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

Redesigning Stone Lion Images with Neural Style Transfer: A Computational Aesthetic Framework for Modernizing Cultural Symbols

Maocong Lin[1,2], *Wenfei Li*[2] *and Gaofeng Mi*[1,*]

[1] College of Design and Art, Shaanxi University of Science and Technology, Xi'an 710000, Shaanxi, China

[2] College of Arts, Qingdao University,Qingdao 266000, Shandong, China

SUMMARY: *In recent years, the application of deep learning in image generation, visual recognition and creative design has deepened, and neural style transfer has gradually become an important technical path connecting computer vision and art design. Aiming at the problems of aging visual expression, limited style update and easy weakening of cultural recognition of traditional stone lion images in contemporary communication scenes, this paper constructs a redesign method of stone lion images based on neural style transfer, and proposes a computational aesthetics framework for cultural symbol modernization. The multi-source data organization is carried out around the traditional stone lion images, modern art style samples and visual design samples. Through image preprocessing, deep feature representation, bidirectional style mapping and aesthetic constraint modeling, the coordination and fusion between the content structure of the stone lion and the modern visual style are realized. The experimental results show that the proposed method can effectively enhance the contemporary expression features of color, texture and composition while maintaining the head contour, mane hierarchy and overall pose recognition of the Shi lion. In the later stage of model training, the cycle consistency loss is reduced to 1.3, the average PSNR reaches 25.33 dB, and the average SSIM reaches 0.844. The results show that the proposed method has good application value in the field of traditional cultural image digital updating and visual redesign.*

KEYWORDS: *neural style transfer; Redesign of stone lion image; Computational aesthetics; Modernization of cultural symbols*

1 Introduction

Image is not only the carrier of visual information, but also an important medium for cultural meaning to be deposited and transmitted. For traditional statues such as the stone lion, its value is not only reflected in the shape itself, but also reflected in the symbol structure shaped by historical rites, regional aesthetics and folk beliefs. With the continuous development of digital image processing and computer vision technology, the redesign of cultural symbols is no longer limited to manual tracing, experience rewriting or local decoration adjustment, but has begun to turn to a computational path with deep feature analysis, style mapping and generative model control as the core. Especially after the introduction of neural style transfer, image content and style can be separated, reorganized and optimized in the convolutional feature space, which makes the modern translation of traditional visual motifs obtain more stable technical support [1-3]. Since then, the adaptive instance normalization method for arbitrary style transfer has

*linmaoc188@163.com

https://doi.org/10.65102/is2026208

Published: 30 April, 2026

further improved the applicability of the model in multi-style scenarios [4, 5], and the Transformer-based style transfer framework has enhanced the ability of long-range dependence modeling, so that the coordination relationship between complex patterns, color order and local semantics can be more fully expressed [6, 7].

Existing studies show that neural style transfer has been widely used in fields such as artistic image generation, visual creative design and digital transformation of cultural heritage, but its focus is mostly on style simulation itself, and it is still insufficient to consider the balance between semantic continuation, recognition stability and aesthetic modernization of cultural symbols [8-10]. If we only pursue the replacement of surface style, it is often easy to cause the weakening of traditional image structure, the drift of symbolic meaning or the distortion of visual recognition. The stone lion image has this typical feature: its head proportion, mane organization, torso posture and decorative texture together constitute a strong cultural recognition. Once the content skeleton and local semantics are ignored in the stylization process, the generated result may be more "new", but it may not still be an accurately recognizable stone lion.

At the same time, the development of computational aesthetic research also provides a new evaluation dimension for cultural image redesign. From early high-level visual feature modeling to deep image aesthetic evaluation networks [11-15], researchers have gradually proved that composition, color hierarchy, visual balance, and semantic clarity can be transformed into computable aesthetic indicators. The research on perceptual similarity measurement further shows that deep features have strong effectiveness in reflecting human visual judgment [16-18]. This means that the modern expression of the stone lion image should not only rely on subjective experience judgment, but also establish a quantifiable evaluation mechanism between content fidelity, style adaptation and aesthetic quality. In recent years, research on visual style transfer for cultural heritage has begun to focus on the coupling relationship between technical methods and cultural expressions [19, 20], but a systematic computational framework for the traditional visual symbol Shi lion is still relatively lacking.

Based on this, this paper focuses on the redesign of the stone lion image, and attempts to construct a cultural symbol modernization framework combining neural style transfer and computational aesthetic evaluation. The research focus is not to simply generate a number of stylized images, but to make the traditional stone lion obtain a formal expression more in line with the contemporary visual communication environment on the basis of maintaining the core identification characteristics through content-style feature decomposition, aesthetic feature extraction and transfer result constraints. In this way, on the one hand, it can provide a reusable method for the digital updating of traditional image resources, and on the other hand, it can also help to promote the research paradigm of cultural symbol design from experience-driven to computation-aided and evaluation-provable.

1.1 Problem Formulation

In the process of visual updating of traditional culture, the redesign of stone lion image has long relied on manual experience, manual description and local modification. Although this method can retain certain cultural semantics, it generally has problems such as unstable efficiency, extensive style control and insufficient consistency of results. Especially when the design goal turns to digital communication, cultural and creative development and cross-media application, traditional methods are often difficult to simultaneously consider symbol recognition, shape integrity and contemporary aesthetic expression. The existing image style transfer technology provides a new calculation path for cultural image reconstruction. However, in complex symbol scenes, there are still some phenomena such as structural boundary loosening, local pattern distortion, and semantic center shift, which lead to the form change of the generated results, but weaken the original cultural identification of the stone lion. Based on this reality gap, this paper summarizes the core issues in the modern representation of the stone lion image into two levels. The first is how to maintain the stability of the key content structures such as the head, mane, and body posture of the stone lion in the deep feature space. The second is how to introduce color, texture and composition features that are more in line with the needs of contemporary visual communication through a computable style mapping mechanism. Aiming at the above problems, this paper proposes a redesign framework combining neural style transfer and computational aesthetic constraints to improve the generation quality, expression efficiency and application controllability of traditional cultural symbols in the digital environment.

1.2 Research contribution of this paper

This paper aims to construct a neural style transfer computing framework for the redesign of stone lion images, so as to improve the generation quality, style adaptation ability and aesthetic evaluation interpretability of traditional cultural symbols in the modern visual communication context.

The research contributions of this paper are mainly reflected in the following aspects:

Focusing on the stone lion, a traditional image with stable cultural recognition characteristics, this paper collects traditional stone lion images, artistic style samples and modern design images to form a multi-source image data basis for cultural symbol redesign.

In the preprocessing stage, the resolution of images from different sources is unified, color space conversion and pixel normalization are performed to ensure the consistency of the input scale of the model and enhance the stability of the training process.

Combined with the deep convolutional feature representation, the content structure, local texture and style information of the stone lion image are decomposed and modeled, which strengthens the retention ability of key visual features such as head contour, mane organization and body posture in the transfer process.

A neural style transfer model for the modern expression of cultural symbols is constructed, so that the style mapping and visual redesign of the stone lion image can be completed without relying on strict paired samples.

The computational aesthetic evaluation is introduced to comprehensively analyze the generated results from the dimensions of structural clarity, style coordination, visual appeal and content recognition, so as to improve the comparability and interpretability of the redesign results.

