ITC 1/55

Information Technology and Control
Vol. 55 / No. 1/ 2026
pp. 280-297
DOI 10.5755/j01.itc.55.1.43111

Image Style Transfer for Visual Design Using StarGANv2 and Attention Mechanism

Received 2025/10/15 | Accepted after revision 2026/01/11

HOW TO CITE: Sun, M., Peng, X. (2026). Image Style Transfer for Visual Design Using StarGANv2 and Attention Mechanism *Information Technology and Control,* 55(1), 280-297. https://doi.org/10.5755/j01.itc.55.1.43111

Image Style Transfer for Visual Design Using StarGANv2 and Attention Mechanism

Maishuang Sun

School of Art and Design, Fuzhou University of International Studies and Trade, Fuzhou, 350202, China

Xueying Peng*

Visual Communications Department, Kyungil University, Daegu, 38428, Korea

Corresponding author: 15039602087@163.com

In visual design, high-quality facial attribute style transfer and diversified painting style transfer are key to meeting the diverse artistic creation needs. To improve the facial image generation effect in vis-ual design, the study first designed a facial attribute style transfer method based on an improved star shaped generative adversarial network version 2. The generator introduced dense connection modules and random noise modules, and a spectral normalization network was added to the discriminator. On the basis of this model, a painting style transfer model was studied and constructed. Its generator adopted distribution shift convolution and attention mechanism, while its discriminator used spectral normalization and Markov dual discriminator. The findings demonstrated that the attribute generation accuracy of the facial attribute style transfer method was 98.56%, and its average time consumption in practical applications was 71.2ms, significantly better than the comparison model. The PSNR and max-imum structural similarity index of the painting style transfer model are 38.68dB and 0.948. Both mod-els had good performance and could provide efficient and high-quality style transfer solutions for visu-al design. This research holds immense importance in fostering the advancement of artistic creation in the field of visual design.

KEYWORDS: StarGANv2; Attention mechanism; Style transfer; Facial attributes; Painting style

1. Overview

With the deep integration of digital technology and creative industries, visual design has become a core link in fields such as advertising and marketing, film and animation, game development, and social content creation. Its demand for diversified and personalized image styles is increasingly prominent

[28, 30]. In visual design practice, facial images, as high-frequency design elements, often need to adjust attributes such as gender, age, and hairstyle according to scene requirements to match brand tone or narrative themes. Moreover, as a key means to enhance the artistic sense of design, the transfer of painting styles requires the natural transformation of different styles such as oil painting, sketching, and cartoons to enrich visual expression forms [11, 15]. However, in traditional visual design, stylization relies on designers manually drawing and adjusting, which not only has a long creation cycle and high labor costs, but also makes it difficult to ensure consistency in the style of batch generated images. Therefore, it is necessary to introduce efficient, accurate, and stable new technologies to achieve image style transfer. Regarding this issue, commonly used technologies include Generative Adversarial Networks (GAN), Convolutional Neural Networks, Star Shaped Generative Adversarial Network version 2 (starGANv2), Transformer-based methods, and CycleGAN [31, 24]. Meanwhile, many scholars have also explored this issue.

Some studies focus on the fidelity of details and semantic consistency in style transfer. Xie et al. [26] proposed a semantic enhanced multi-modal style transfer algorithm for exhibition hall design to address the limitations of existing style transfer techniques in single dimensional feature transformation, insufficient detail preservation, and inconsistent semantic region style features. In addition, the algorithm adopts a multi-modal encoder architecture and introduces an enhanced Transformer encoder. The outcomes demonstrated that the Fréchet Inception Distance (FID) of this method was 87.9, which performed well. Sun et al. [19] introduced a novel image style transmission method based on neural network models to alleviate the problem of common neural style transmission techniques being difficult to fully transmit texture colors and prone to introducing visible errors. The core of this method is significance constraint and adopts scale invariant feature transformation. The results indicated that this method could avoid artifacts and demonstrate excellent performance by utilizing the salient feature maps (FMs) of the source image.

Some studies have also conducted field reviews and development reviews. To systematically sort out the research context of image style transfer and clarify future directions, Lian et al. [10] adopted the methods of analyzing the current research status, reviewing the development process and latest progress, and comparative analysis. The results indicated that this study clearly presented the evolution path of image style transfer from traditional methods to deep learning driven methods, and clarifies future research directions.

The advantage of GAN-based style transfer method lies in breaking through paired data dependencies and achieving efficient cross domain style transfer. Zhao [29] introduced the CycleGAN algorithm to overcome the challenges of high design costs and dependence on paired data in traditional image style transfer, and proposed a method for efficient image style transfer without the need for paired data. The results showed that the model achieved Structural Similarity Index measures (SSIM) values of 0.85, 0.84, and 0.87 in classical painting, modern art, and retro poster style transfer, respectively.

The existing style transfer methods often rely on simple normalization or linear mapping of the feature space, which makes it difficult to fully capture the texture and lighting details of complex styles. Moreover, there is a problem of transfer distortion caused by the coupling of content and style. Meanwhile, attribute editing methods have insufficient generalization ability in multi style cross domain migration, which can easily lead to artifacts or loss of details. Furthermore, most current state-of-the-art methods require adjusting model parameters for specific styles during the testing phase, resulting in high computational costs and difficulty in adapting to the efficient requirements of actual design scenarios. To address these issues, a method for image style transfer that integrates StarGANv2 and attention mechanism is proposed. Through the collaborative design of multi domain style generation architecture and spatial attention module, accurate extraction of style features and effective preservation of content structure are achieved. This design also avoids additional parameter adjustments during testing, providing a more flexible and efficient solution for visual design.

