figure shows that the accuracy results of YOLOv7 and other models are larger, indicating that the performance of other YOLO models in the target detection task is not stable, which again highlights the strong stability and robustness of this paper's method.

3.2 Experiments based on an asymmetric circular consistency style migration study

3.2.1 Visualization of hidden spatial features

Second, the performance of the suggested art-style migration loss in the framework of a generative adversarial network under the cycle-consistency assumption is considered. To enhance the migration of style when transitioning between the domain of real-world images and that of folk-art element illustrations, this paper incorporates two additional losses namely feature-level cycle-consistency loss and saliency-edge loss when training the generator network. The loss, which is the distance between the generated image and the target image related to the style, is determined based on the latent space. With latent space training, it may be possible to use the encoder to learn content-image data depending on the associated style.
The distribution of the latent-space features of both the content and style images during pre-training and training are illustrated in Fig. 8. Specifically, the former one is shown in Fig. 8(a), while the latter one is demonstrated in Fig. 8(b).



(a)Initial latent space (b)Trained latent space

Figure 8: The latent space features of the content and style images

From the diagram in Figure 8, we see that initially the content characteristics and the style characteristics are independent of each other, with the content characteristics being mostly found to the left-hand side of the diagram and the style characteristics on the right-hand side of the diagram, where x>0. After training the latent space using the cycle-consistency loss function, the encoder maps the content image to a latent space close to the style characteristics. From Figure 8(b), it can be seen that there is a clear crossover between these two sets of characteristics. From this analysis, we see that the encoder no longer just separates domains. Instead, through the strict constraint of feature-level cycle-consistency loss, it learns to encode the structure of the content image near the stylistic characteristics manifold.

3.2.2 Analysis of style migration comparison results

In order to quantitatively examine the effectiveness of the asymmetric cycle-consistency style migration approach for extracting and transferring the characteristics of the Chinese folk art style, this paper designs a new image dataset which contains images from three folk art styles, namely Van Gogh, Picasso, and Lôerh, each containing 1,000 images, for the purpose of evaluating contemporary illustration.

Five metrics are used for the quantitative measurement of the proposed style migration technique in terms of structural similarity (SSIM), perceptual image quality (HOSA), and Fréchet inception distance (FID). Six baselines will be compared with our approach, including Gatys, AdaIN, AvatarNet, WCT, Sanakoyeu, and CycleGAN. Detailed experimental results of the above-mentioned seven methods on our dataset are summarized in Table 1.

Table 1: The experimental results of the 7 models on the dataset

	SSIM	HOSA	FID	Classification Accuracy/%
Gatys	0.3539	20.574	120.623	4.811
AdaIN	0.3135	24.714	138.825	38.633
AvatarNet	0.3873	40.335	171.687	27.402
WCT	0.2523	26.785	180.892	10.377
Sanakoyeu	0.3027	23.824	203.705	13.271
CycleGAN	0.5471	30.134	146.484	47.916
OURS	0.7087	45.629	121.966	70.723

From the results, it is evident that the model developed in this research provides images that possess the closest resemblance to the actual photos concerning their overall structure and contour. This means that the problem of structural distortion associated with most models in style transfer has been adequately addressed using the proposed model. Considering the HOSA measure that evaluates the perceptual quality of the images, it has been observed that the proposed model achieves the highest HOSA score of 45.629 and comes first, followed by AvatarNet with a HOSA score of 40.335. The high HOSA score suggests that the images produced by both models have high perceptual quality. For the evaluation of style similarity using FID, it was found that the proposed model ranked second with a low score of 121.966, barely ahead of the Gatys approach. Nevertheless, the images generated by the proposed model have closer similarities with real-life folk-art images concerning their style distribution. As far as the classification accuracy is concerned, the proposed model attained a score of 70.723%, the highest among all the models. The method in this paper finds an excellent balance in the multi-objective task of style resemblance, content clarity, picture beauty, and identification accuracy through the design of asymmetric structure and targeted loss function.

3.2.3 Classification Genre Confusion Matrix

In order to further analyze the classification accuracy of the models on different style genres, four models with high classification accuracy, namely, AdaIN, AvatarNet, CycleGAN, and this paper's asymmetric cyclic consistency-based art style migration, are selected to study the style classification of its three genres, namely, Van Gogh style, Picasso style, and Loehl's style, and Fig. 9 illustrates the classification confusion of the four models Matrix.