The experimental results show that the proposed method can realize the style transformation with more contemporary visual characteristics while maintaining the core culture recognition characteristics of the stone lion, and provide a reusable technical path for the digital update of traditional cultural images.

The rest of this paper is arranged as follows: Section 2 reviews the related research, Section 3 introduces the method design process, Section 4 presents the experimental results and discussion, and Section 5 concludes the paper and proposes future research directions.

2 Related Research

The proposal of neural style transfer has shifted the research of image generation from pure pixel transformation to joint modeling of content features and style statistics. Based on the deep features of convolutional neural network, Gatys et al. [1] used content loss and style loss to realize the separate expression of image semantic structure and artistic style texture, which proved that deep features can effectively undertake the task of visual transfer. This method is of pioneering significance in theory, but its optimization process relies on iterative solution, and its computational overhead is large, which is difficult to directly adapt to high-frequency design and production scenarios. Subsequently, Johnson et al. [2] constructed a feedforward style transfer network by replacing the successive optimization process with perceptual loss, which significantly improved the generation efficiency. Ulyanov et al. [3] and Dumoulin et al. [4] further promoted rapid stylization and unified modeling of multiple styles, laying an engineering foundation for image style transfer into practical design applications.

In any direction of style transfer, Huang and Belongie proposed an adaptive instance normalization method to achieve fast alignment between content images and style images by matching feature mean and variance [5]. Li et al. [6] used the feature transformation mechanism to enhance the universality of style adaptation. These methods have outstanding performance in style switching efficiency and model flexibility, but there are still problems such as edge drift, local pattern weakening and semantic focus deviation on structure-sensitive objects. For cultural symbols such as stone lions, the head contour, mane organization, limb proportion and decorative details form the basis of their recognition. If only texture replacement is emphasized while the stability of content skeleton is ignored in the transfer process, the generation result is prone to the phenomenon of "style but loss of image".

In order to improve detail preservation and global dependency modeling, multi-scale and attention mechanism have been introduced into related researches. Gatys et al. [7] tried to control the perception factors to make the style transfer results have higher adaptability. Sheng et al. [8] used multi-scale feature decoration to enhance local texture representation. Park and Lee[9] incorporated the style attention mechanism into the transfer network to establish a more targeted matching relationship between the content region and the style region. Recently, Deng et al. proposed StyTr2 model, which uses Transformer for feature reorganization in style transfer process, effectively enhancing the expression ability of long-distance dependencies [10]. This kind of research is more helpful for maintaining complex patterns, comprehensive color layers and spatial layouts, but the cost of model training and inference also increases. It is still necessary to find a balance between expression richness and controllability in cultural image redesign scenarios.

In addition to the generation mechanism, the evaluation problem of style transfer has gradually become the focus of research. Jing et al. [11] and Cai et al. [12] systematically sorted out neural style transfer from the perspective of method evolution and task type respectively, and pointed out that although the current research has continued to improve the visual effect, it still lacks a unified standard for the evaluation of "whether the effect is reasonable". Ioannou and Maddock[13] further pointed out that the quality of style transfer results should not only rely on subjective viewing experience, but also combine multi-dimensional indicators such as content fidelity, style consistency and user perception. In response, image aesthetic evaluation research has developed from high-level feature design [14] to deep learning modeling based on large-scale aesthetic datasets [15-20], which provides a reference path for the quantification of visual appeal, composition balance and semantic clarity. At the same time, research on deep perceptual distance shows that convolutional features have strong effectiveness in reflecting human visual similarity judgment [22], which provides a technical basis for establishing a joint evaluation mechanism of "conformal" and "style change" in the redesign of cultural symbols.

From the application level, research on style transfer of cultural heritage images has begun to focus on the combination between technical generation and cultural expression. Relevant reviews have pointed out that visual style transfer has high potential in digital display of cultural heritage, re-interpretation of traditional images and cultural and creative visual development, but it still generally faces problems such as insufficient semantic continuity, loss of cultural characteristics and imperfect evaluation system [23]. Most of the existing studies regard objects as general images, and few of them establish a special content constraint and aesthetic analysis framework for high symbolic images such as stone lions. It can be seen that how to maintain the core image of Shishi, realize contemporary style transfer, and verify the quality of the results with computable indicators in the process of deep feature transfer is still a problem worthy of further progress. Table 1 summarizes the representative studies related to this paper.

Table 1: Comparison of related studies

Reference	Research Objective	Main Findings	Limitations
Gatys et al. [1]	To establish a neural style transfer method based on CNN features	Achieved separate modeling of content and style, laying the foundation for neural style transfer	Relies on iterative optimization and has high computational cost
Johnson et al. [2]	To improve the real-time performance of style transfer	Significantly improved generation efficiency through feed-forward networks and perceptual loss	The range of styles and model adaptability remain limited
Huang and Belongie [5]	To achieve fast arbitrary style transfer	AdaIN improved adaptability to multiple styles	Edge preservation for complex structured objects is insufficient
Park and Lee [9]	To introduce a style attention mechanism to improve matching performance	Improved coordination between local regions and global style	Training complexity is relatively high, and stability is affected by data distribution
Deng et al. [10]	To apply Transformer to style transfer	Enhanced long-range dependency modeling and complex style representation ability	The parameter scale is large, resulting in high deployment cost
Ioannou and Maddock [13]	To summarize evaluation issues in style transfer	Emphasized the importance of multidimensional evaluation	Lacks unified and directly reusable evaluation standards
Talebi and Milanfar [20]	To build a deep image aesthetic assessment model	Showed that aesthetic quality can be predicted through deep networks	Focuses more on general image scoring and is not targeted at cultural symbols
Wang et al. [23]	To review visual style transfer applications in cultural heritage	Confirmed the potential of this technology in cultural heritage digitization	Specialized frameworks for specific cultural symbols are still insufficient

3 Method Design

The method design of this paper does not regard the stone lion image as a general decorative pattern, but as a visual object with both form recognition and cultural semantics. On this basis, the calculation process of "data organization, feature representation, style transfer and result evaluation" is established. As shown in Figure 1, the study first collects traditional stone lion images, artistic style samples and modern visual design samples to form a cross-source image collection. Since images from different sources have obvious differences in resolution, light and dark distribution, color space and background complexity, the preprocessing stage needs to complete size unification, color conversion, noise suppression and pixel normalization to reduce the influence of irrelevant disturbances on subsequent feature learning and ensure that the model input has good statistical consistency.

