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

Research on Meticulous Flower-and-Bird Painting and the Expression of National Cultural Elements in Chinese Painting Creation Based on Deep Learning

Yawan Chen[1] *and Lihong Qin*[1,*]

[1] Guangxi Minzu Normal University Chongzuo 532200, Guangxi, P.R.China

SUMMARY: *Under the background of continuous integration of digital art generation and intelligent inheritance of traditional culture, how to maintain the style of meticulous flower-and-bird painting and accurately express national cultural elements in Chinese painting creation has become an important topic of intelligent art research. This paper constructs a deep learning method framework for Chinese painting creation, and designs the system from visual semantic feature representation, multi-modal feature extraction and fusion, style generation constraints to interactive feedback closed loop. The experiment is based on 3600 valid images and 10800 groups of samples after amplification. The results show that the style similarity of the complete model reaches 0.901, the accuracy of cultural element expression reaches 91.8%, the coordination degree of composition reaches 89.6%, and the user satisfaction score reaches 9.1. In addition, the accuracy of cultural expression still maintains 86.7% under 30% disturbance. The research shows that this method can improve the style stability, cultural consistency and creation auxiliary value of digital generation of meticulous flower-and-bird painting, which has positive significance for the intelligent creation of Chinese painting and the digital communication of traditional culture.*

KEYWORDS: *Deep learning; Meticulous flower-and-bird painting; National culture element; Chinese painting creation*

1 Introduction

With the continuous evolution of generative artificial intelligence, computer vision and deep learning technologies, digital art creation is moving from pure image generation to a new stage that focuses on style understanding, cultural expression and human-computer collaborative design. As an important carrier of Chinese excellent traditional culture, Chinese painting has distinctive Oriental artistic characteristics in its ink language, modeling mode and aesthetic spirit. Among them, meticulous flower-and-bird painting is famous for its rigorous shape, subtle color and concise image. It not only carries the formal aesthetic sense of flower-and-bird theme, but also contains the auspicious concept, etiquette consciousness, regional aesthetic and national culture symbol. In the context of digital creation, how to accurately extract the visual features of meticulous brushwork flower-and-bird paintings with the help of deep learning technology, and further realize the effective integration and orderly expression of national cultural elements, has become a problem worthy of in-depth discussion in the research of intelligent art generation.

From the existing research, image generation model, style transfer method and multi-modal

*13207800742@163.com
https://doi.org/10.65102/is2026224 Published: 30 April, 2026

fusion technology have achieved good results in the fields of oil painting generation, illustration design, image inpainting and visual reconstruction, which provide a technical path for the digital creation of Chinese painting. However, Chinese painting, especially meticulous brushwork flower and bird painting, is not a natural image or plane decoration image in the general sense. Its line drawing structure, color hierarchy, white space relationship and artistic conception creation have strong regularity and cultural characteristics. If only relying on the general visual model for feature learning, it is often prone to problems such as weakening of ink texture, convergence of schema expression, and shallow embedding of cultural symbols, resulting in the generation of results that are close to Chinese painting in form, but difficult to reach a high level in spiritual temperament and cultural connotation. We believe that the expression of national cultural elements is not simply superimposed patterns, colors or symbols, but needs to establish a deeper correspondence between the theme organization, composition logic, modeling language and aesthetic meaning.

Based on this, we focus on the problems of visual semantic modeling and the expression of national cultural elements in the creation of meticulous flower-and-bird painting, and try to construct a method framework of Chinese painting based on deep learning. We will conduct research from the aspects of data representation, multi-modal feature extraction, style generation constraints and interactive feedback mechanisms, and explore how to improve the accuracy, coordination and generation quality of national cultural elements while maintaining the artistic characteristics of meticulous flower-and-bird painting. By introducing the computer modeling method and experimental evaluation mechanism, we hope to provide research ideas with more technical support and cultural interpretation power for the intelligent creation of Chinese painting, and also provide reference for the digital transformation and innovation dissemination of traditional art resources.

2 Related work

In recent years, the research on digital protection, intelligent generation and interactive creation of traditional Chinese paintings has been continuously promoted. Li et al. sorted out the application path of visual computing in the arrangement, analysis and representation of traditional art resources from the perspective of Chinese cultural heritage computing, and pointed out that digital technology is promoting traditional art from static preservation to computational expression [1]. Feng et al. proposed iPoet, an interactive poetry and painting creation method, which enhances semantic linkage in the process of art generation through visual multimodal analysis, providing new ideas for traditional painting content understanding and cross-media expression [2]. Geng et al. constructed MCCFNet to classify traditional Chinese paintings by multi-channel color fusion, indicating that color structure and cognitive features have important value in painting identification and style analysis [3]. Chung et al. used the boundary enhanced generative adversarial network to realize the interactive conversion of Chinese ink painting to real images, which expanded the technical boundary of visual mapping research of Chinese painting [4].

In terms of style transfer and scene generation, Hong et al. transferred the aesthetic style of Chinese landscape painting to the virtual scene of classical gardens, proving that the deep network can retain the formal characteristics and aesthetic temperament of Oriental paintings in cross-scene tasks [5]. Guo et al. proposed ArtVerse parallel human-computer collaborative painting paradigm, emphasizing that art generation should not stay at the level of automatic output, but should pay attention to the participation of creators and feedback regulation mechanism [6]. Zhang et al. systematically summarized the computational methods of

traditional Chinese painting from the perspective of "six methods", and promoted the correspondence research between technique rules, formal languages and algorithmic models [7]. Cheng et al. further discussed the role of deep neural networks in traditional landscape painting creation, showing the application potential of deep models in composition organization and style construction [8].

At the same time, Li et al. studied the style transfer and creation algorithm of Chinese painting based on artificial intelligence, which provided a relatively complete implementation path for computer-aided generation of Chinese painting [9]. Yan et al. started with stroke recognition and simulation, and strengthened the supporting role of computer vision in the analysis of detail of Chinese painting [10]. In general, the existing research has formed a certain foundation in the classification, style transfer, interactive generation and stroke simulation of Chinese painting. However, there is still a lack of more targeted and systematic research on how to achieve stable expression of national cultural elements under the deep learning framework, especially for meticulous flower-and-bird painting, which has the characteristics of fine shape, heavy color and cultural symbol.

3 Model design of meticulous flower-and-bird painting and national cultural elements in Chinese painting creation based on deep learning

3.1 Visual semantic features and data representation of ethnic cultural elements in meticulous flower-and-bird painting

The computable features of meticulous brushwork flower and bird paintings not only come from visible objects such as flowers, birds, branches and leaves, and artifacts, but also come from implicit information such as line drawing density, color level, white space ratio, center of gravity of composition, and implied meaning of patterns. Therefore, this study divides the data representation into two parts: visual semantic feature coding and structural annotation of ethnic cultural elements, and establishes a unified mapping between image layer, symbol layer and semantic layer.

In the visual feature encoding stage, we first divide the work into several semantic regions, and extract the line, color and composition information respectively. The line part focuses on delineating clarity, density rhythm and edge continuity. The color part focuses on comprehensive color scale, cold and warm contrast and heavy color distribution. The composition part records the position of the main body, the proportion of the accompanying body and the intensity of the blank. For the ith region, let the line features, color features and composition features be l_i, c_i and q_i respectively, then the local visual semantic vector can be expressed as follows.

$$v_i = W_v[l_i; c_i; q_i] + b_v \tag{1}$$

where, W_v is the projection matrix, b_v is the bias term, and $[\cdot]$ represents the vector concatenation. This formula is used to compress the originally scattered heterogeneous visual information into a unified feature space, so as to facilitate the subsequent network recognition of the fine form and aesthetic structure of meticulous flower and bird paintings.