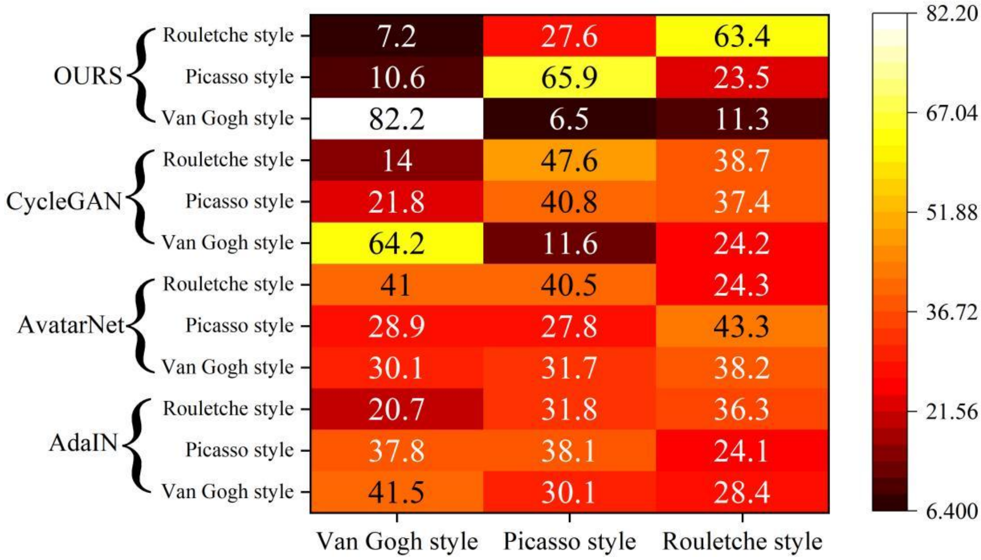


Figure 9: Confusion matrices of the 4 models

The correct classification accuracies of both AdaIN and AvatarNet models for the three genres of Van Gogh, Picasso and Roerich styles are generally low, mostly hovering around 30%-40%, and the errors are very scattered. That is, for 1000 Van Gogh style images, only 415 were correctly categorized by AdaIN and retained the content information for style migration, while 378 and 207 were misidentified as Picasso or Roerich styles, respectively. The style features learned by these two models are vague and the internal representations are confusing, making it difficult to clearly distinguish these three genres.

In contrast, CycleGAN's performance is greatly improved. It achieves a 64.2% correct rate for Van Gogh style image recognition, proving that its stylization process produces more discriminative features. However, it still has significant mutual confusion between Picasso and Roerich styles, each with about 37% misclassification.

In contrast, this paper's improved model, based on an asymmetric circular consistency structure, demonstrates more accurate discrimination. The accuracy rates of the three genre styles are 82.2%, 65.9% and 63.4%. It shows that the asymmetric structure and feature level loss effectively strengthen the uniqueness of the style representations. The model correctly categorized 634 Rohreh style images, but at the same time there were still the remaining 276 images in this category that were misclassified as Picasso style. In the model's learning feature space, there may be some kind of visual or semantic similarity between the Roerich style and the Picasso style itself, e.g., both contain certain structural brushstrokes, leading to easy confusion in the model's categorization.

4 Feasibility Analysis of Contemporary Illustration Creation Incorporating Chinese Folk Art Elements

The fusion between folk culture and modern visual representation has become inevitable, and such a phenomenon has played a role in the maintenance and dissemination of the folk culture. The current research, based on the aforementioned analysis, has developed a questionnaire for the evaluation of the possibilities for innovative design concerning the symbolism of Chinese folk arts in modern illustration. A total of 154 effective responses were collected from the questionnaire survey through offline means.

The questionnaire survey analyzing the feasibility of incorporating Chinese folk arts in illustration focuses on two aspects: the respondent's familiarity with illustration and the respondent's expectation towards the incorporation of Chinese folk arts in illustration creation.

4.1 Exposure to illustration

Exposure to illustrations was investigated in terms of the degree of frequency of exposure to illustrations and the source of exposure. The results are shown in Figure 10.