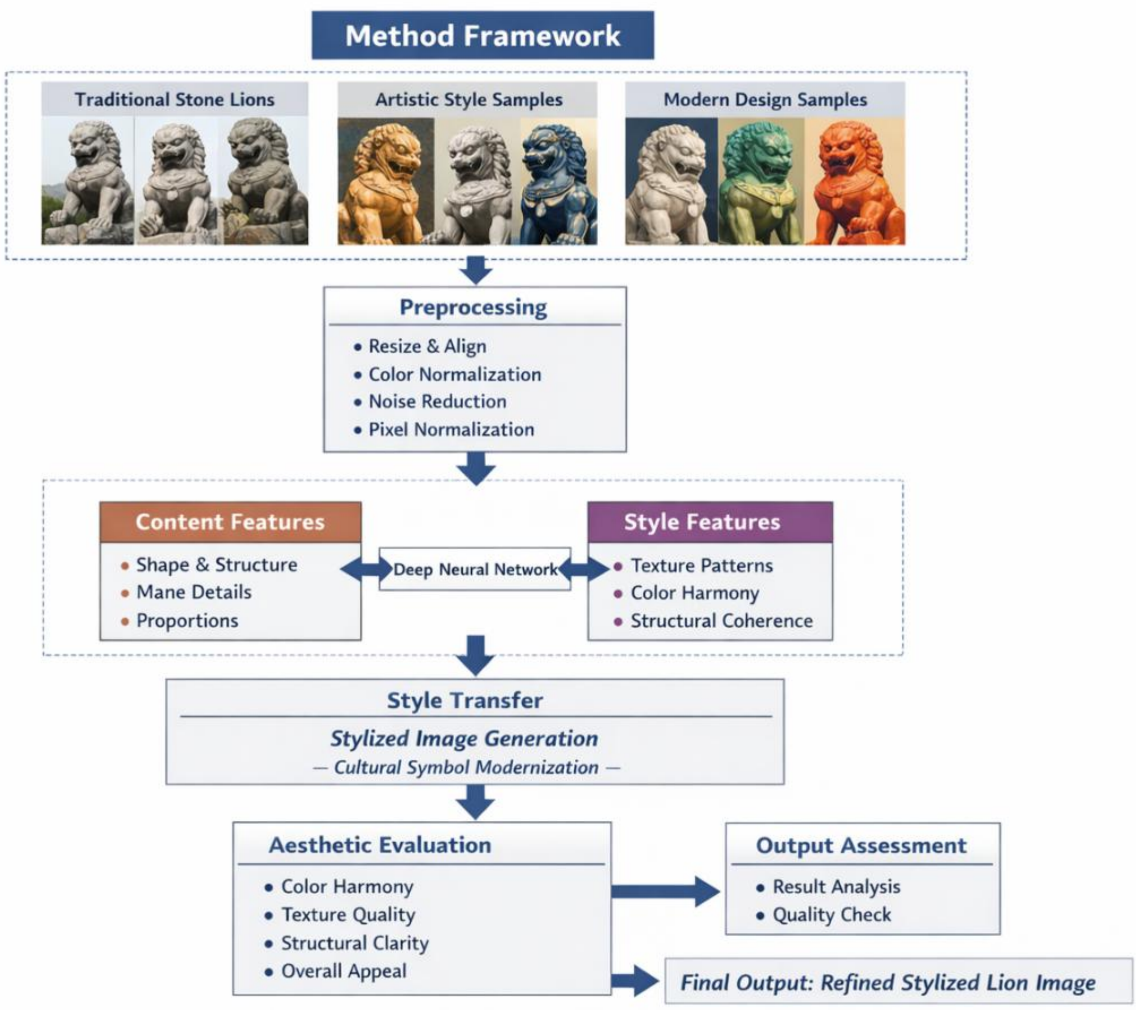

Figure 1: Flow chart of method design

In the feature modeling part, this paper uses deep network to extract content features and style features, and takes the head contour, mane direction, body proportion and decorative boundary of the stone lion as the key preservation objects. Different from the practice of simply pursuing texture replacement, this study emphasizes more on the stable transmission of content skeleton, so that style transfer is based on the premise of "image recognition". Therefore, in addition to content constraints, an aesthetic analysis idea for visual coordination is introduced, which takes color harmony, texture delicacy, structure clarity and overall appeal into the unified scope of investigation, so that the generated results not only have form changes, but also have strong readability and design completion. In the implementation of the model, we construct a neural style transfer framework for the modern expression of cultural symbols. The content preservation term and the style mapping term jointly drive the image redesign, and the aesthetic evaluation results are used to assist the judgment of the output effect. The method path thus formed has two characteristics. On the one hand, it can introduce contemporary visual style without changing the core recognition characteristics of the stone lion. On the other hand, it makes the redesign process of traditional imaging transform from experience-led to computation-assisted, and provides a repeatable technical basis for subsequent experimental comparison and application expansion.

3.1 Stone lion image and style sample collection

The data collection of this paper does not directly apply the general image style transfer dataset, but focuses on the redesign task of the Shishi image, and constructs a multi-source image collection composed of content images, style images and auxiliary evaluation samples. The content images are mainly used to carry the basic form and cultural identification information of the stone lion. The sources include the digital collection of public museums, the image data of ancient buildings, the local cultural protection exhibition platform, and the field images collected during the research process. During the collection, the differences in head structure, mane organization, limb proportion, base decoration and expression characteristics of stone lions in different regions, different ages and different materials are preserved, so as to avoid too single training samples, which leads to style solidifaction or shape deviation in the generated results.

Style samples serve the visual transfer stage, mainly selecting modern illustrations, decorative paintings, flat posters, comprehensive color images and digital art works with obvious contemporary design language. Such images do not assume the task of content expression of the stone lion, but provide style information such as color configuration, texture rhythm, line and plane relationship, and overall atmosphere. In order to ensure the comparability of style input, the images with low resolution, severe subject occlusion and strong visual noise are eliminated during sample selection, and the preliminary classification is carried out according to color tendency, texture complexity and composition method. In addition to the required training images, this paper also collates a part of aesthetic evaluation auxiliary samples to support the comparative analysis of the transfer results in visual appeal and style coordination.

From the perspective of computer processing, this collection method has two direct effects. On the one hand, it makes the Shishi content domain and modern style domain form a clearer function division at the data level, which is convenient for subsequent content-style decoupling modeling. On the other hand, the introduction of multi-source samples can improve the generalization ability of the model in the face of complex shapes and diverse styles, and reduce the risk of texture distortion, edge collapse or cultural feature weakening in local areas of the generated results. After comprehensive arrangement, the study finally forms the image sample set that can be used for training, validation and result analysis, and its composition is shown in Table 2.

Table 2: The composition of stone lion images and style samples

Sample Type	Main Source	Number of Samples / Images	Main Purpose	Key Control Points
Stone Lion Content Images	Museum digital collections, ancient architecture archives, and field photography	820	Provide the main structural features and cultural identity characteristics of stone lions	Complete contours, clear poses, and low background interference
Artistic Style Samples	Modern illustrations, decorative paintings, and digital art images	640	Provide color, texture, and visual style information	Distinct stylistic differences and stable resolution
Modern Design Samples	Posters, cultural and creative graphics, and visual communication works	310	Serve as the reference domain for modernized expression	Clear composition and contemporary design language
Aesthetic Evaluation Support Samples	High-quality visual design cases and screened images	180	Support result comparison and aesthetic analysis	High visual quality and distinguishable styles

3.2 Image preprocessing and data standardization

Image preprocessing is a key step before neural style transfer into effective training. Its role is not only to "arrange samples", but to weaken the irrelevant statistical fluctuations between images from different sources, so that the model can focus on the learning of the morphology, texture organization and style mapping relationship of the stone lion. The data used in this paper includes traditional stone lion images, artistic style samples and modern visual design samples. Different sources differ significantly in resolution, brightness range, color distribution, background complexity and edge sharpness. If the original image is directly fed into the network, the convolutional features are easily disturbed by scale inconsistency and pixel distribution shift, which affects the stability of content representation and the controllability of style transfer. Therefore, a unified preprocessing process is set up before model training, including size scaling, color space unification, pixel normalization and basic data standardization, so as to construct a more stable input domain.