Compared with the current state-of-the-art methods, the proposed method has the following gaps: in terms of core mechanism, the Attribute Conditional

GAN (AttGAN) of attribute editing class relies on binary attribute labels to achieve feature mapping, and the Spatio Temporal Autoregressive Regression (STAR) aggregates frame level information through four memory modules. However, this study constructs a multi domain generation architecture based on StarGANv2 mapping network, combined with attention mechanism to mine key region features, avoiding the limitations of single attribute mapping or memory aggregation. In terms of applicable scenarios, AttGAN and STAR are more suitable for single attribute editing or video frame level information processing. Adaptive Instance Normalization (AdaIN) and linear style transfer excel at fast transfer of simple styles, while art fusion focuses on controllable fusion of artistic details. This study does not require adjusting model parameters during testing and can flexibly adapt to diverse needs such as multi style cross domain transfer and small sample image stylization in visual design, balancing generality and practicality. In terms of expected advantages, this study can improve style generalization ability through the flexible mapping network of StarGANv2, and enhance the distribution of differentiated features in the target domain through attention mechanism. It achieves a balance between content structure and style details, reduces computational and storage costs, and solves the problems of transfer distortion, limited generalization, and high testing costs in existing methods. The research aims to break through the limitations of existing technology, provide high-precision, high-efficiency, and low resource consumption style transfer solutions for visual design, and promote the improvement of artistic creation efficiency and quality.

Compared with existing works, the unique contribution of the research is the construction of a "dual task collaborative optimization" style transfer system. It uses StarGANv2 as a unified baseline, and improves targeted modules to simultaneously adapt to the needs of facial attribute editing and painting style transfer. This approach breaks the limitations of existing methods' single task adaptation. In terms of technical implementation, dense connections, random noise, spectral normalization and distribution shift convolution, and dual attention mechanism are hierarchically fused. The first three techniques enhance the stability of facial detail preservation and generation. The latter two techniques optimize the global consistency and local delicacy of painting style, achieving a dual improvement in effect and efficiency. In terms of application value, it can meet the diverse scene requirements of multi-style cross-domain transfer and small-sample stylization in visual design without adjusting model parameters. It solves the pain points of traditional methods that require separate modeling for different tasks and high resource consumption, providing a more universal and efficient solution for visual design.

2. Research Design

Aiming at the problem of improving the generation effect of facial images in visual design, a facial attribute style transfer method based on an improved starGANv2 model was studied and designed, introducing dense connection module and random noise module. On the basis of this model, a painting style transfer model was constructed to transfer the entire facial image to different painting styles.

2.1. Design of Facial Attribute Style Transfer Method Considering StarGANv2 Model

To improve the generation effect of facial images in visual design, a facial attribute style transfer method was first designed, and a painting style transfer model for facial images was constructed based on this method to better meet different forms of artistic creation. The core motivation for this task decomposition is the essential difference in the needs of facial attribute adjustment and painting style transfer in visual design. The former focuses on precise changes of discrete attributes such as gender and age, and needs to prioritize the consistency of facial identity features and key structures. The latter focuses on global artistic expression such as color, stroke, texture, etc., and needs to balance style purity and detail fusion.

However, the existing unified model has a performance bottleneck that is difficult to balance: the unified model needs to optimize attribute accuracy and style consistency at the same time, which is easy to cause gradient conflicts, leading to the distortion of non-target areas during attribute migration, and the destruction of facial structure during style migration. Furthermore, the unified model also faces challenges such as the difficulty of feature conflict decoupling, the surge of training complexity, and the difficulty of performance balancing. It is difficult to surpass the accuracy and efficiency of the independent model in the short term. Its potential is only reflected in reducing the repeated costs of model maintenance and deployment, which is difficult to offset the shortcomings of performance loss. After model decomposition, the structure can be optimized for different tasks, and memory usage and inference time can be reduced through individual calls. It can adapt to diverse scenarios such as batch design and real-time adjustment, and its performance advantages and application flexibility far exceed the complexity and cost of building two independent models.

The theoretical basis of task decomposition comes from the feature decoupling theory, where facial attributes rely on structured features of key facial regions and painting styles rely on global unstructured features. The two can be effectively separated through different network structures [2, 12].