The representation of national cultural elements cannot stay on the static mark of "there or there", but also needs to reflect the strength of association between elements and screen objects. For example, peony, peacock, crane, pomegranate, twigs, group patterns, and ethnic minority

dress colors bear different cultural meanings in different works. To do this, we build an attribute vector ej for each type of cultural element and measure the degree of semantic coupling between it and the visual area by bilinear relationship:

$$r_{ij} = \frac{v_i^{\top} M e_j}{\|v_i\|_2 \|M e_j\|_2} \tag{2}$$

Here, M is the relational mapping matrix, and r_{ij} represents the matching coefficient between the ith visual area and the JTH national culture element. The larger this index is, the more likely the region is to carry the corresponding cultural implication, which is helpful for the subsequent model to distinguish between "surface decoration" and "effective expression".

In the sample-level representation stage, the visual information of the whole image and the cultural information need to be integrated into a unified input. Assuming that the whole work contains n semantic regions and m types of cultural elements, the final sample representation is defined as follows.

$$x = \left[\frac{1}{n} \sum_{i=1}^{n} v_i ; \frac{1}{m} \sum_{j=1}^{m} e_j ; \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} \right] \tag{3}$$

where, the first term reflects the global visual style, the second term reflects the collection characteristics of cultural elements, and the third term reflects the association structure of the two. The sample constructed in this way not only retains the fine features of meticulous brushwork flower and bird paintings at the image level, but also supplements the interpretation information of national culture elements at the semantic level.

As shown in Table 1, this paper divides the main data sources of meticulous flower-and-bird painting into six categories, and assigns corresponding coding methods to them respectively. Such processing can reduce the interference of single image feature on model judgment, and make the data input more in line with the real needs of Chinese painting creation tasks. Compared with the direct use of general image tags, this representation method emphasizes the internal relationship between objects, strokes, colors and cultural meanings, which provides a more stable data basis for subsequent multi-modal feature extraction and generative constraint design. On the whole, the data representation scheme constructed in this section does not simply expand the number of features, but tries to transform the formal language of fine brushwork flower and bird painting and the meaning structure of national cultural elements into learnable, comparable and optimized computational expressions.

Table 1: Coding scheme of visual features and ethnic cultural elements of meticulous brushwork flower and bird paintings

Feature Encoding Category	Specific Content	Data Source	Representation Form	Main Function
Line-Drawing Features	Outline thickness, edge continuity, turning density, contour clarity	Original image edges and brush-line regions	Numerical vectors	Represent the fineness of meticulous brushwork modeling and the order of line organization
Color Features	Dominant color distribution, composite tonal levels, warm–cool contrast, heavy-color area	Image color channels and regional statistics	Numerical vectors	Reflect coloring style and ethnic color tendencies
Composition Features	Subject position, blank space ratio, hierarchical relationship, visual center of gravity	Spatial distribution information of the picture	Numerical vectors	Describe compositional organization and picture balance
Subject-Object Features	Flower species, bird posture, branch–leaf combination, object accompaniment	Object recognition results and manual verification	Category labels and semantic encoding	Clarify the main semantic objects in the picture
Pattern-Symbol Features	Round floral patterns, scrolling branch patterns, cloud motifs, feather ornaments, etc.	Local pattern regions and manual annotation	Label encoding and feature representation	Strengthen the recognition ability of ethnic decorative elements
Cultural-Meaning Features	Auspicious symbolism, regional aesthetics, ethnic ritual orientation	Expert annotation and cultural semantic lexicon	Semantic labels and relational descriptions	Establish the correspondence between image content and cultural meaning

3.2 Multi-modal Deep feature extraction and fusion mechanism for Chinese painting Creation

The intelligent generation of meticulous flower-and-bird paintings is not a task that can be completed simply by relying on image texture learning. Compared with general visual objects, the creation of Chinese painting involves multi-source information such as the structure of brush lines, the relationship between colors, the details of patterns, and the semantics of national culture. Therefore, after the basic data representation is completed, it is necessary to construct a multimodal deep feature extraction and fusion mechanism for Chinese painting creation tasks. The image content, stroke information, color structure and cultural semantics are co-modeled through different branches to enhance the recognition ability of the model for the internal relationship between the style characteristics of fine brushwork flower and bird painting and

national cultural elements. Figure 1 shows the network structure of multimodal feature extraction and fusion.

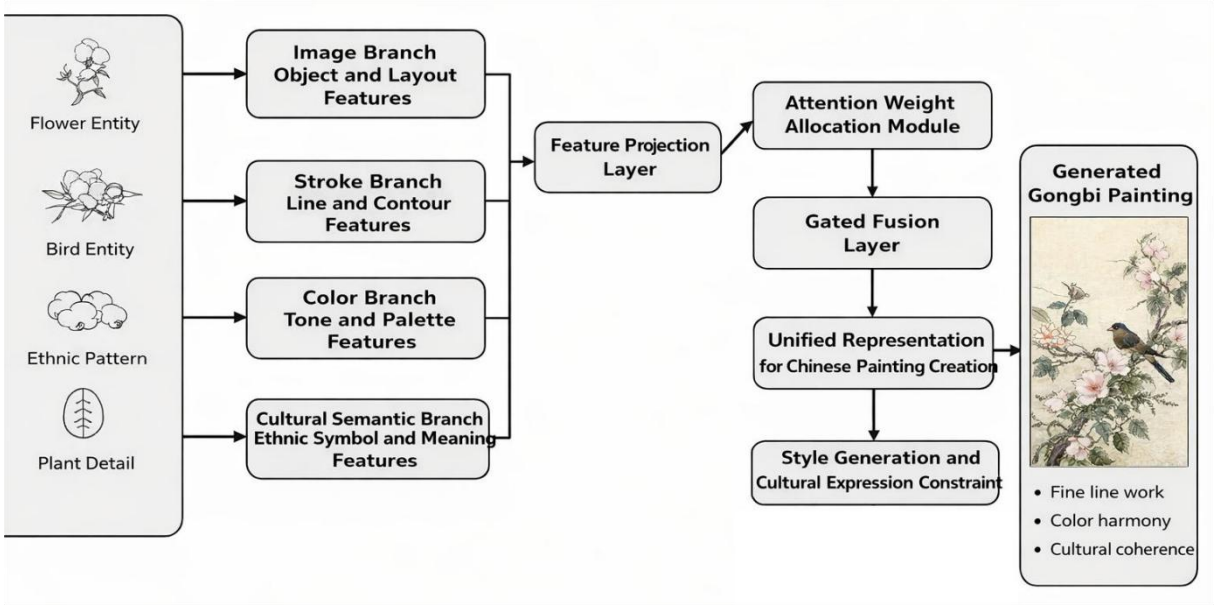

Figure 1: Network structure diagram of multi-modal feature extraction and fusion

As shown in Figure 1, the multi-modal feature extraction and fusion network we constructed consists of four parts: image branch, pen line branch, color branch and cultural semantic branch. The image branch is mainly responsible for extracting the main body of flowers and birds, the level of branches and leaves, and the overall composition information. The brush line branch is used to depict the fine, uniform and coherent line drawing features in Gongbi painting. The color branch focuses on the analysis of color hierarchy, comprehensive hue and national decorative color tendency. The cultural semantic branch obtains high-level semantic representation based on pattern labels, subject meanings and ethnic symbol descriptions. After local coding, each branch enters the unified fusion layer, and forms the global creation feature through weight allocation and gated screening, which finally provides stable input for subsequent style generation and expression constraints.