In the dimension processing stage, all images are uniformly scaled to 256×256 pixels. In this way, on the one hand, the video memory overhead and training time can be controlled, so that the content image and the style image maintain the same tensor structure during batch processing. On the other hand, it is also beneficial to the core visual elements such as the head contour, mane distribution, limb posture and base boundary of the stone lion to be extracted by the network at the same scale. In the process of scaling, bilinear interpolation is used, and the image with large width/height ratio deviation is cropped first and then resampling is performed to avoid facial compression, body elongation or local texture distortion caused by direct stretching. For the samples with strong background noise, lightweight edge-preserving filtering is also performed during preprocessing to reduce the interference of shadows, masonry background and stray decorations on the main area.

In the color processing stage, the original image is uniformly converted to RGB color space, and the three-channel input form is maintained. This is not just for the sake of format uniformity, but more important to ensure that the color statistics of different source images are comparable. The surface of limestone, bluesite, copper or painted commonly seen in stone lion images is prone to color temperature drift and local exposure uneven under changing shooting conditions. However, style images often have strong subjective color organization and contrast. If the two types of data are not uniformly processed, the network is easy to misidentify the device difference as the style difference, which weakens the learning effect of real style features. Based on this, in this paper, after the color space is unified, the brightness truncation and contrast compression of the abnormal exposure samples are performed to make the input distribution closer to the trainable state.

Normalization processing is used to further eliminate pixel value range differences. Let the value of a pixel in the original image be x, and the minimum and maximum values of this channel in the data set be x_{min} and x_{max}, respectively, then the min-max normalization result is defined as:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

After the transformation of equation (1), all pixel values are mapped to the interval [0,1]. This processing can reduce the numerical oscillation caused by different brightness baselines between different images, so that the generator and the discriminator receive a more stable input distribution in the adversarial training. For neural style transfer, normalization also has a more direct significance: it makes the response range of content features and style features closer, which helps to reduce the phenomenon that one feature is too strong to suppress the other, thus improving the visual coordination of transfer results.

In addition to pixel-level standardization, the data input is tensorized with batch standard organization in this paper. Let the normalized image be represented as $I \in \mathbb{R}^{H\times W\times C}$, then the data input to the network after tensor transformation is of the form:

$$X \in \mathbb{R}^{N\times C\times H\times W} \tag{2}$$

where N is the batch size, C is the number of channels, H and W are the image height and width, respectively. This representation facilitates the subsequent convolutional layers to extract local edges, textures and high-level semantic structures uniformly, and also enables the data in the content domain and style domain of Shishi to complete feature alignment in the same computational graph. Practice shows that after the above preprocessing, the dispersion degree of the input image in scale, color and numerical distribution is significantly reduced, the loss oscillation in the model training process is smaller, and the generated results are easier to achieve a balance in structure clarity, edge integrity and style expression stability. This provides a reliable data basis for the subsequent content, the style representation and the modernization migration of the stone lion image.

3.3 Content-style Representation and Computational Aesthetic Feature Extraction Based on Deep Features

In order to ensure that the stone lion image still has stable cultural recognition after style transfer, this paper no longer understands the content and style as a simple pixel replacement relationship, but uses the deep convolutional network to hierarchical represent the image. After the image of the stone lion is input into the feature extraction network, the shallow response mainly retains the edge, turning point and texture details, which can be used to describe the fluctuation of the mane, the outer contour of the ear and the decorative boundary of the base. The middle and high level responses are closer to semantic structure, which can reflect head proportion, body posture and visual center of gravity. This method of feature organization from local to global is more suitable for dealing with traditional statues such as Shi lion, which are both decorative and symbolic, and also provides a more stable content constraint for subsequent style mapping.

In the content representation, the original stone lion image I_c and the generated image I_g are input into the pre-trained network at the same time, the feature map is extracted in the l layer, and the degree of content retention is measured by the feature difference. The content loss is defined as:

$$L_{content}^{(l)} = \frac{1}{C_l H_l W_l} \left\| F^{(l)}(I_g) - F^{(l)}(I_c) \right\|_2^2 \tag{3}$$

where, $F^{(l)}(\cdot)$ represents the feature map of the LTH layer of the network; C_l, H_l and W_l denote the number of channels, height and width of the feature map of the layer, respectively. $\|\cdot\|_2^2$ is the squared Euclidean distance. This formula is used to constrain the generation result to be consistent with the original stone lion at the structure level, avoiding problems such as head compression, loose contour or posture deviation in the style transfer process.

The style representation is completed by feature correlation modeling. Different from content features that emphasize "shape", style is closer to the statistical relationship between channels, reflecting color organization, texture density and stroke rhythm. In this paper, Gram matrix is used to characterize the style distribution of a certain layer feature, which is defined as:

$$G_{ij}^{(l)} = \sum_{k=1}^{H_l W_l} F_{ik}^{(l)} F_{jk}^{(l)} \tag{4}$$

where, $G_{ij}^{(l)}$ represents the correlation between the ith channel and the JTH channel of the LTH layer; $F_{ik}^{(l)}$ and $F_{jk}^{(l)}$ are the characteristic responses of the corresponding channel at position k, respectively. The closer the Gram matrix is, the more similar the generated image is to the target style image in texture structure and color rhythm. Based on this, the style loss can be expressed as:

$$L_{style} = \sum_{l \in \Omega} \frac{1}{C_l^2} \left\| G^{(l)}(I_g) - G^{(l)}(I_s) \right\|_F^2 \tag{5}$$

where Ω is the set of feature layers participating in style constraints; I_s is the style reference image; $\|\cdot\|_F^2$ denotes the square of the Frobenius norm. This formula is used to control the absorption degree of the generated image to the target style, so that the stone lion image can obtain more color and texture expression of contemporary visual language while retaining the main skeleton.

Targeting only content loss and style loss may still yield results that are "transferable but not good-looking enough", so we further introduce computational aesthetic features. Aesthetic extraction does not rely on subjective descriptions, but constitutes a comprehensive scoring vector from the dimensions of composition balance, clarity, color harmony, and visual appeal.

Let the aesthetic feature vector of the generated image be a_g and the target aesthetic distribution vector be a_r, then the aesthetic constraint term is written as:

$$L_{aesthetic} = \|a_g - a_r\|_2^2 \tag{6}$$

where, a_g represents the generated image features output by the aesthetic evaluation network; a_r represents the aesthetic reference features of high-quality modern visual samples. The function of this item is not to pursue abstract "high score image", but to avoid color imbalance, local stacked too dense or disordered visual center of gravity in the generated results, so that the modern expression of the stone lion image is more in line with the viewing habits of the design communication scene.

In general, through the joint extraction of deep content features, style statistical features and aesthetic evaluation features, the redesign process of Shishi is transformed from empirical modification to a feature reorganization process that can be calculated and controlled. The representation mechanism established in this way can not only protect the core contour of the stone lion as a cultural symbol, but also provide a clearer optimization direction for the subsequent style transfer model.