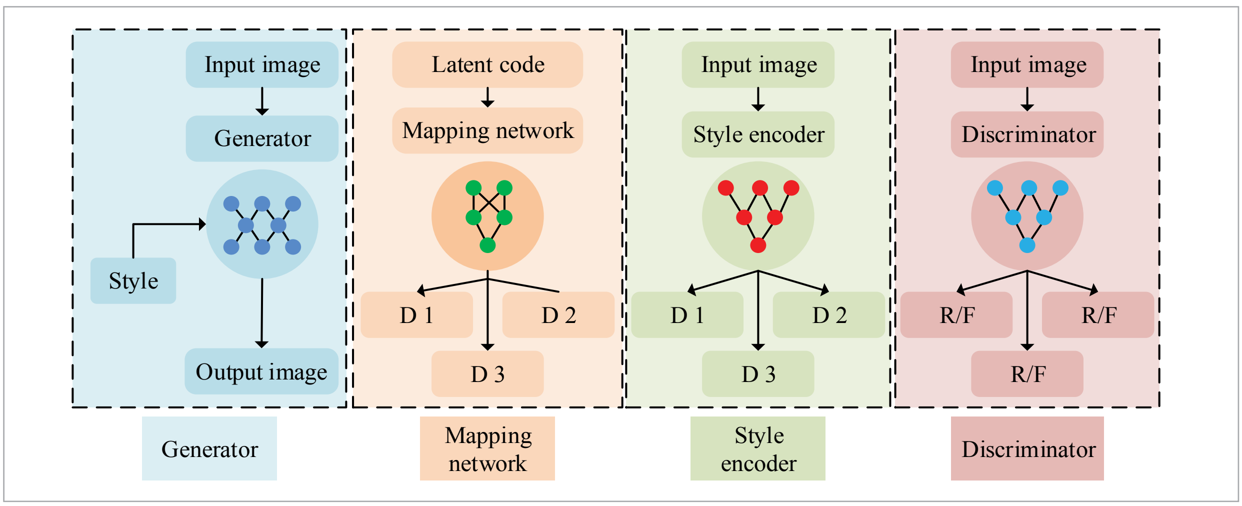
In regard to facial attribute style transfer method, the starGANv2 baseline model was adopted in the study. StarGANv2 is a single framework image to image conversion model designed to acquire the mappings among various visual domains, concurrently ensuring the diversity of the generated images and the scalability across numerous domains. It has advantages such as high diversity of generated images, strong style extraction and transfer capabilities [13-14, 18]. The structure of the starGANv2 baseline model is illustrated in Figure 1.

From Figure 1, the starGANv2 baseline model mainly consists of four modules: generator, mapping network, style encoder, and discriminator. Among them, the generator is responsible for receiving input images and encoding styles, and generating images with the target style through a series of convolution, normalization, and activation operations. The mapping network takes random vectors as input and generates style codes corresponding to different style domains. In addition, the style encoder is mainly used to guide the generation process of the generator, and the discriminator is mainly responsible for determining the authenticity of the input image and the style domain it belongs to.

However, the starGANv2 model also has certain shortcomings, such as the tendency for blurry or distorted images generated in fine facial parts. Therefore, the study made three improvements to the starGANv2 baseline model. The first improvement is the introduction of a dense connection module to enhance feature propagation and reuse, improving the quality of generating fine facial parts. The dense connection module is a core component of dense connected convolutional networks, which can alleviate the problem of gradient vanishing, enhance feature propagation, improve feature utilization, and promote deep training of the network [4, 16]. In this module, the expression of the input x_l of the l th layeris given in Equation (1).

Figure 1

Structure of starGANv2 baseline model.