In the single-modal feature extraction stage, the original inputs are set as image information, pen line information, color information and cultural semantic information, and the corresponding depth coding results are defined as follows.

$$f_k = \phi_k(x_k), k \in \{1,2,3,4\} \tag{4}$$

Here, x_k represents the KTH modality input, $\phi_k(\cdot)$ represents the feature extraction network of the corresponding modality, and f_k is the encoded deep feature. This equation reflects the multimodal parallel modeling process, that is, the information from different sources completes the preliminary representation through relatively independent subnetworks, so as to avoid the compression loss of complex artistic information by a single path.

In order to highlight the difference in contributions of different modalities in different samples, this paper introduces an adaptive attention allocation mechanism before fusion to dynamically model the importance of features in each branch. Let the fusion weight be αk, then it is calculated as follows.

$$\alpha_k = \frac{\exp(u^\top \tanh(W_a f_k + b_a))}{\sum_{t=1}^{4} \exp(u^\top \tanh(W_a f_t + b_a)} \tag{5}$$

where W_a and b_a are trainable parameters and u is the weight evaluation vector. The function of this formula is to automatically judge which type of modality is more expressive according to the feature structure of the current work. For example, when the ethnic patterns and decorative color blocks are prominent in the picture, the weight of the color branch and the cultural semantic branch will be increased accordingly. When the work emphasizes more on the accuracy of drawing and the shape of flowers and birds, the contribution of the pen line branch and the image branch will be more obvious.

After the attention allocation is completed, this paper further adopts the gated fusion strategy to compress the different modal features into a unified representation. Let the integrated feature be z, then:

$$z = \sum_{k=1}^{4} \alpha_k \odot \sigma(W_g f_k + b_g) \tag{6}$$

Here, $\sigma(\cdot)$ is the nonlinear activation function, W_g and b_g are the gating mapping parameters, and $\odot$ represents element-wise modulation. Compared with the simple stitching method, the proposed mechanism can actively suppress redundant information in the fusion stage and retain the most valuable deep features for Chinese painting creation. The unified representation not only contains the detailed description characteristics of meticulous flower-bird paintings, but also incorporates the high-level semantic constraints of national culture elements, which makes it easier for the subsequent generative network to maintain the consistency between object modeling, ink order and cultural meaning during style learning.

In order to improve the stability of cross-modal representation, we also introduce a consistency adjustment objective in the training process, so that the distribution difference of different modalities after mapping to a unified space is reduced as much as possible. Let the feature centers between any two modes be μ_p and μ_q respectively, then the consistency term is defined as follows.

$$L_c = \sum_{p=1}^{4} \sum_{q=p+1}^{4} \| \mu_p - \mu_q \|_2^2 \tag{7}$$

This term is used to constrain the representation offset of each modal feature before and after fusion, and reduce the overall expression imbalance caused by a certain modality being too strong. For the creation task of meticulous flower-and-bird paintings, this processing helps to prevent the phenomena of "similar shapes but discordant colors", "bright colors but insufficient meaning" or "prominent patterns but fragmented picture relationships" in the generated results.

In summary, the multi-modal deep feature extraction and fusion mechanism proposed in this section establishes a complete process of branch extraction, dynamic weighting, gated fusion and consistency adjustment around the creation law of Chinese painting. Through this mechanism, the model can more carefully understand the complex structure of meticulous flower-and-bird paintings in form, color and cultural semantics, and provide a more solid feature basis for the accurate expression of national culture elements in the generated results.

3.3 Generation of meticulous flower-and-bird painting style and expression constraint design of ethnic cultural elements

After the multi-modal feature extraction and fusion, although the model has been able to obtain relatively complete creation information of meticulous flower and bird paintings, there may still be two problems in the output results if there is no constraint design in the generation stage: One is that the style tends to be generalized, and the generated image is close to general decorative painting in outline, color and rules, which is difficult to reflect the delicacy and implicacy that meticulous flower and bird painting should have. The other is that the elements of national culture stay in the local splicing level. Although patterns, colors or symbols appear in the picture, there is no organic connection between them and the main structure. Based on this, we also introduce a composite constraint mechanism for style consistency, cultural expression integrity and picture balance in the generation stage, so that the model not only pays attention to visual visibility, but also pays attention to the coordination relationship between cultural semantics and formal order.

In the generation structure, the fused unified representation, target style prior and cultural semantic conditions are input into the generator together to form a controlled generation process. Assuming that the fusion feature is z, the target style prior is s, and the cultural condition vector is m, the generated result can be expressed as follows.

$$\hat{y} = G(z, s, m) \tag{8}$$

where $G(\cdot)$ represents the generation network and $\hat{y}$ is the output image. This formula shows that the generation of meticulous flow-bird painting is not an unconditional mapping, but a directional generation completed under the joint action of style information and cultural conditions, which helps the model to retain the main shape while stably presenting the semantic characteristics of national culture elements.

In order to ensure that the generated results maintain the fine line outline and color level of meticulous flower and bird painting, this paper sets the style preservation constraint. Different from the methods that only compare pixel differences, the consistency between the generated image and the reference style image is measured from both texture statistics and hierarchical response levels, which is defined as:

$$L_s = \sum_{h=1}^{H} \| \Phi_h(\hat{y}) - \Phi_h(y_s) \|_1 + \eta \sum_{h=1}^{H} \| C_h(\hat{y}) - C_h(y_s) \|_F \tag{9}$$

Here, $\Phi_h(\cdot)$ represents the feature response of the h-th layer, $Ch(\cdot)$ represents the channel correlation structure of the corresponding layer, y_s is the reference Gongbi style image, and η is the balance coefficient. The constraint focuses on controlling the fineness of line drawing, color transition and local texture organization, so that the generated result is closer to the meticulous flower and bird painting in visual temperament.

Style consistency alone is still not enough to reflect the effective expression of national cultural elements, so this paper further establishes cultural semantic constraints. Let the response area of the JTH cultural element in the generated image be Ωj, and the corresponding activation map be A_j. Then the cultural expression loss is defined as follows.

$$L_m = \frac{1}{K} \sum_{j=1}^{K} \left(1 - \frac{\sum_{p \in \Omega_j} A_j(p)}{\sum_p A_j(p) + \varepsilon} \right) \tag{10}$$

Here, K is the number of cultural element categories, p represents the image location, and ε is a tiny constant to avoid the denominator being zero. This formula is used to restrict the cult-related response to concentrate on the due object area and composition position, avoid the disordered drift of national patterns, color symbols and moral elements, so as to improve the pertinence of cultural expression and the semantic consistency within the picture.