3.4 Neural Style Transfer Model for Modern Expression of Cultural Symbols

The redesign of stone lion image is not a problem of style replacement between ordinary images, but a generative task that requires "image recognizability", "style transferability" and "aesthetic controllability" at the same time. There is usually a lack of strict pairing relationship between traditional stone lion images and modern style samples, so it is difficult for the model to obtain a stable correspondence map if the supervised image translation method is directly used. Based on this characteristic, this paper constructs a dual-domain neural style transfer model for the modern expression of cultural symbols, and completes the bidirectional learning between the content domain X of the stone lion and the modern style domain Y under the condition of no paired samples. As shown in Figure 2, the model consists of two generators and two discriminators: generator G:X→Y is used to map the traditional stone lion image to the modern style domain, and generator F:Y→X is used to reverse map the style domain image back to the stone lion content domain. The discriminators DY and DX are responsible for judging the authenticity of the target domain image respectively. This structure does not pursue to complete style coverage at one time, but uses bidirectional constraints to make the generated results form a relatively stable balance between visual changes and content preservation.

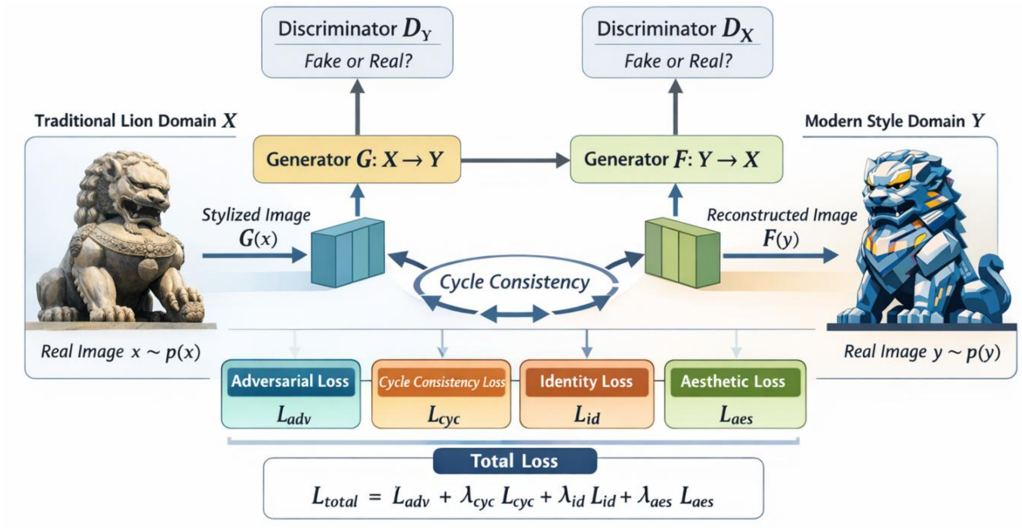

Figure 2: Model structure diagram

The model training is based on adversarial learning, and its overall optimization objective can be written as:

$$\min_{G,F} \max_{D_X,D_Y} L_{total} (G, F, D_X, D_Y) \tag{7}$$

where G and F denote two generators; D_X and D_Y denote the two discriminators; L_{total} is the overall loss function. This equation shows that the generator wants to generate images that are enough to fool the discriminator, while the discriminator tries to distinguish between real images and generated images, and the two constantly update their parameters in the game.

In order to enable the model to complete the style transfer without destroying the structural identifiability of the stone lion, the total loss is defined as:

$$L_{total} = L_{adv} + \lambda_{cyc} L_{cyc} + \lambda_{id} L_{id} + \lambda_{aes} L_{aes} \tag{8}$$

where, L_{adv} represents adversarial loss, which is used to ensure the realism of the generated image on the target domain. L_{cyc} represents cycle consistency loss, which is used to constrain the reconstruction result after bidirectional mapping. L_{id} stands for identity constraint loss, which is used to mitigate unnecessary color drift and structural perturbation. L_{aes} represents aesthetic constraint loss, which is used to improve the stability of the generated results in visual coordination. λ_{cyc}, λ_{id} and λ_{aes} are the weight coefficients of each loss term.Here, the adversarial loss is defined as:

$$\begin{aligned} L_{adv} = \mathbb{E}_{y\sim p(y)}[\log D_Y (y)] + \mathbb{E}_{x\sim p(x)}[\log(1 - D_Y(G(x)))] + \\ \mathbb{E}_{x\sim p(x)}[\log D_X (x)] + \mathbb{E}_{y\sim p(y)}[\log(1 - D_X(F(y)))] \end{aligned} \tag{9}$$

where, p(x) and p(y) represent the data distribution of Shishi content domain and modern style domain respectively. G(x) is the modern style result generated by the stone lion image. F(y) is the content domain result obtained from the style image reconstruction. The function of this term is to push the generator to learn the distribution characteristics of the target domain, so that the output image is closer to modern visual design samples in color organization, texture expression, and overall temperament.

Cycle consistency is the key to maintain the continuity of cultural symbols in this model. Its expression is:

$$L_{cyc} = \mathbb{E}_{x\sim p(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y\sim p(y)}[\|G(F(y)) - y\|_1] \tag{10}$$

where, F(G(x) represents the result of the Shishi image after forward transfer and then reverse reconstruction, G(F(y) represents the result of the style image after reverse mapping and then forward restoration, and $\|\cdot\|_1$ is the L_1 norm. This requirement requires the image to return to the original state as far as possible after the bidirectional conversion, so as to reduce the risk of the head contour, mane trend and body proportion of the stone lion being overrewritten during the transfer process. Considering that cultural symbol images are more dependent on local recognition cues than general natural images, this paper adds the identity constraint term:

$$L_{id} = \mathbb{E}_{y\sim p(y)}[\|G(y) - y\|_1] + \mathbb{E}_{x\sim p(x)}[\|F(x) - x\|_1] \tag{11}$$

where, G(y) and F(x) represent the output result of the generator when the image itself belongs to the target domain. This term is used to limit the invalid perturbation of the model on the existing style features, preventing the generator from over-" stylizing "all the inputs, thus improving color stability and edge integrity. Based on this, the computational aesthetic features extracted in Section 3.3 are introduced into the optimization process to construct the aesthetic constraint loss as follows.

$$L_{aes} = \|a(G(x)) - a_r\|_2^2 \tag{12}$$

where, $a(G(x))$ represents the aesthetic feature vector of the generated image, a_r represents the reference aesthetic vector of the modern visual sample, and $\|\cdot\|_2^2$ is the squared Euclidean distance. This item does not directly replace the content and style constraints, but complements and corrects the generated results from the perspectives of composition balance, visual clarity, color coordination and attractiveness, so that the problem of "strong texture and unstable picture" will not appear in the process of modern expression of the stone lion image.