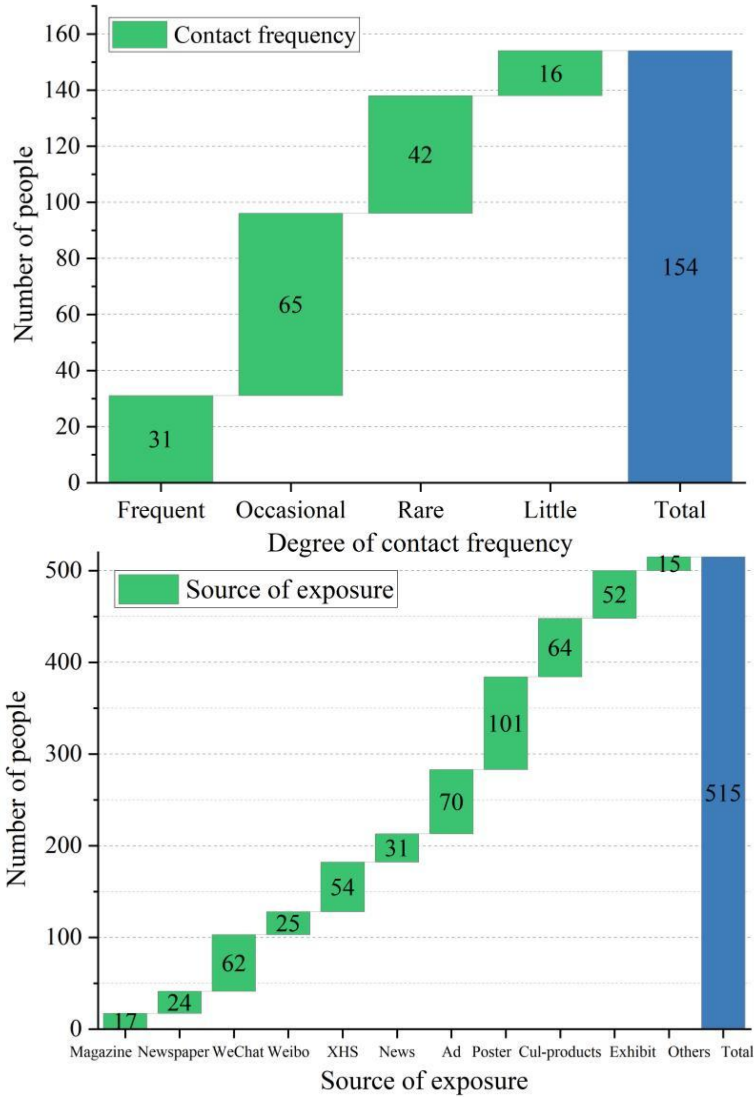


Figure 10: Investigation and analysis of the exposure to illustrations

More than 42% of the respondents were exposed to illustration only occasionally, and only 31 of the 154 people surveyed were exposed to illustration on a regular basis, or 20.13%. Illustration is not common to many people. "Posters" and "advertisements" are the most important forms of exposure to illustration, accounting for 65.58% and 45.45% respectively, confirming the strong penetration of illustration in commercial publicity and public vision. There is also no lack of people who come into contact with illustration through social media platforms such as WeChat and Xiaohongshu. What's more, 52 people said they came across illustration through offline exhibitions. Together, these channels depict the multifaceted ecology of contemporary illustration communication, and also provide a carrier reference for subsequent fusion creation.

4.2 Expectations for the Integration of Folk Art Elements into Illustration Creation

The expectations about incorporating folk art elements in illustration creation primarily pertain to two aspects: (1) the extent of willingness to incorporate folk art elements in the illustration design process, and (2) the extent of preference for this type of illustration form. The findings are depicted in Figure 11.