Meticulous flower-and-bird painting not only attaches importance to the subject matter and brushwork, but also emphasizes the stability of the rules and the proper density. To this end, this paper adds a screen balance constraint to jointly control the subject distribution, white space ratio and visual center of gravity. Suppose that the main body proportion, blank proportion and visual center shift of the generated image are u_g, w_g and d_g respectively, and the corresponding statistics of the reference sample are u_r, w_r and d_r, then:

$$L_b = |u_g - u_r| + |w_g - w_r| + \gamma \|d_g - d_r\|_2 \tag{11}$$

where, γ is the weight parameter. This method can suppress the common imbalance problem of composition in the generation process, and is especially suitable for the correction of excessive deviation of the main body, too little white space or the accumulation of the accompanying body in the fine brushwork flower and bird paintings.

Based on the above design, the overall optimization objective of the generation phase is written as follows.

$$L = \lambda_s L_s + \lambda_m L_m + \lambda_b L_b + \lambda_a L_a \tag{12}$$

Here, L_a represents the base generative adversarial term, and λ_s, λ_m, λ_b, and λ_a are the weights of each part. Through the joint constraints of style preservation, cultural expression, picture balance and generation authenticity, the objective function makes the model meet the formal requirements of meticulous flower and bird paintings and the expression requirements of national culture elements at the same time. The constraint design and function description in the generation stage of meticulous flower-and-bird painting are shown in Table 2.

Table 2: Constraint design and function description of meticulous flower-and-bird painting generation stage

Constraint Type	Controlled Object	Main Content	Expected Effect
Style Preservation Constraint	Line drawing, coloring, texture hierarchy	Control fine-line delineation, tonal transition, and local texture consistency	Preserve the rigorous and delicate visual style of meticulous flower-and-bird painting
Cultural Expression Constraint	Ethnic patterns, symbolic objects, semantic regions	Constrain cultural elements to appear stably within appropriate regions	Improve the accuracy and concentration of ethnic cultural element expression
Picture Balance Constraint	Subject position, blank space relationship, visual center of gravity	Adjust object distribution and spatial rhythm	Ensure compositional stability and enhance overall picture harmony
Basic Generation Constraint	Overall image quality	Improve image clarity, naturalness, and structural	Ensure that the generated result has good visual appeal and realism

		integrity	

3.4 Chinese painting creation generation process and interactive feedback mechanism design

After the completion of data representation, multi-modal feature extraction and style constraint design, the model has a strong ability to generate images. However, without the flow organization and feedback correction oriented to the creation process, the generated results may still stay at the level of one-time output, which is difficult to adapt to the actual law of repeated deliberation and gradual adjustment in Chinese painting creation. The formation of meticulous flower-and-bird painting needs to form a continuous iterative closed loop between theme setting, cultural implication selection, picture generation, result evaluation and manual correction. Therefore, we further construct a generation and interactive feedback mechanism for Chinese painting creation, so that the system can dynamically adjust the generation direction according to the evaluation results and user preferences, and improve the comprehensive performance of the work in style stability, cultural consistency and aesthetic completion. Figure 2 shows the closed-loop process of Chinese painting creation generation and interactive feedback.

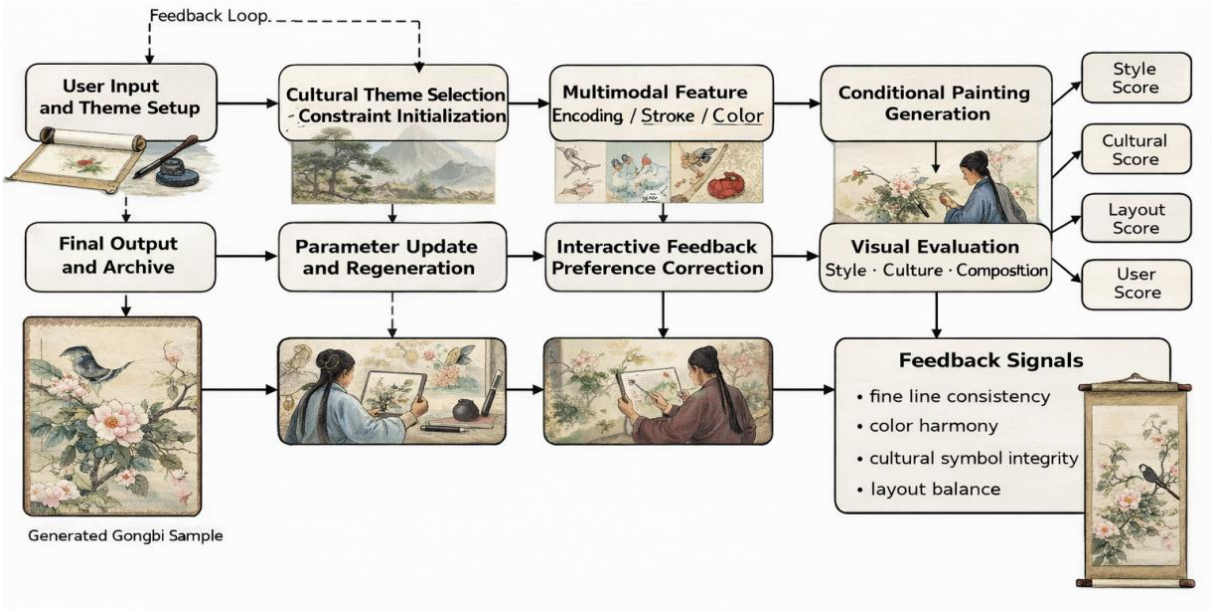

Figure 2: Closed-loop flow chart of Chinese painting creation generation and interactive feedback

The closed-loop process designed in this paper consists of eight steps: theme setting, cultural constraint initialization, multi-modal feature coding, condition generation, visual evaluation, interactive feedback, parameter update and regeneration. Users input the creation theme, subject object and cultural information, and the system filters the cultural theme according to the preset vocabulary and semantic mapping rules. The multimodal coding module jointly represented the image, pen line, color and semantic information, and fed the results into the generator to obtain the initial work. The initial results first enter the evaluation module, and their style consistency, cultural expression, composition balance and subjective satisfaction are jointly judged. Then the feedback signal is sent to the update module, and the condition vector, constraint weight and local parameters are corrected. The closed-loop mechanism formed in this way is closer to the process of "generation-inspection- adjustment-regeneration" in real creation.

In order to quantify the comprehensive quality of single-round generation results, let the style score, culture score, composition score and user feedback score of the TTH output image be s_t, c_t, p_t and u_t respectively, then the comprehensive evaluation function is defined as follows.

$$Q_t = \omega_1 s_t + \omega_2 c_t + \omega_3 p_t + \omega_4 u_t \tag{13}$$

Here,ω_1,ω_2,ω_3,ω_4 are the weight coefficients. In the feedback update phase, the generation condition needs to be adjusted according to the difference between the current evaluation results and the results of the previous round. Let the feedback gain of the current round be Δ_t, then:

$$\Delta_t = Q_t - Q_{t-1} \tag{14}$$

When $\Delta_t>0$, the current generation result is improved compared with the previous round, and the system can appropriately retain the existing parameter direction. When $\Delta_t<0$, it indicates that the current round of adjustment has weakened the quality of the work, and the weight configuration and local constraints need to be rolled back or re-corrected.

In order to optimize the condition vector round by round driven by feedback, this paper sets the creation control state in round t as h_t and the learning rate as ρ. Then the update rule is written as follows.

$$h_{t+1} = h_t + \rho \Delta_t g_t \tag{15}$$

Here, g_t represents the direction vector composed of user preference, cultural error and visual evaluation. This formula transforms the abstract feedback results into executable control modifications, so that the model can converge around the user's intention and the art specification.