4 Results and discussion

4.1 Experimental environment Setup

In order to verify the feasibility and stability of the neural style transfer framework in the redesign task of the Shishi image, this paper completes the model training, result generation and index evaluation in a unified software and hardware environment. The experimental platform uses Windows 11 64-bit operating system, Intel Core i7-12700 processor, 32 GB memory, and NVIDIA RTX 3080 10 GB graphics processor. The model development language is Python 3.10, the deep learning framework is PyTorch 2.1, and the image processing and data organization rely on tool libraries such as NumPy, OpenCV, Pillow and torchvision. The configuration can not only support batch training of unpaired images in the process of dual-domain mapping, but also meet the memory requirements of the generator, discriminator and aesthetic evaluation module when calling in parallel. In the process of experimental implementation, independent data reading pipelines are established for the Shishi content image domain and the modern style image domain, and the forward transfer, reverse reconstruction and loss return are completed in the same training round. In the model training phase, GPU acceleration is used to improve the execution efficiency of convolutional feature extraction, adversarial update and loop consistency calculation. In the testing phase, the random seeds and inference parameters are fixed to ensure comparability between different experimental groups. The overall environment setting takes into account the training stability, computational efficiency and result reproduction requirements, which provides a reliable experimental basis for subsequent parameter configuration, effect evaluation and comparative analysis.

4.2 Setting of model parameters

In order to balance the structure stability, texture expression and training convergence speed in the style transfer process of the stone lion image, the model training parameters are configured uniformly. Considering that the stone lion image has strong contour recognition requirements, and the modern style samples have large fluctuations in color and texture, the parameter setting does not simply pursue the generation speed, but pays more attention to the controllability and result consistency in the dual-domain mapping process. In the experiment, the total network training rounds are set to 100, the initial learning rate is set to 0.0002, the optimizer uses Adam, the first-order momentum coefficient is 0.5, and the second-order momentum coefficient is 0.999. This set of Settings helps alleviate gradient oscillations in adversarial training, keeping the generator and discriminator in a relatively balanced update rhythm.

The input image size is uniformly 256×256 and the batch size is set to 8. In this way, on the one hand, it can meet the stable training requirements under the limitation of GPU memory, on the other hand, it is also convenient for the model to obtain a reasonable feature resolution between local texture and global contour. The main body of the generator adopted a residual structure, the number of residual blocks was set to 9, and the number of basic convolution channels was set to 64. The discriminator is in the form of PatchGAN to enhance the discrimination ability of local texture authenticity and edge details. To avoid excessive destruction of the head, mane and body boundaries of the stone lion during migration, the weight of the cycle consistency loss is set to 10, the weight of the identity loss is set to 5, and the weight of the aesthetic constraint is set to 2. The parameter pre-experiment results show that the model under the above configuration has a more stable loss change in the later training stage, and the generated images are also more in line with the research objectives of this paper in terms of structure clarity and style coordination. The specific parameters are shown in Table 3.

Table 3: Model parameter Settings

Parameter	Setting
Input Image Size	(256 × 256)
Number of Training Epochs	100
Batch Size	8
Initial Learning Rate	0.0002
Optimizer	Adam
β1	0.5
β2	0.999
Base Number of Convolution Channels	64
Number of Residual Blocks in Generator	9
Discriminator Type	PatchGAN
Activation Functions	ReLU, Leaky ReLU
Cycle-consistency Loss Weight	10
Identity Loss Weight	5
Aesthetic Constraint Loss Weight	2

4.3 Evaluation of redesign effect

In the redesign task of the stone lion image, the judgment of the model effect cannot only be based on whether the style change is obvious, but more importantly, whether the original recognition features of the stone lion are stably retained after the transfer is observed. Based on this consideration, this paper takes the cycle consistency loss as the core index in the evaluation of the redesign effect, and combines the visual performance of the generated results to comprehensively analyze the structure preservation ability and style transfer stability during the training process of the model. The cycle consistency loss reflects how different an image is from the original input after undergoing a forward style mapping and a backward reconstruction. The lower the index is, the better the model can maintain the key content information of the stone lion's head contour, mane level, body posture and base boundary without excessive damage while introducing modern visual language.

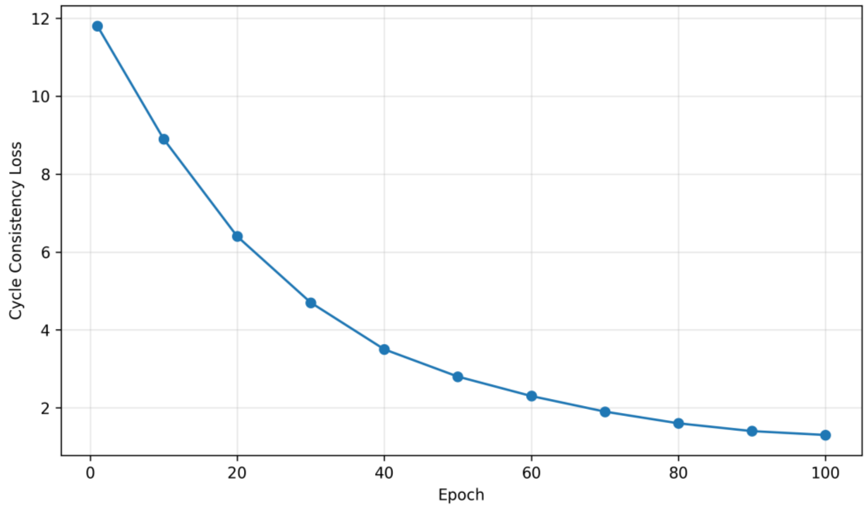

Figure 3: Change diagram of cycle consistency loss

As shown in Figure 3, the cycle consistency loss is at a high level in the early stage of model training, which is about 11.8 in the first round. This indicates that when the parameters have not yet fully converged, the generator's grasp of the coordination between the content structure and the target style is still unstable, and the reconstruction error is large. As the training continues to progress, the loss value shows a continuous downward trend, which decreases to 4.7 in the 30th round, further decreases to 2.8 in the 50th round, and stabilizes around 1.3 in the 100th round. This change indicates that the model gradually learns to retain the backbone structure of the stone lion image in the dual-domain mapping, while absorbing the color rhythm and texture features in the modern style image. In other words, the generated results in the later stage of training are no longer pure surface style coverage, but achieve a balance between content skeleton and style expression to a large extent.

From the actual observation of the generated samples, the loss decrease is basically consistent with the visual quality improvement. In the early stage of training, some results will show loose local edges, adhesion of decorative patterns and imbalance of light and dark relations. After entering the middle and late stages, the facial structure of the stone lion is clearer, the proportion of the mane organization and body is more stable, and the embedding of modern colors and textures is more natural. This shows that the cycle consistency loss can not only reflect the convergence state of the model, but also better correspond to the retention degree of the recognition features of the stone lion culture. Overall, the evaluation results show that the proposed method has good structure preservation ability and style transfer stability in the redesign task of the stone lion image, which provides a reliable basis for subsequent comparative experimental analysis.

4.4 Comparative experimental analysis

In the comparative experiment stage, this paper compares the proposed model with three representative methods: AdaIN, standard CycleGAN and StyTr2, in order to observe the quality differences of the redesign of the stone lion image under different style transfer frameworks. Considering that the task of this paper is not only to complete the image style replacement in a general sense, but also to preserve the structure identification of the stone lion as a cultural symbol in the process of modern expression, the evaluation indicators are selected to evaluate the peak Signal-to-Noise Ratio (PSNR) and structural similarity index (SSIM). PSNR mainly reflects the reconstruction fidelity between the generated image and the reference result, the higher the value, the smaller the image distortion. SSIM pays more attention to the consistency of brightness, contrast and structural information, which is more suitable to measure the maintenance effect of the head contour, mane organization and body posture of the stone lion after migration.