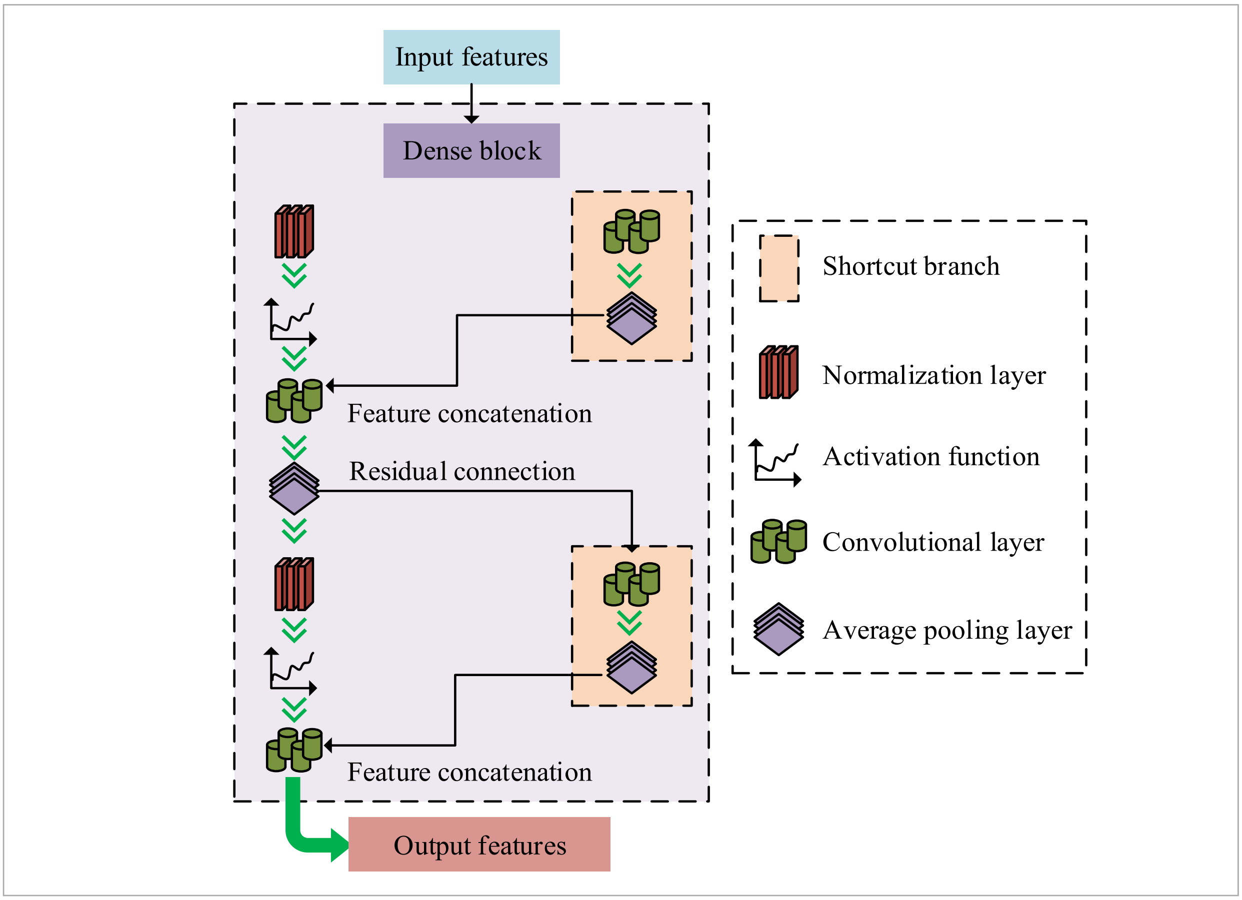


$$x_l = H_l([x_0, x_1, ..., x_{l-1}]) . \quad (1)$$

In Equation (1), H_l represents the operation of the l-th layer, such as convolution, activation function, etc. $[x_0, x_1, \ldots, x_{l-1}]$ represents the output of all preceding layers. The structure of the dense connection module is shown in Figure 2.

From Figure 2, the structure of the dense connection module includes convolutional layers, feature concatenation, shortcut branches, normalization layers, activation functions, residual connections, and average pooling layers. Specifically, the convolutional layer undertakes the task of capturing local features from the inputFM, while feature concatenation concatenates the output FMs of different layers together, thereby enhancing the feature expression ability of the model. Moreover, the Shortcut branch can prevent the loss of feature information. The residual connection adds the concatenated features to further enhance feature propagation, and finally compresses the feature dimension through convolutional layers, and outputs the fused output features of multiple rounds of information. The second improvement is the introduction of a random noise module to increase the diversity of generated images. The random noise module is a tool utilized in deep learning to enhance model robustness and generalization ability, with simple operation [6, 23]. The core mechanism of the random noise module is "controllable noise generation feature collaborative constraint", which is not simply adding noise: a dynamic noise feature map that matches the resolution of the input image is generated through bilinear interpolation algorithm, and the noise is accurately injected into the feature layer to ensure that the

Figure 2
The structure of densely connected modules.



noise intensity does not cover the effective features. The coordination between this module and the style code is achieved through dual constraints. On the one hand, the target style encoding generated by the mapping network and the style features extracted by the style encoder will jointly guide the generator to perform style consistency calibration on the feature map after noise injection. This ensures that random changes always revolve around the feature distribution of the target style. On the other hand, the style reconstruction loss and cyclic consistency loss in the loss function will filter out the variations introduced by noise, retain effective variations that match the target style, and discard invalid noise that may cause style deviation. The expression of noise parameter θ is shown in Equation (2) [3].

$$\theta = \mu + \sigma \odot \varepsilon \,. \tag{2}$$

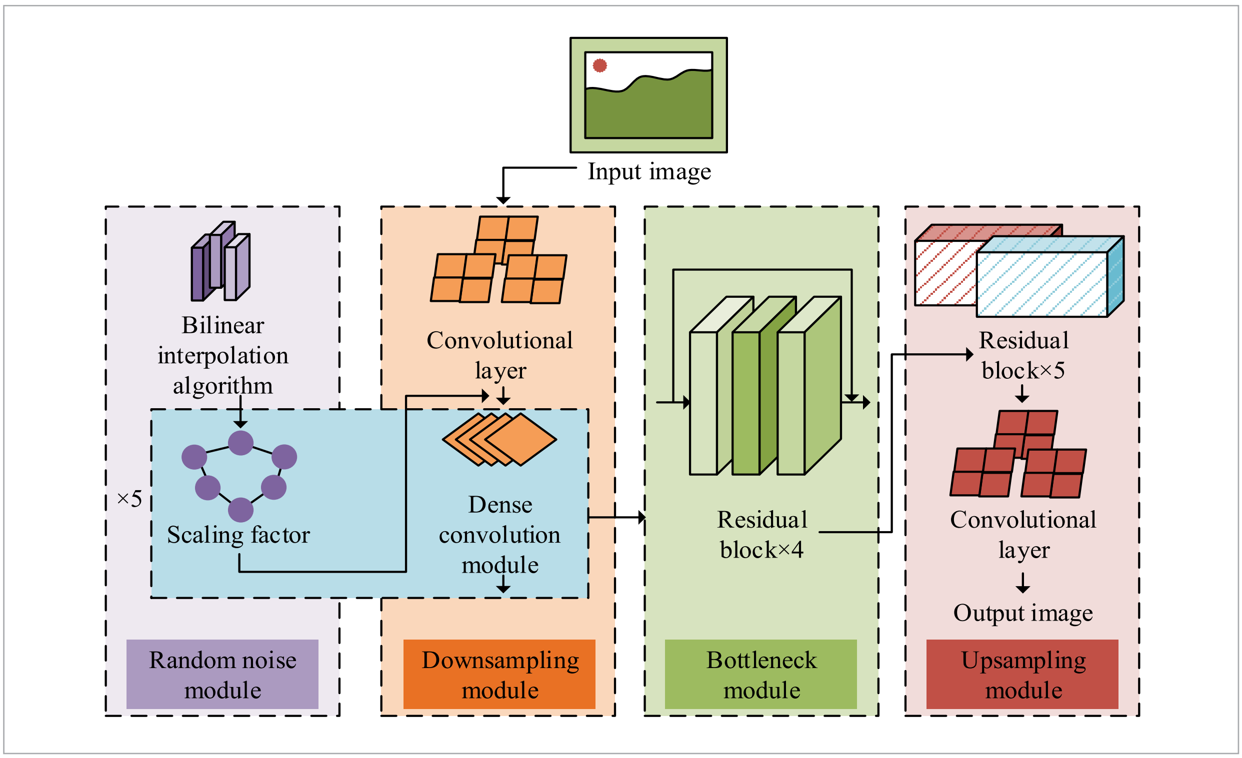
In Equation (2), both μ and σ represent learnable noise parameters, ε representsthe zero mean noise vector with fixed statistical properties. $\odot$ means element wise multiplication. The FM with noise is expressed as shown in Equation (3).