Considering that repeated iteration will increase the computational overhead, this paper also sets the stopping criterion. When the output difference between two consecutive rounds is small and the comprehensive quality reaches the set threshold, the system ends the closed-loop iteration. Its condition can be expressed as follows.

$$|Q_t - Q_{t-1}| < \delta, \quad Q_t \geq \tau \tag{16}$$

Here, δ is the convergence threshold and τ is the quality threshold. This criterion not only ensures that the generated results are stable enough, but also avoids the increase of time consumption caused by invalid iterations.

In general, the creation generation and interactive feedback mechanism proposed in this section integrates the theme setting, cultural constraints, evaluation results and user preferences into the same control chain through the closed-loop process. This not only enhances the adaptability of the generation process of meticulous flower-and-bird painting, but also improves the pertinence of the expression of national cultural elements and the artistic completion of the final output, which provides a method basis for the evaluation of interactive effects and the analysis of application value in the subsequent experimental part.

4 Experimental design and result analysis

4.1 Data set construction and experimental environment configuration

In order to ensure the verifiability of model training and result analysis, we construct a multi-modal experimental dataset for meticulous flower-and-bird painting creation and the expression

of national cultural elements. The data mainly came from the publicly available digital image resources of Chinese painting, image data collected in the library and teaching research samples. After resolution screening, duplicate samples elimination and theme consistency check, 3600 valid images were finally retained. Among them, 2860 master samples of meticulous flower and bird paintings contain complete composition information such as flowers, birds, branches and leaves and artifacts. 740 reinforcement samples containing labels of national patterns, regional colors or cultural meanings were used to improve the model's ability to identify and constrain national cultural elements. According to the content annotation results, a total of 18 types of flower themes, 12 types of bird posture types, 9 types of national patterns and 8 types of cultural semantic tags are set in the dataset, such as auspicious meanings, etiquette symbols, regional aesthetics and decorative styles. In order to enhance the robustness of the model, the original image is further processed by rotation, cropping, mirroring and color scale perturbation, and the total number of samples that can be used for training reaches 10800 groups after amplification. The dataset is divided into training, validation and test sets with a 7:2:1 ratio of 7560 groups, 2160 groups and 1080 groups, respectively.

In terms of experimental environment, this paper uses Ubuntu 22.04 operating system, Python 3.10 as the development environment, and PyTorch 2.1 as the deep learning framework, supporting CUDA 12.1 and cuDNN acceleration library. The hardware platform is configured with an Intel Xeon processor, 32 GB memory, and an NVIDIA RTX 4090 graphics card with 24 GB of video memory. During model training, the batch size was set to 16, the initial learning rate was set to 0.0001, AdamW was used as the optimizer, the weight decay coefficient was set to 0.01, and the maximum number of training rounds was 180. In the training process, cosine annealing strategy and early stopping mechanism are introduced to improve the efficiency of parameter convergence and reduce the risk of overfitting. The above data scale and experimental configuration provide a stable experimental basis for the subsequent verification of multi-modal feature fusion, style generation constraints and interactive feedback mechanisms.

4.2 Data preprocessing and style label labeling method

In order to improve the recognition stability of the model for the detail information of meticulous flower and bird paintings and national cultural elements, this paper performs a unified preprocessing of the image before sample input. It includes resolution normalization, edge enhancement, background noise cleaning and color space correction, and the original image is uniformly scaled to 512×512 pixels. Considering the subtle differences in line drawing level and color transition of meticulous flower-and-bird painting, this paper uses channel standardization to reduce the fluctuations in brightness and saturation of images from different sources. The processing form is as follows:

$$\tilde{x}_{u,v,c} = \frac{x_{u,v,c} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} \tag{17}$$

Here, $x_{u,v,c}$ represent the original value of the pixel on the CTH channel, μ_c and σ_c represent the mean and standard deviation of the channel, and ϵ is the smoothing term. After processing, the images are better comparable in texture response, edge sharpness, and comprehensive hue.

In terms of style label labeling, this paper adopts a three-level labeling method of "manual initial labeling + rule verification + expert review", and establishes a style label system from four dimensions of line drawing features, color design methods, composition patterns and cultural symbols. The content of the label includes the fineness of meticulous brushwork, the

degree of heavy color, the combination type of flowers and birds, the category of patterns, and the direction of cultural meaning. For the samples with cross style features, the primary label and secondary label are recorded in parallel to enhance the adaptability of the training stage to complex style samples.

4.3 Evaluation index system and comparative experimental scheme

In order to comprehensively evaluate the actual effect of the model in the generation of meticulous flower-and-bird paintings and the expression of national cultural elements, we construct an evaluation index system from four levels of style maintenance, cultural expression, picture structure and comprehensive generation quality. Specifically, the style similarity is used to measure the consistency between the generated results and the target meticulous brush samples in terms of line drawing features, color levels and overall temperament. The matching effect of ethnic patterns, symbols and semantic labels in the generated images was evaluated by the expression accuracy of cultural elements. The composition coordination reflects the rationality of the layout of the main body, the relationship between the white space and the distribution of the visual center of gravity. The user satisfaction score is used to test the comprehensive performance of the generated results in terms of aesthetic acceptance and creative assistance value. At the same time, peak signal-to-noise ratio and structural similarity are used as auxiliary image quality indicators to supplement the basic performance of the generated results in terms of clarity and structural integrity. However, this paper still focuses on the style, culture and composition indicators that better reflect the characteristics of Chinese painting creation tasks.

In the design of comparison experiment scheme, four groups of methods are set up for horizontal comparison: the traditional style transfer method, the single-modal convolution generation method, the basic generation model without cultural constraints, and the multi-modal fusion generation model proposed in this paper. All methods are run under the same data set division, training rounds and hardware environment to ensure that the experimental results are comparable. On this basis, we further design ablation experiments to remove cultural semantic branches, expression constraint modules and interactive feedback mechanisms, respectively, and investigate the specific effects of each component on style maintenance, cultural expression and composition stability.

4.4 Ablation experiment

In order to verify the actual contribution of the three parts of multimodal fusion, expression constraints and interactive feedback to the performance of the model, this paper carries out ablation experiments in the same training environment, and selects style similarity, cultural element expression accuracy, composition coordination and user satisfaction as the core indicators for comparison. The results of ablation experiments are shown in Table III. When the cultural semantic branches are removed, the model can still complete the basic generation of flowers and birds, but the recognition ability of national patterns and cultural meanings is significantly reduced. After removing the expression constraint module, the image color and composition stability are greatly affected. After removing the interactive feedback mechanism, the final output fluctuates to some extent at the subjective evaluation level. In contrast, the full model remains optimal in all indicators, indicating that the multi-modal fusion and closed-loop correction mechanism constructed in this paper has a synergistic effect on the maintenance of meticulous flower-and-bird painting style and the expression of national cultural elements.

Table 3: Comparison of ablation experiment results

Model Setting	Style Similarity	Cultural Element Expression Accuracy / %	Composition Coordination / %	User Satisfaction / Score
Without Cultural Semantic Branch	0.842	81.6	84.3	8.1
Without Expression Constraint Module	0.857	84.9	82.7	8.3
Without Interactive Feedback Mechanism	0.866	86.5	85.8	8.5
Full Model	0.901	91.8	89.6	9.1

On the whole, the style similarity of the complete model reaches 0.901, which is 0.059 higher than that of removing the cultural semantic branch. The expression accuracy of cultural elements reaches 91.8%, which is 6.9 percentage points higher than that of the module removing expression constraints. The coordination degree of composition reaches 89.6%, which is 3.8 percentage points higher than that without interactive feedback mechanism. User satisfaction also increased from 8.1-8.5 to 9.1. These results show that each module forms a relatively obvious synergistic gain in generation quality, cultural expression and aesthetic stability.