To avoid contingency in a single style scenario, the test samples are divided into three groups: decorative style, poster style and illustration style. Figure 4 shows the PSNR comparison results of different models on the three types of style sets. It can be seen that the PSNR of AdaIN in the three groups of tasks is 22.4dB, 21.9dB and 22.7dB respectively, and the corresponding PSNR of standard CycleGAN is 23.1dB, 22.8dB and 23.4dB. StyTr2 further improves to 24.0 dB, 23.6 dB, and 24.3 dB, while the proposed model achieves 25.3 dB, 24.9 dB, and 25.8 dB. According to the three sets of results, the average PSNR of the proposed model is 25.33 dB, which is higher than 23.10 dB of the standard CycleGAN and 23.97 dB of StyTr2, indicating that after the introduction of content preservation and aesthetic constraints, the average PSNR of the proposed model is 25.33 dB. The phenomena of edge blurring, local texture stacking and light and dark imbalance in the style transfer process of stone lion images are more effectively suppressed.

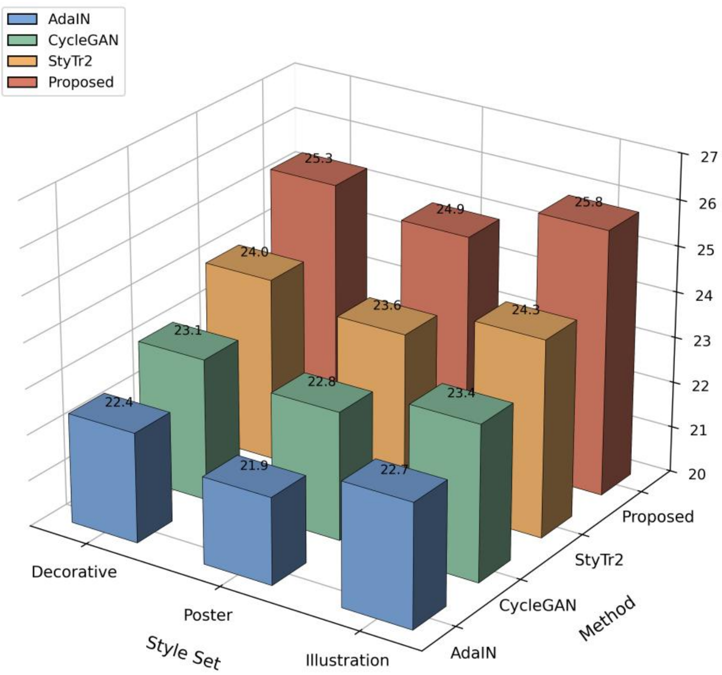

Figure 4: Comparison of PSNR

Figure 5 illustrates the comparison results of SSIM. The SSIM of AdaIN in the three groups of styles are 0.742, 0.731 and 0.748, 0.781, 0.769 and 0.786 for standard CycleGAN, 0.812, 0.801 and 0.818 for StyTr2, and 0.846, 0.834 and 0.851 for the proposed model. The corresponding average SSIMs are 0.740, 0.779, 0.810 and 0.844, respectively. This result shows that the proposed model not only maintains high quality at the pixel level, but also maintains the stability of the Shishi visual prototype at the structural level. Especially in the test set of decorative style and illustration style, the retention of stone lion mane layers, eye boundaries and base contours is more complete. Although the color and texture of modern style are obviously embedded, it does not cause the problem of "losing the object if you change the style".

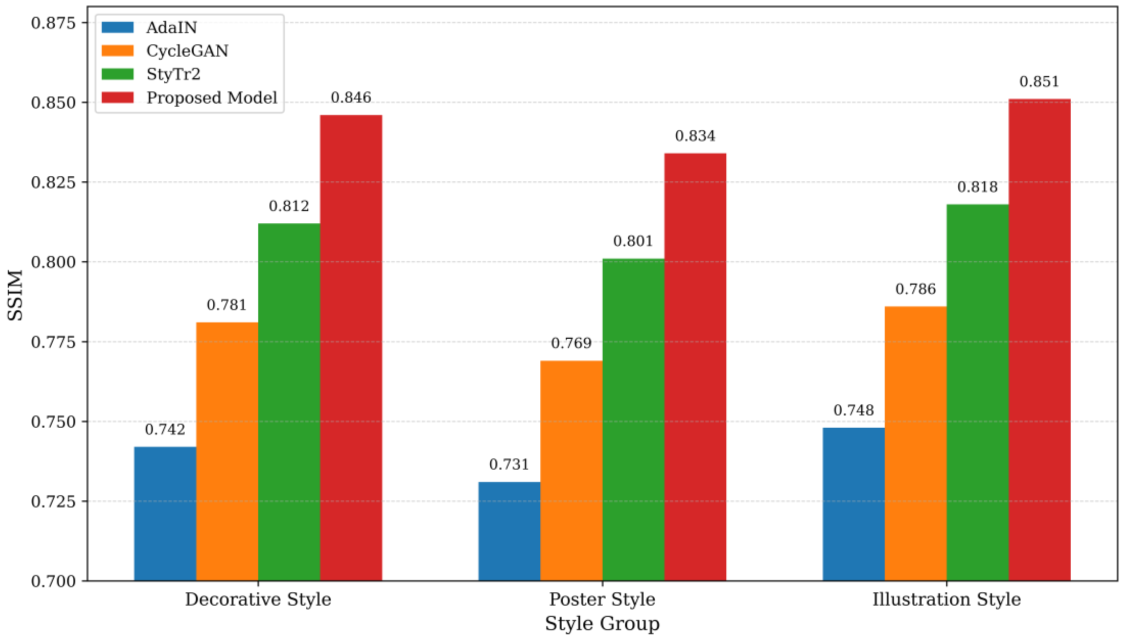

Figure 5: Comparison of SSIM

Combining the two indicators, it can be seen that AdaIN has faster style injection ability, but it is easy to sacrifice structural accuracy on high symbolic objects. Standard CycleGAN is more stable on dual-domain mapping, but it still has details loose in complex pattern areas. StyTr2 is more expressive of global style relations, but there is still a slight drift of local edges on objects with distinct contour features such as stone lions. In contrast, the proposed model achieves a more appropriate balance between structure preservation and visual update through the combined effects of cycle consistency, identity constraints, and computational aesthetic guidance. In other words, it does not simply generate images with "stronger style", but more effectively completes the transformation of stone lion cultural symbols from traditional visual forms to modern communication forms, which is also consistent with the research objectives proposed in this paper.

4.5 Comparison of result data

In order to further investigate the adaptability of the model under different data conditions, this paper divides the test results into two types of data scenarios for comparison. One is the standard shape data set of the stone lion. The samples are mainly traditional stone lion images with complete contour, clear posture and less background interference, which are more suitable for testing the ability of the model to maintain the core recognition features such as head structure, mane distribution and base boundary. The other is the Shi Lion cross-style transfer dataset, whose content images are used with modern illustrations, poster design and art Deco style samples, with a larger style span, more obvious texture levels and color fluctuations, which can better reflect the transfer stability of the model under complex visual conditions.