$$f' = f + \alpha n \,. \tag{3}$$

In Equation (3), f represents the initial FM, α is the noise intensity coefficient, and n represents the generated random noise. Through this technology, it is possible to inject random noise into the FMs of starGANv2 generators, enabling the generator to generate more variations in the generated images, thereby enhancing the diversity of the generated images. Both improvement one and improvement two are carried out in the generator of the starGANv2 baseline model, so the improved generator structure is shown in Figure 3.

From Figure 3, the improved generator includes a random noise module, multiple dense convolution modules, downsampling module, upsampling module, and bottleneck module. Specifically, the random noise module generates noise through bilinear interpolation algorithm and injects dynamic perturba-

Figure 3
Improved generator structure.



tions into the network by combining scaling factors to enhance the diversity of generated images. The downsampling module first uses convolutional layers to extract features, and then uses multiple dense convolutional modules to enhance feature reuse and detail preservation, and compress the spatial dimension of the image. Furthermore, the bottleneck module is composed of residual blocks to further extract abstract features. The upsampling module uses residual blocks to restore the spatial resolution of the image and outputs the style transferred image through convolutional layers.

The third improvement is the introduction of a spectral normalization network in the discriminator to achieve 1-Lipschitz constraints and enhance the stability of the discriminator. Spectrum normalization network is a type of generative adversarial network that uses spectral normalization techniques to stabilize the training of discriminators and prevent them from becoming too powerful [9, 21]. Specifically, the spectrum normalization network normalizes the weight matrix of the discriminator to ensure that the spectral norm of the weight matrix does not exceed 1, thereby satisfying the 1-Lipschitz condition. In this network, the definition of spectral norm $\sigma(W)$ is shown in Equation (4) [5].

$$\sigma(W) = \max_{\|y\|=1} \|Wy\| . \tag{4}$$

In Equation (4), W is the weight matrix and $\|y\|$ is the norm of the vector y. The expression of the weight matrix W' after spectrum normalizationis shown in Equation (5).

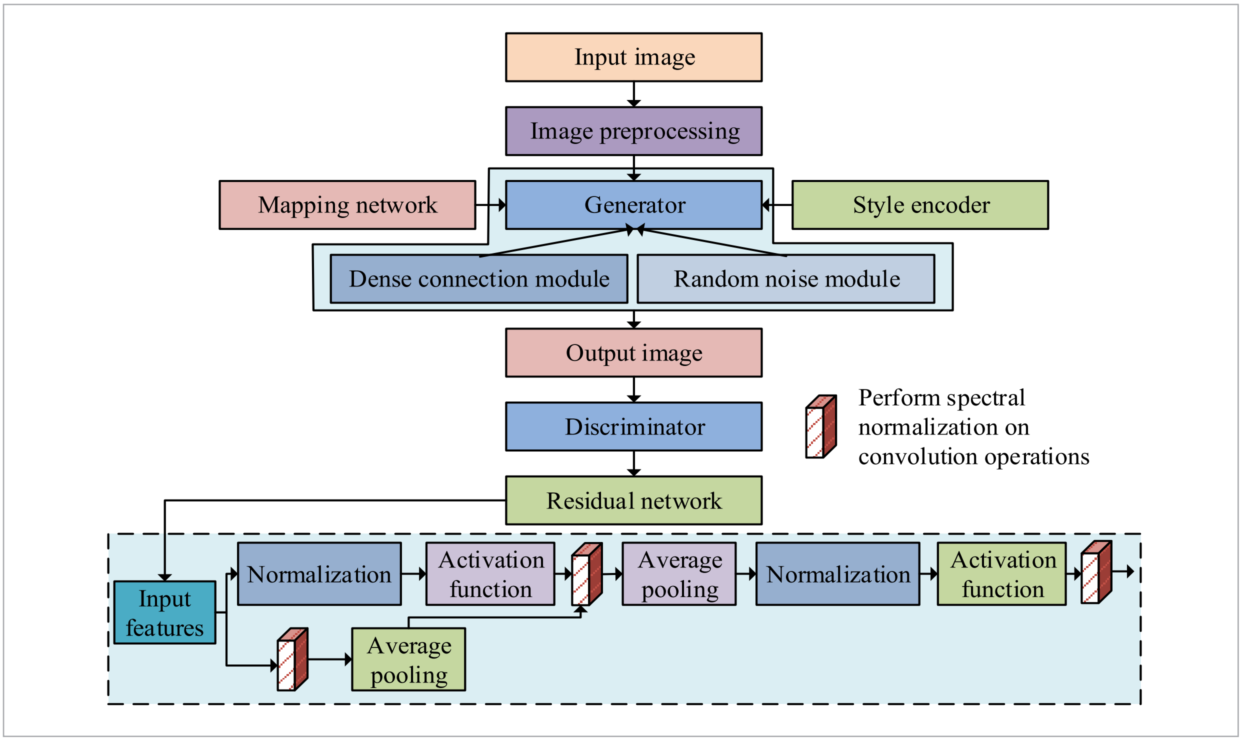
$$W' = \frac{W}{\sigma(W)} . \tag{5}$$

The structure of the facial attribute style transfer method is shown in Figure 4.