5 Discussion

5.1 Comparison and analysis with existing Chinese painting generation methods and style transfer methods

In order to further verify the comprehensive advantages of the proposed method in the generation of fine brushwork flower-and-bird paintings and the expression of national cultural elements, we select the traditional style transfer method, the single-modal generation method, and the basic model without cultural constraints for horizontal comparison with the proposed method, focusing on the three indicators of style similarity, the accuracy of cultural element expression, and the coordination degree of composition. The comprehensive performance comparison of different methods is shown in Figure 3. From the overall results, the traditional style transfer method has a certain effect in the visual transfer level, but the ability to maintain the fine lines in the meticulous flower and bird paintings and national cultural symbols is relatively limited. The single-modal generation method can improve the local texture performance, but there is still the problem of insufficient cultural expression depth. Although the model without cultural constraints is stable in general generation quality, it is still not prominent enough in the directional expression of national culture elements. In contrast, the proposed method shows better overall coordination under the joint effect of multi-modal fusion and expression constraints.

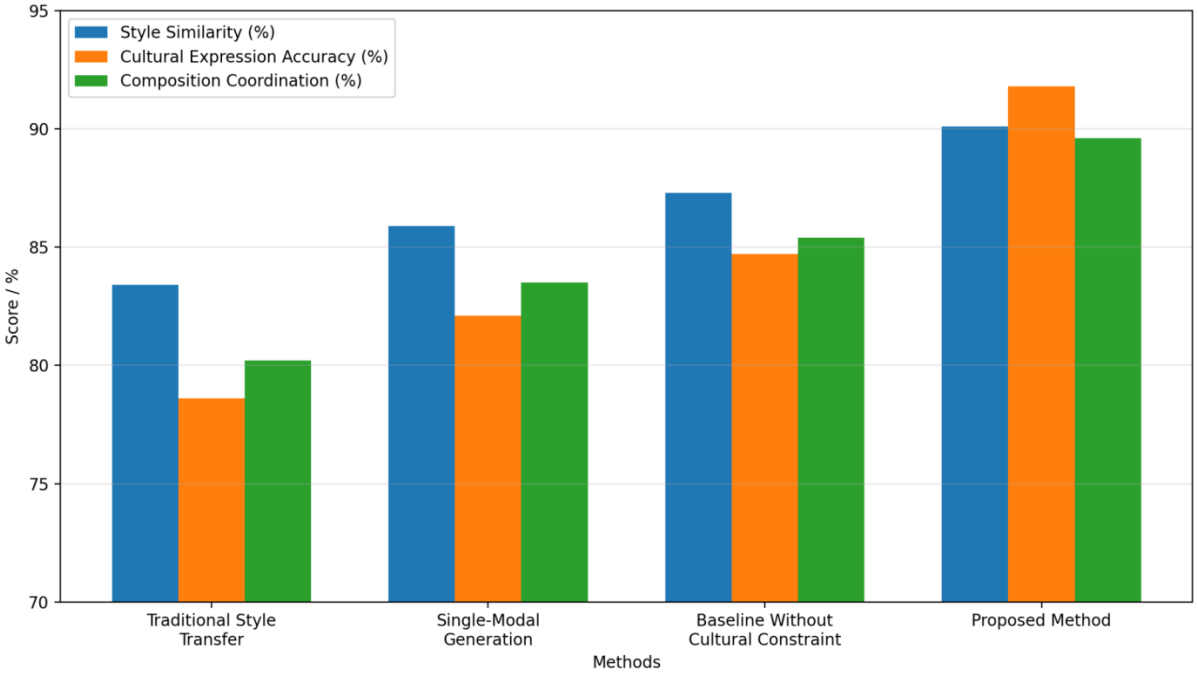

Figure 3: Bar chart of comprehensive performance comparison of different Chinese painting generation methods

Overall, the style similarity of the proposed method reaches 90.1%, which is 6.7 percentage points higher than that of the traditional style transfer method. The expression accuracy of cultural elements reaches 91.8%, which is 9.7 percentage points higher than that of the single-modal generation method. The coordination degree of composition reaches 89.6%, which is 4.2 percentage points higher than that of the model without cultural constraints. This shows that the proposed method can not only better maintain the formal characteristics of meticulous flower-and-bird painting, but also achieve more stable generation effects in the integration of national cultural elements and the overall organization of the picture.

5.2 Verification of accuracy of ethnic cultural elements expression and model stability

In order to test the ability of the proposed method to maintain national cultural elements under complex conditions, we also set up perturbation experiments with different intensities, including partial occlusion, color shift and pattern information weakening, and compare the proposed method with the single-modal generation method and the model without cultural constraints. Figure 4 shows the variation of the expression accuracy of national cultural elements under different disturbance conditions. It can be seen from the results that as the disturbance intensity gradually increases, the expression accuracy of each model decreases, but the overall decline of the proposed method is relatively small, indicating that it has better stability in cultural semantic modeling and multi-modal fusion. Especially under medium and high disturbance conditions, the proposed method can still better maintain the corresponding relationship between national patterns, symbols and cultural meanings, and avoid the problems of missing cultural elements or expression deviation in the generated results.