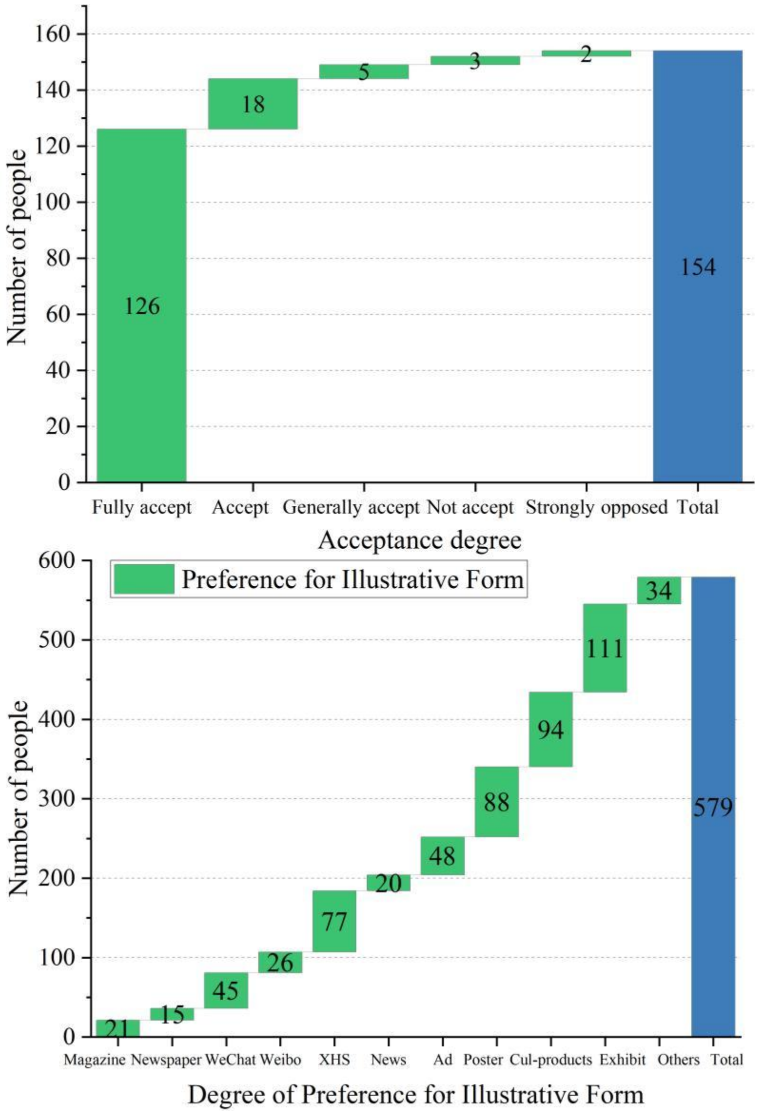


Figure 11: The expectations for integrating folk art elements into illustration creation

The results of the survey show that the public is highly receptive to and positively expects the incorporation of folk art elements into contemporary illustration. In terms of the degree of acceptance, 81.82% of the respondents said “very much”, and with 11.69% of the respondents “accept”, the total acceptance rate is over 93%. This shows that this innovative direction of cultural integration has a deep psychological foundation in the audience.

In terms of favorite forms, “offline exhibition” and “cultural products” are the most popular, with 111 and 94 audiences respectively, suggesting that the public is not only willing to watch, but also looks forward to experiencing and owning. Meanwhile, among digital media, the high proportion of “Xiaohongshu” and “WeChat” indicates an effective path for online visualization. And “posters” are still important, with 88 people preferring this form of illustration. The preference data from the survey provides valuable market insights into the contemporary use of folk art elements in illustration, from creative positioning to communication strategies.

5 Conclusion

Through the combination of “feature guidance” and “asymmetric migration”, this study realizes the in-depth analysis and high-quality stylized regeneration of Chinese folk art elements, and verifies its feasibility in terms of technical performance and social acceptance.

In terms of the generation quality of style migration, this paper's method reaches 40.629 on the HOSA index, which reflects the naturalness of the image, and the style of the generated result is the closest to the real folk art distribution, with FID=121.966. When classifying three different style genres, the model in this paper correctly recognizes specific styles up to 82.2%, which is nearly 20 percentage points better than the average performance of the next best CycleGAN. This suggests that asymmetric structure with feature loss can efficiently distinguish different style genres.

The questionnaire survey explains the public's expectation of this type of fusion creation, clearly pointing to physical scenes such as offline exhibitions and cultural products that can be experienced as owned. It reveals that the technological endpoint should not just be a generated image, but a revitalized interface that can be integrated into cultural consumption and empower creative industries.

About the Author

Yanran Liang was born in Nanning, Guangxi, China, in April 1990. She obtained a doctoral degree from Silpakorn University in Thailand. She is currently working as a lecturer at Guangxi Arts University in China. Her main research directions are design and illustration.

Yumeng Yan was born in Suqian, Jiangsu, China, in January 1990. She obtained a doctoral degree from Silpakorn University in Thailand. She is currently working at Guangxi Arts University in China as a senior arts and crafts master. Her main research directions are visual communication and decorative patterns.

References

[1] Blaiklock, D. (2025). Contemporary Illustration: Understanding. The Language of Art and Artists, 155.

[2] 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.

[3] Lyu, Q., & Sangiamvibool, A. (2025). Cultural implications and educational literacy of the fusion between Chinese landscape painting and impressionism. Academic Committee (Peer Reviewers), 405.

[4] Wang, L., & Wang, B. (2025). Between emptiness and reality: from the white space in Chinese painting to the philosophy of form in western art. Trans/Form/Ação, 48(3),

e025010.

[5] Sun, S., & Wang, K. (2023). An Introduction to the Influence of Neo-expressionism on Contemporary Chinese Oil Painting. Journal of Humanities, Arts and Social Science, 7(8).

[6] Yang, G. (2021). The imagery and abstraction trend of Chinese contemporary oil painting. Linguistics and Culture Review, 5(S2), 454-471.