From Figure 4, the input image first undergoes image preprocessing steps, including face pose correction, face cropping, and image scaling, to ensure the consistency and effectiveness of the input image. The preprocessed image will enter the generator, which contains dense connection modules and ran-

Figure 4

Structure of facial attribute style transfer method.



dom noise modules for generating facial images with target attributes. Moreover, the mapping network receives 16 dimensional random vector z and generates style code , while the style encoder extracts style code s from the input dependent images, and these two style code s' together guide the generator's generation process. Finally, the introduction of spectral normalization discriminator will discriminate the images generated by the generator and continuously optimize the performance of the generator through adversarial training. The loss function L_A of the facial attribute style transfer method is shown in Equation (6).

$$L_A = L_p + \lambda_q L_q - \lambda_r L_r + \lambda_s L_s \, . \quad (6)$$

In Equation (6), L_p represents generative adversarial loss, L_q represents style reconstruction loss, L_r represents style diversification loss, and L_s represents cyclic consistency loss. λ_q, λ_r, and λ_s are the weight coefficients of L_q, L_r and L_s, respectively.

2.2. Construction of Painting Style Transfer Model Tak-ing into Account Distribu-tion Shift Convolution

The study has designed a facial attribute style transfer method based on an improved starGANv2 model. However, in visual design, it is also necessary to transfer the entire facial image to different painting styles, such as oil painting, sketching, cartoons, etc. The face attribute style transfer method is difficult to directly meet the needs of painting style transfer, because painting style transfer needs to consider the overall style features of the image, including color, stroke, texture, etc., rather than just the attributes of the face. Therefore, it is necessary to construct a specialized model for transferring painting styles. In the construction of the painting style transfer model, the starGANv2 baseline model was also used in the study, and another improvement was made to it that is more suitable for painting style transfer, mainly focusing on the generator and discriminator. The first improvement is the introduction of distributed shift convolution in the generator to reduce the number of model parameters, accelerate training speed, and ensure the quality of generated images. Distributed shift convolution is an improved convolution operation that performs distributed shift processing on the convolution kernel, which can reduce the number of parameters in the convolution kernel while maintaining the receptive field of the convolution operation, thereby improving computational efficiency [8, 25]. The output of the distributed shift convolution is shown in Equation (7) [7].

$$Z = \sum_i K_i \, \hat{A} D_{t_i} \quad (7)$$

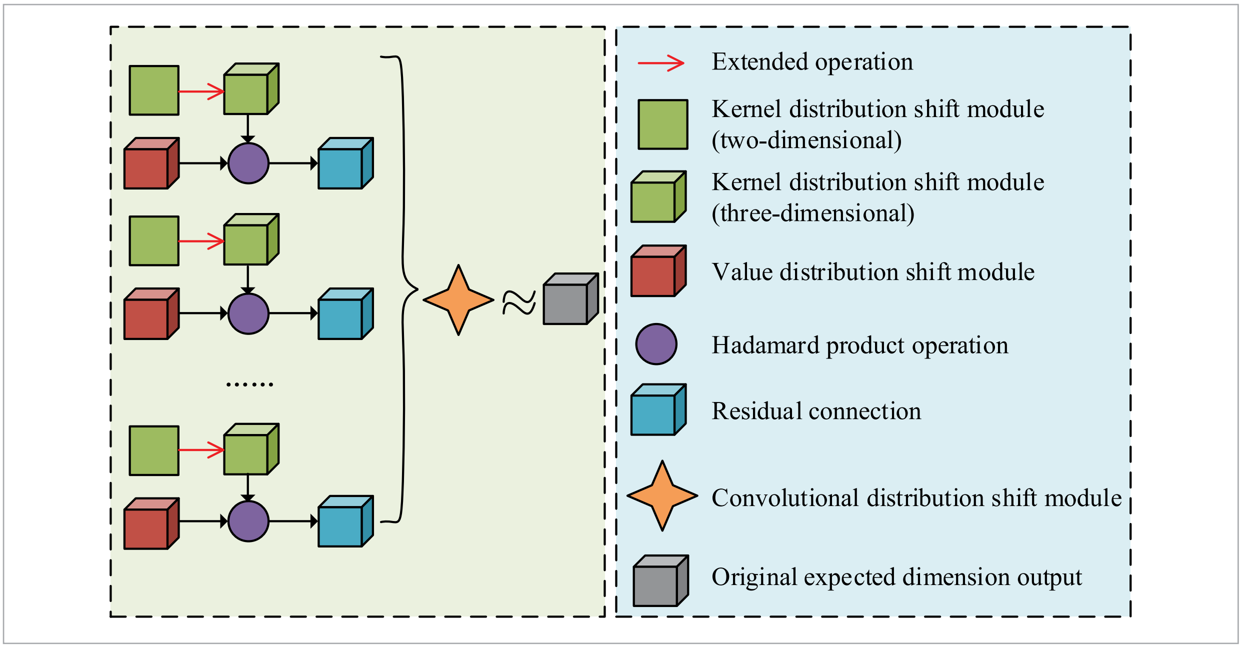
In Equation (7), Z is the output of the distribution shift convolution, $\hat{A}$ represents the convolution operation, t_i represents the shift amount of the sub kernel, and K is the traditional convolution kernel. K_i represents the i-th sub kernel, D is the input FM. The structure of distributed shift convolution is shown in Figure 5 [20].