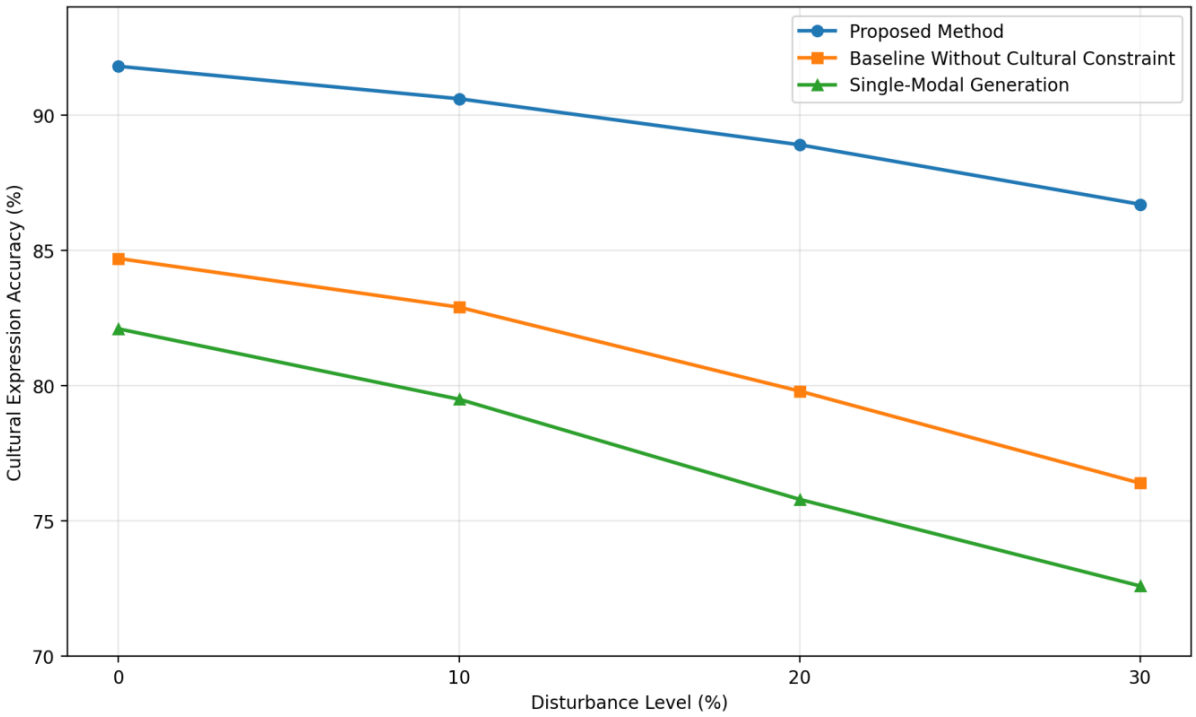

Figure 4: Line chart of expression stability of ethnic cultural elements

In general, under the condition of no disturbance, the accuracy of cultural element expression of the proposed method is 91.8%. When the perturbation intensity is increased to 30%, the accuracy still remains at 86.7%, and the overall decrease is only 5.1 percentage points. In contrast, the model without cultural constraints decreased from 84.7% to 76.4%, a decrease of 8.3 percentage points; The single-modal generation method decreased from 82.1% to 72.6%, a decrease of 9.5 percentage points. This indicates that the proposed method has more advantages in the stability and anti-interference ability of the expression of national cultural elements.

5.3 Computing resource consumption and feasibility evaluation of model application

In addition to the generation effect, the computational resource consumption of the model is also an important basis to measure the practical value of the model. For the creation task of meticulous flower-and-bird painting, although the model has high generation quality, it requires too long training cycle and high hardware threshold, and its promotion value in teaching assistance, digital exhibition and interactive creation will be limited. Therefore, this paper comprehensively compares the traditional style transfer method, the single-modal generation method, the cultural-free model and the proposed method from four aspects of training time, single inference time, video memory occupation and parameter scale. The computational resource consumption and application feasibility evaluation of different methods are shown in Table 4.

Table 4: Computational resource consumption and application feasibility evaluation of different methods

Method	Training Time / h	Single-Inference Time / s	GPU Memory Usage / GB	Parameter Size / M	Application Feasibility
Traditional Style Transfer Method	5.8	1.2	7.4	24.6	Suitable for basic demonstration and offline generation
Single-Modal Generation Method	9.6	1.8	10.9	38.2	Suitable for routine digital creation
Model Without Cultural Constraints	11.3	2.1	12.6	44.7	Suitable for general generation tasks
Proposed Method	13.1	2.4	14.2	51.3	Suitable for interactive creation and cultural expression tasks

In general, the single inference time of the proposed method is 2.4 s, which is only 0.3 s higher than that of the model without cultural constraints, but the improvement in generation quality and cultural expression accuracy is more obvious. It occupies 14.2GB of video memory and can run stably in RTX 4090 environment. The parameter scale is 51.3M, which is still in the deployable range. It shows that although the method increases a certain amount of computational overhead, it does not significantly weaken the application feasibility, and is more suitable for Chinese painting creation scenes with high requirements for style accuracy and cultural expression.

5.4 Application value analysis of model generation results in meticulous flower-and-bird painting creation

The application value of the model generation results in the creation of meticulous flower-and-bird paintings is not only reflected in the watchability of the image output level, but also reflected in its supporting role in the reorganization of traditional themes, cultural semantic expression and creation efficiency improvement. Different from the previous section, which focuses on the horizontal comparison of algorithm performance, this section pays more attention to the application performance of the model in the actual creation scene. Therefore, the traditional style transfer method, the uncultural constraint model and the proposed method are selected as representative objects for analysis. The method in this paper can better maintain the line drawing order, color level and composition rules of the fine brushwork flower and bird painting, and make the national cultural elements integrate into the picture structure in a more natural way. It shows that if the deep learning model can form a synergy in the three levels of feature extraction, expression constraint and feedback correction, the deep learning model can effectively improve the performance of the model. To a certain extent, the function from image generation to creative assistance can be transformed. This method has good application potential for digital teaching, cultural creative design and redevelopment of traditional painting resources.

The application value evaluation of the model generation results is shown in Table 5. The results show that compared with the traditional style transfer method and the model without cultural constraints, the proposed method performs better in the aspects of meticulous style maintenance, national culture recognition, composition integrity, aesthetic acceptance and

creative assistant adaptability. Among them, the meticulous style preservation score reaches 9.3 points, the national culture recognition score reaches 9.4 points, and the creative assistant adaptability reaches 9.5 points, which indicates that the method proposed in this paper is not only closer to the artistic requirements of meticulous flower and bird painting in the visual level, but also more suitable for the creation task of Chinese painting with clear cultural expression goals.

Table 5: Model generation result application value evaluation table

Evaluation Dimension	Traditional Style Transfer Method	Model Without Cultural Constraints	Proposed Method
Meticulous Style Preservation / Score	8.2	8.6	9.3
Ethnic Cultural Recognition / Score	7.8	8.1	9.4
Composition Integrity / Score	8.0	8.4	9.1
Aesthetic Acceptance / Score	8.3	8.5	9.2
Creative Assistance Adaptability / Score	8.1	8.4	9.5

In general, the meticulous style preservation score of the proposed method reaches 9.3 points, which is 1.1 points higher than that of the traditional style transfer method. The national culture recognition score reached 9.4 points, which was 1.3 points higher than that of the model without cultural constraints. Authoring assistance adaptability reached 9.5 points, which was the highest among the three categories of methods. This shows that the proposed method can not only improve the artistic performance of the generated results, but also reflect stronger practical value in application scenarios such as digital creation of meticulous flower-and-bird painting, teaching demonstration and cultural communication.

5.5 Discussion on method innovation and research limitations

From the perspective of research ideas, the innovation of this paper is mainly reflected in three levels. First, our research object does not stop at the general generation task of Chinese painting, but further focuses on the specific category of meticulous flower-and-bird painting, which has the complex characteristics of fine line drawing, heavy color decoration and cultural meaning, so that the model design is closer to the real art style. Second, in terms of method design, we do not rely on a single image generation framework, but integrate visual semantic feature representation, multi-modal deep feature extraction, ethnic cultural elements expression constraints, and interactive feedback mechanisms into the same technical chain, forming a relatively complete auxiliary process of Chinese painting creation. This approach enhances our ability to characterize the relationship among meticulous language, cultural symbols and composition order. Thirdly, in the evaluation level, we take into account the style similarity, cultural expression accuracy, composition coordination, resource consumption and application value analysis, so that the pros and pros of the model are no longer limited to a single generation quality judgment, but can be examined from two dimensions of artistic expression and practical application.

At the same time, we recognize that there are still some limitations in this paper. Although the data set has been specifically constructed around fine brushwork flower-and-bird paintings and ethnic cultural elements, the sample sources are still mainly digital image data, and some works have differences in shooting conditions, color reproduction and resolution, which will have a certain impact on the stability of model learning. Although the labeling of national

cultural elements has been screened by rules and manually reviewed, the cultural meaning itself has a certain openness and context, and it is still difficult to use fixed labels to completely exhaust. In addition, although the interactive feedback mechanism we constructed has the ability of closed-loop correction, its feedback sources still mainly rely on preset evaluation indicators and users' explicit preferences, and it is not sufficient to respond to more complex aesthetic judgments and changes in creative intentions. In the follow-up research, we can continue to expand the size of high-quality samples, refine the cultural semantic level, and enhance the dynamic interaction modeling, so as to further improve the expressiveness and interpretability of intelligent creation of Chinese painting.