The experimental results show that the model performs better on the standard morphological data set, with PSNR reaching 25.62 dB, SSIM reaching 0.86, and the average aesthetic score reaching 7.94. On the cross-style transfer dataset, the PSNR is 24.18 dB, the SSIM is 0.81, and the average aesthetic score is 7.48. The difference between the two groups of results shows that when the input image has clearer structural boundaries and more stable content distribution, the model is more likely to complete the high-fidelity style mapping, and the facial contour, body proportion, and decoration level of the stone lion are less likely to deviate. In contrast, due to the large internal differences in the style domain, some samples will suffer from detail compression and edge weakening during local texture injection in the cross-style dataset, so the related indicators slightly decrease.

However, from the perspective of the overall data, the indicators under the two types of scenes do not fluctuate greatly, indicating that the proposed method maintains a good robustness to the structure of the stone and lion cultural symbols, and also shows that the introduced aesthetic constraints can still play a certain regulatory role in complex style environments. In other words, our model is not only suitable for regular samples, but can still maintain relatively stable redesign quality in the face of modern visual materials with stronger style changes. The specific results are shown in Table 4.

Table 4: Comparison of results on different datasets

Metric	Standard Stone Lion Morphology Dataset	Cross-style Stone Lion Transfer Dataset
PSNR / dB	25.62	24.18
SSIM	0.86	0.81
Aesthetic Score	7.94	7.48
Content Preservation Rate / %	91.3	87.6

4.6 Discussion

This paper aims to construct a neural style transfer method for the modern expression of stone and lion cultural symbols, so that the generated results can obtain contemporary visual features without weakening the recognition basis of the original image. The experimental results show that although the model can complete color and texture replacement only by relying on the general style transfer mechanism, it is prone to problems such as edge drift, local decoration adhesion and pose center deviation on high symbolic objects such as stone lions. Comparison experiments show that the average PSNR of the proposed model reaches 25.33 dB, which is 2.23 dB higher than that of the standard CycleGAN, and the average SSIM reaches 0.844, which is 0.065 higher than that of the standard CycleGAN, indicating that a more stable balance between structure preservation and style mapping is achieved. At the same time, the content recognition retention rates on the Shi lion standard shape dataset and the cross-style transfer dataset reach 91.3%and 87.6%respectively, which also shows that the proposed method has a good ability to protect the core contour of cultural symbols. However, this result does not mean that the model has completely solved all the problems in cultural image redesign. When the texture density of the style samples is too high and the color contrast is too strong, some generated results still have detail compression and local level overload, and the aesthetic score also drops from 7.94 to 7.48. This shows that the modern expression of cultural symbols cannot be simply understood as the accumulation of style strength, but should be based on the collaborative control of content structure, visual order and communication adaptability. In this regard, the significance of our method is not only to improve the generation quality, but also to provide a relatively clear technical path for the computational redesign of traditional image resources.

5 Conclusion

Focusing on the problem of structure preserving and style updating in the modern expression of the stone lion image, this paper constructs a redesign framework combining neural style transfer and computational aesthetic evaluation. Based on the traditional stone lion images, modern art style samples and visual design samples, this study completes multi-source image acquisition, preprocessing standardization, content-style depth representation, and dual-domain generative mapping modeling, so that the stone lion, a traditional cultural symbol, obtains a more contemporary visual expression in the digital environment. Different from the practice of simply pursuing texture replacement, this paper emphasizes the continuation of core recognition features such as head contour, mane organization, and body posture, so as to control the style transfer within the scope of "updating without losing shape". Experimental results show that the model has good training stability and structure preservation ability. In the later stage of training, the cycle consistency loss is reduced to 1.3, the average PSNR reaches 25.33 dB, the average SSIM reaches 0.844, and the content recognition retention rate on the Shi Lion standard shape dataset reaches 91.3%. These results show that the proposed method can effectively retain the original cultural identity of the stone lion while introducing modern color, texture and composition language. However, when the style sample has strong abstraction or high density texture, the local area may still suffer from detail compression and level overload. Future research can continue to deepen the attention constraint, adaptive style strength control, and feature guidance oriented to cultural semantics, so as to improve the generalization ability and design flexibility of the model in more complex visual scenes.

Funding

Annual Project of Philosophy and Social Science Research in Shaanxi Province, Grant No. 2025YB0043.

References

[1] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2414-2423.

[2] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Cham: Springer International Publishing, 2016: 694-711.

[3] Ulyanov D, Lebedev V, Vedaldi A, et al. Texture networks: Feed-forward synthesis of textures and stylized images[J]. arXiv preprint arXiv:1603.03417, 2016.

[4] Dumoulin V, Shlens J, Kudlur M. A learned representation for artistic style[J]. arXiv preprint arXiv:1610.07629, 2016.

[5] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance

normalization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 1501-1510.

[6] Li Y, Fang C, Yang J, et al. Universal style transfer via feature transforms[J]. Advances in neural information processing systems, 2017, 30.

[7] Gatys L A, Ecker A S, Bethge M, et al. Controlling perceptual factors in neural style transfer[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3985-3993.

[8] Sheng L, Lin Z, Shao J, et al. Avatar-net: Multi-scale zero-shot style transfer by feature decoration[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8242-8250.

[9] Park D Y, Lee K H. Arbitrary style transfer with style-attentional networks[C]// proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 5880-5888.

[10] Deng Y, Tang F, Dong W, et al. Stytr2: Image style transfer with transformers[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11326-11336.

[11] Jing Y, Yang Y, Feng Z, et al. Neural style transfer: A review[J]. IEEE transactions on visualization and computer graphics, 2019, 26(11): 3365-3385.

[12] Cai Q, Ma M, Wang C, et al. Image neural style transfer: A review[J]. Computers and Electrical Engineering, 2023, 108: 108723.

[13] Ioannou E, Maddock S. Evaluation in neural style transfer: a review[C]//Computer Graphics Forum. 2024, 43(6): e15165.

[14] Ke Y, Tang X, Jing F. The design of high-level features for photo quality assessment[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, 2006, 1: 419-426.

[15] Murray N, Marchesotti L, Perronnin F. AVA: A large-scale database for aesthetic visual analysis[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 2408-2415.

[16] Lu X, Lin Z, Jin H, et al. Rapid: Rating pictorial aesthetics using deep learning[C]//Proceedings of the 22nd ACM international conference on Multimedia. 2014: 457-466.

[17] ai, L., Jin, H., & Liu, F. Composition-Preserving Deep Photo Aesthetics Assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 497–

[18] Kong S, Shen X, Lin Z, et al. Photo aesthetics ranking network with attributes and content adaptation[C]//European conference on computer vision. Cham: Springer International Publishing, 2016: 662-679.

[19] Kao Y, He R, Huang K. Deep aesthetic quality assessment with semantic information[J]. IEEE Transactions on Image Processing, 2017, 26(3): 1482-1495.

[20] Talebi H, Milanfar P. NIMA: Neural image assessment[J]. IEEE transactions on image processing, 2018, 27(8): 3998-4011.

[21] Deng Y, Loy C C, Tang X. Image aesthetic assessment: An experimental survey[J]. IEEE Signal Processing Magazine, 2017, 34(4): 80-106.

[22] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 586-595.

[23] Wang Y, Wu G, Wang S, et al. Visual style transfer for cultural heritage: A comprehensive review of techniques and applications[J]. Design and Artificial Intelligence, 2025: 100027.