From Figure 5, distribution shift convolution includes expansion operation, kernel distribution shift module, value distribution shift module, residual connection, Hadamard product operation, and convolutional distribution shift module. Among them, the kernel distribution shift module is responsible for simulating the distribution changes of convolutional kernels and performing expansion operations to simulate different shift amounts, thereby enhancing the robustness of the model to input feature translation. The value distribution shift module is responsible for simulating the distribution changes of input feature values, while the residual module is used to alleviate the problem of gradient vanishing. In addition, the Hadamard product operation is responsible for multiplying the convolution results of different sub kernels element by element to simulate the convolution effect of different sub kernels at different positions. Finally, the convolution distribution shift module compares the Hadamard product processed results with the original convolution results to simulate the adaptability of convolution operations under different distributions, and outputs the processed FMs.

The second improvement is the introduction of attention mechanisms in the generator, namely channel attention (CA) module and spatial attention (SA) module, to enable the generator to place greater emphasis on crucial areas within the image and enhance the effect of painting style transfer. The CA module and SA module will respectively weight the

Figure 5

The structure of distributed shift convolution.



channels and spatial positions of the FM. The FM F' after CA processingis shown in Equation (8).

$$F' = F \times M_c(F) = F \times \left\{ J(R(F_{avg}^c) + R(F_{max}^c)) \right\}. \quad (8)$$

In Equation (8), F is the input FM, J represents the sigmoid activation function, and R represents the multilayer perceptron. F_{avg}^c represents the channel descriptor obtained by global average pooling, and F_{max}^c is the channel descriptor obtained by global maximum pooling. $M_c(F)$ is channel weight. The FM F'' after SA processingis shown in Equation (9).

$$F'' = F' \times M_s(F') = F' \times \left\{ J(f^{7\times7}([F_{avg}^s; F_{max}^s])) \right\}. \quad (9)$$

In Equation (9), $M_s(F')$ is the spatial weight, $f^{7\times7}$ represents the 7×7 convolution operation. F_{avg}^s and F_{max}^s are the spatial descriptors obtained by global average pooling and global maximum pooling, respectively. Therefore, in the improved generator, the input image is first processed by the initial convolutional layer, and then enters multiple distribution shifted convolution residual blocks for feature extraction. In the process of feature extraction, the FM is weighted using CA module and SA module to highlight important feature information. The relationship between attention module and painting style transfer is closely related and targeted. The channel attention module focuses on the core elements of color distribution in the style, and through weight calibration of feature channels, strengthens the channel response that characterizes the proportion of target style tones and color saturation, ensuring global consistency of color style. The spatial attention module focuses on local style elements related to texture and shape. By generating spatial weight heat maps, it accurately locates and enhances key areas such as stroke direction, texture density, and contour lines, ensuring the delicate restoration of local style details. Both anchor the global color and local texture and shape features of style transfer, forming a dual dimensional collaborative capture mechanism. The various modules of the generator work together to reduce the quantity of parameters while selectively extracting and enhancing important features of the image, laying the foundation for high-quality painting style transfer.

The third improvement is the introduction of spectral normalization network and Markov discriminator in the discriminator to enhance its ability to distinguish the style of generated images and ensure consistency between the generated images and the target painting style. Markov discriminator is a classification model based on Markov properties, which has the advantage of strong locality and can better capture the detailed information of images [1, 17, 22]. Therefore, research is being conducted on using Markov discriminators to distinguish local regions in images, in order to better capture the local style features of the image and improve the performance of the discriminator. The improved discriminator structure is illustrated in Figure 6.

From Figure 6, the discriminator is mainly composed of a spectrum normalization discriminator and a Markov discriminator. The input image will be simultaneously fed into two discriminators, where the spectral normalization discriminator will judge the authenticity and style domain of the image from a global perspective, while the Markov discriminator will judge from a local perspective. The output results of the two discriminators are used together to guide the training of the generator, so that the images generated by the generator can conform to the target painting style both globally and locally. The dual discriminator operates in a collaborative mode of "global control local refinement", without a dominant relationship. The two output complementary outputs to jointly optimize generator training, and its core collaborative mechanism is reflected in two aspects: division of labor positioning and interaction logic. In terms of division of labor, the spectral normalization discriminator focuses on the global dimension and is mainly responsible for determining the overall authenticity, style domain matching, and global style consistency of the generated image, ensuring that the generated image conforms to the macro features of the target painting style. The Markov discriminator relies on the strong locality advantage of Markov properties, focuses on local regions of the image, compensates for the shortcomings of the spectral normalization discriminator in local detail judgment, and focuses on optimizing the local style purity and detail fusion of the generated image. In terms of collaborative logic, during the training process, the two synchronously receive the images output by the generator and calculate the global loss and local loss separately. Moreover, the loss value feedback of the spectrum normalization discriminator is used to guide the generator to adjust the distribution of global style features, making it closer to the real style domain. The loss value feedback of the Markov discriminator guides the generator to optimize the style expression of local details, avoiding local texture distortion or style deviation. The structure of the painting style transfer model is illustrated in Figure 7.

Figure 6

Improved discriminator structure.