6 Conclusion

Focusing on the expression of national cultural elements in the creation of fine brushwork flower-and-bird paintings, this paper proposes a generation method of Chinese paintings that integrates visual semantic modeling, multi-modal deep feature extraction, style-preserving constraints, cultural semantic constraints and interactive feedback mechanism. The experimental results show that the proposed method is superior to the traditional style transfer method, the single-modal generation method and the model without cultural constraints in terms of style similarity, cultural expression accuracy, composition coordination and application evaluation. The style similarity reaches 90.1%, the cultural element expression accuracy reaches 91.8%, and the creative assistant adaptability reaches 9.5. The results show that multimodal fusion and closed-loop feedback can effectively enhance the fineness, stability and cultural expression depth of meticulous flower-and-bird painting generation. In general, this paper provides a feasible technical path for the intelligent creation of Chinese painting, and also provides a new research reference for the structured expression and innovative transformation of national cultural elements in digital art scenes.

Funding

This work was supported by Research on the Construction of a Painting Resource Database of Guangxi Frontier Regional Characteristics and Innovative Practice in Chinese Painting, 2025SBNGCC018

References

[1] Li M, Wang Y, Xu Y Q. Computing for Chinese Cultural Heritage[J]. Visual Informatics, 2022, 6(1): 1-13. DOI: 10.1016/j.visinf.2021.12.006.

[2] Feng Y C J, Chen J Z, Huang K Y, Wong J K, Ye H, Zhang W, Zhu R C, Luo X N, Chen W. iPoet: interactive painting poetry creation with visual multimodal analysis[J]. Journal of Visualization, 2022, 25(3): 671-685. DOI: 10.1007/s12650-021-00780-0.

[3] Geng J, Zhang X, Yan Y, Sun M, Zhang H, Assaad M, Ren J, Li X. MCCFNet: Multi-channel Color Fusion Network for Cognitive Classification of Traditional Chinese Paintings[J]. Cognitive Computation, 2023, 15: 2050-2061. DOI: 10.1007/ s12559-023-10172-1.

[4] Chung C Y, Huang S H. Interactively transforming Chinese ink paintings into realistic

images using a border enhance generative adversarial network[J]. Multimedia Tools and Applications, 2023, 82(8): 11663-11696. DOI: 10.1007/s11042-022-13684-4.

[5] Hong S, Shen J, Lü G, Liu X, Mao Y, Sun N, Tang L. Aesthetic style transferring method based on deep neural network between Chinese landscape painting and classical private garden's virtual scenario[J]. International Journal of Digital Earth, 2023, 16(1): 1491-1509. DOI: 10.1080/17538947.2023.2202422.

[6] Guo C, Dou Y, Bai T, Dai X, Wang C, Wen Y. ArtVerse: A Paradigm for Parallel Human-Machine Collaborative Painting Creation in Metaverses[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(4): 2200-2208. DOI: 10.1109/TSMC.2022.3230406.

[7] Zhang W, Zhang J W, Wong K K, Wang Y, Feng Y C J, Wang L, Chen W. Computational Approaches for Traditional Chinese Painting: From the "Six Principles of Painting" Perspective[J]. Journal of Computer Science and Technology, 2024, 39(2): 269-285. DOI: 10.1007/s11390-024-3408-x.

[8] Cheng L, Wang H, Wang T. The role of deep neural network in the creation of traditional Chinese landscape painting[J]. Journal of Computational Methods in Sciences and Engineering, 2024, 24(4-5): 2815-2830. DOI: 10.3233/JCM-247516.

[9] Li J, Liu H. Computer Aided Style Transfer and Creation Algorithm of Chinese Painting Based on Artificial Intelligence[J]. Computer-Aided Design and Applications, 2024, 21(S14): 187-201. DOI: 10.14733/cadaps.2024.S14.187-201.

[10] Yan X, Chen J, Li W, Zhang Z. Application of Computer Vision-based Chinese Painting Stroke Recognition and Simulation System[J]. Computer-Aided Design and Applications, 2024, 21(S15): 35-53. DOI: 10.14733/cadaps.2024.S15.35-53.

[11] Tang Y. Style Transfer of Chinese Art Works Based on Dual Channel Deep Learning Model[J]. Computational Intelligence and Neuroscience, 2022: 1-11. DOI:10.1155/2022/4376006.

[12] Chen B. Classification of Artistic Styles of Chinese Art Paintings Based on the CNN Model[J]. Computational Intelligence and Neuroscience, 2022: 1-7. DOI:10.1155/2022/4520913.

[13] Gui X, Zhang B, Li L, Yang Y. DLP-GAN: learning to draw modern Chinese landscape photos with generative adversarial network[J]. Neural Computing and Applications, 2024: 5267-5284. DOI:10.1007/s00521-023-09345-8.

[14] Ding Y, Wang H, Liu N, Li T. TCP-RBA: Semi-supervised learning for traditional chinese painting classification with random brushwork augment[J]. Journal of Intelligent & Fuzzy Systems, 2024, 46(4): 10653-10663. DOI:10.3233/JIFS-236533.

[15] Wang W, Li Y, Ye H, Ye F, Xu X. Ink painting style transfer using asymmetric cycle-consistent GAN[J]. Engineering Applications of Artificial Intelligence, 2023, 126: 107067. DOI:10.1016/j.engappai.2023.107067.

[16] Hu Q, Peng X, Li T, Zhang X, Wang J, Peng J. ConvSRGAN: super-resolution inpainting

of traditional Chinese paintings[J]. Heritage Science, 2024, 12(1): 176. DOI:10.1186/s40494-024-01279-1.

[17] Lyu Q, Zhao N, Yang Y, Gong Y, Gao J. A diffusion probabilistic model for traditional Chinese landscape painting super-resolution[J]. Heritage Science, 2024, 12(1): 1-12. DOI:10.1186/s40494-023-01123-y.

[18] Xiao B, Li H. Painting style classification and art image analysis model by fusion of convolutional neural network and principal component analysis[J]. Journal of Computational Methods in Sciences and Engineering, 2025, 25(5): 4048-4060. DOI:10.1177/14727978251337910.

[19] Li H, Wang L, Liu J. A review of deep learning-based image style transfer research[J]. The Imaging Science Journal, 2025, 73(4): 504-526. DOI:10.1080/13682199.2024.2418216.

[20] Sun Y, Xie X, Li Z, Zhao H. Image style transfer with saliency constrained and SIFT feature fusion[J]. The Visual Computer, 2025, 41(7): 4915-4930. DOI:10.1007/s00371-024-03698-4.

[21] Liang Y, Lin F, Xie W, Wang J, Nie T. Research and design of image style transfer technology based on multi-scale convolutional neural network feature fusion[J]. Electronics Letters, 2024, 60(11): e13250-e13254. DOI:10.1049/ell2.13250.

[22] Wang X, Du Y, Pang Y. Inkartgan: a deep learning-based approach for digital generation of ink painting[J]. Journal of Intelligent Manufacturing, 2025: 1-17. DOI:10.1007/s10845-025-02674-6.