[7] Jiang, F., & Daud, M. (2024). The Application of Impressionist Colour Techniques Contemporary Chinese Painting. Pakistan Journal of Life & Social Sciences, 22(2).

[8] Xuan, J. (2023). Chinese Painting: Exploration and Comparison. International Journal of Arts & Humanities Studies, 3(2).

[9] Li, L., & Wang, Y. (2021, January). The Integration and Development of Chinese Style Illustrations and Western Painting Aesthetics. In The 6th International Conference on Arts, Design and Contemporary Education (ICADCE 2020) (pp. 238-242). Atlantis Press.

[10] He, Z. (2023). Exploration on Localization of China's Oil Paintings-Research on Image Oil Painting. Art and Performance Letters, 4(8), 40-43.

[11] Hu, X., Lai, Y., Zhao, D., Tong, F., Hu, Y., & Li, Y. (2022). Ceramic Painting and Traditional Cultural Element Fusion Composition Design Based on Virtual Reality. Journal of Nanomaterials, 2022(1), 3781448.

[12] Sun, K., & Zhao, X. (2025). The Inheritance and Development of Different Ethnic Religious Cultures in Chinese Folk Art. Cultura: International Journal of Philosophy of Culture and Axiology, 22(3).

[13] Lijuan, F., & Rahman, A. R. A. (2024). A bibliometric analysis of review on folk art. Cogent Arts & Humanities, 11(1), 2426363.

[14] Yang, Z., Liu, Y., Wang, W., Li, Y., & Yuan, X. (2024). An empirical study of feature extraction for Hubei folk art of carved paper-cutting. Plos one, 19(10), e0311923.

[15] Asif, M., & Ali, M. (2019). Chinese traditions folk art, festivals and symbolism. International Journal of Research, 6(1), 1-20.

[16] Yuanyuan, L. I. (2024). THE ART OF WOODBLOCK NEW YEAR PAINTINGS IN ZHUXIAN TOWN, CHINA FROM THE CULTURAL DIMENSIONS OF CONNOTATION, BELIEFS, SYMBOLS AND FOLK ART. Procedia of Multidisciplinary Research, 2(10), 8-8.

[17] Liu, C., Bakar, S. A. S. A., & Ismail, I. (2024). Application Study of Chinese Traditional Mural Materials in Comprehensive Material Painting. International Journal of Creative Future and Heritage (TENIAT), 12(2), 1-16.

[18] Li, S., & Ghani, D. B. A. (2024, March). Modern Visual Presentation of Traditional Patterns in Chinese Folk Art. In Proceedings of the 3rd International Conference on Culture, Design and Social Development (CDSD 2023) (Vol. 834, p. 390). Springer Nature.

[19] Lin, Y., & Zhang, D. (2024). Historical Inheritance and Folklore Memory-Development and Innovation of Imagery Expression in Chinese Painting. Cultura: International Journal of Philosophy of Culture and Axiology, 21(1).

[20] Peipei, L., & bin Mohd Rafee, Y. (2024). Integrating Miao Traditional Elements into Chinese Figure Painting: Exploring Significance and Cross-Cultural Exchange. Art and Society, 3(1), 1-12.

[21] Yi, D., & Saichai, K. (2024). The Application of Meishan Cultural Symbols in Painting. Journal of Roi Kaensarn Academi, 9(10), 1867-1884.

[22] Zhou, H. (2019, November). The Application of Chinese Folklore Intangible Cultural Heritage Elements in Contemporary Commercial Illustration Design. In 3rd International Conference on Art Studies: Science, Experience, Education (ICASSEE 2019) (pp. 245-248). Atlantis Press.

[23] Gu, J., Zhang, W., Wang, X., Zhou, Q., Zhang, J., Xie, F., ... & Wang, H. (2024). Bridging cultures: Chinese elements in scientific illustrations. Chinese medicine, 19(1), 103.

[24] Ying, X. (2018, December). The Creative Value of Zhuhai Folk Art Illustration. In 5th International Conference on Education, Language, Art and Inter-cultural Communication (ICELAIC 2018) (pp. 585-587). Atlantis Press.

[25] Hu, Y. (2017, December). Inheritance and Innovation of Chinese Folk Art Elements in Modern Illustration Design. In 2017 International Conference on Art Studies: Science, Experience, Education (ICASSEE 2017) (pp. 215-217). Atlantis Press.

[26] Qi, F., Fauzi, T. A., & Yahaya, S. R. (2025). Representing tradition: The construction of culturally-specific visual narratives in Chinese picture books and hand scroll paintings. Journal of Early Childhood Literacy, 25(3), 607-636.