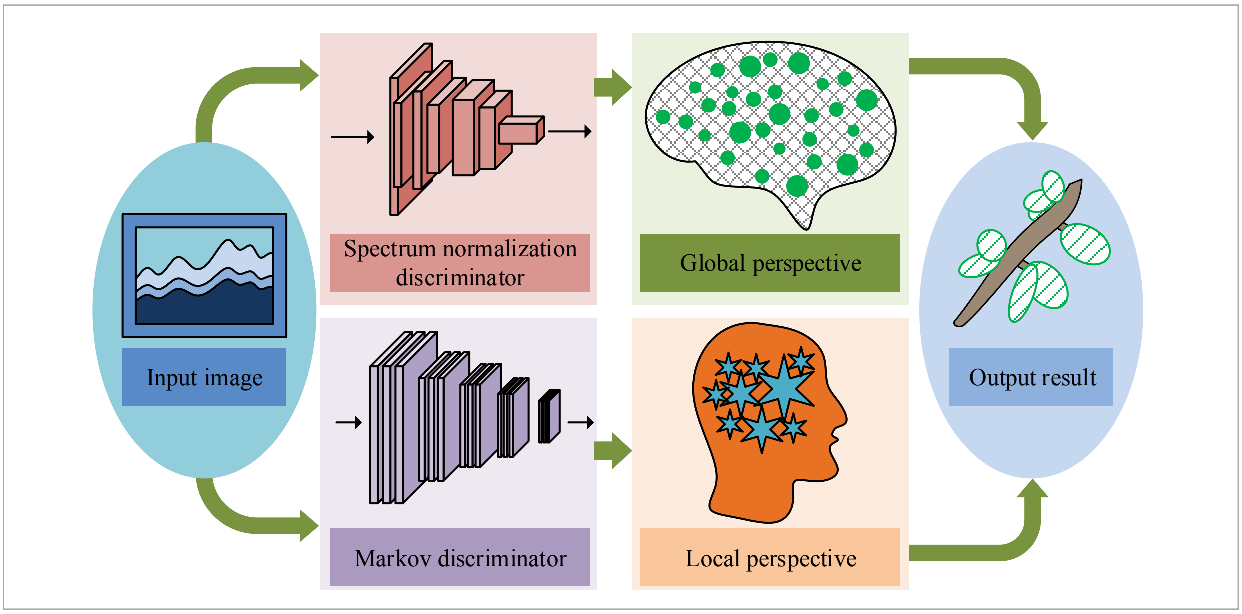
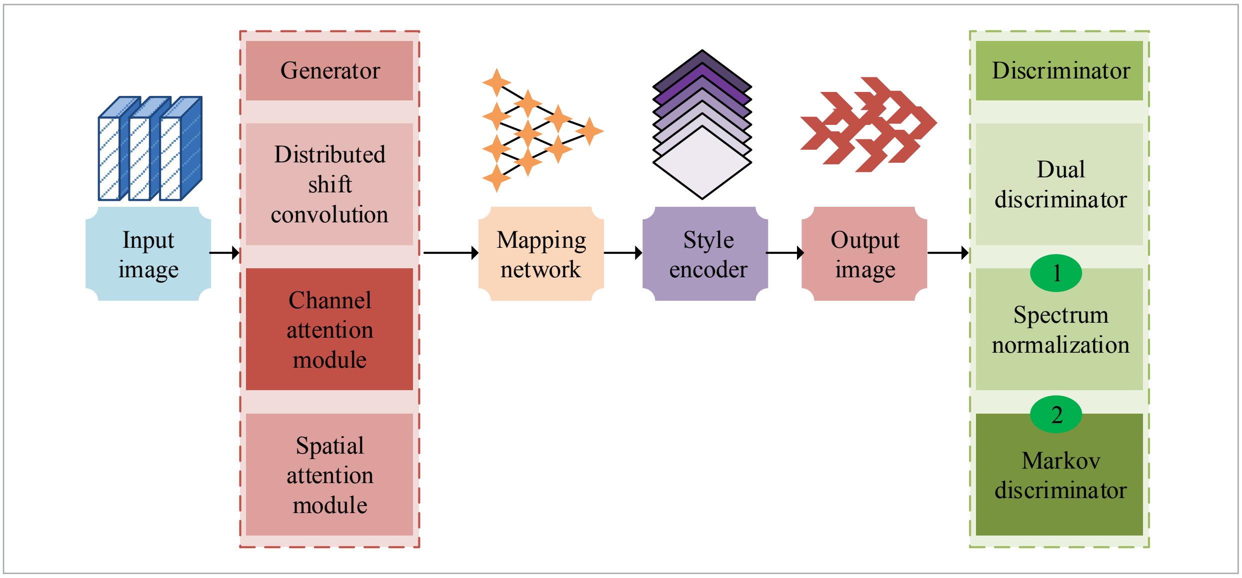


Figure 7

The structure of painting style transfer model.



From Figure 7, the structure of the painting style transfer model mainly includes input images, mapping networks, generators, style encoders, output images, and discriminators. Among them, the input image will undergo preprocessing before being input into the generator. The generator utilizes distribution shift convolution, CA module, and SA module to generate images with the target painting style. The generation process of the generator relies on the style encoding generated by the mapping network and the style encoding extracted by the style encoder. Afterwards, the discriminator uses a dual discriminator to discriminate the generated images, in order to improve the quality and style consistency of the generated images. The loss function of the painting style transfer model is shown in Equation (10).

$$L_B = L_p + \lambda_q L_q - \lambda_r L_r + \lambda_u L_u \, . \quad (10)$$

In Equation (10), L_u is the cross domain consistency content loss. λ_u is the weight coefficient of L_u.

3. Results and Analysis

To validate the performance of the style transfer method used in the research design, the experimental dataset was described and the experimental environment was set up. Subsequently, a comparative model was selected for the study, and the parameters of the model were set. In addition, the study also provided explanations for the evaluation indicators.

3.1. Performance Verification of Facial Attribute Style Transfer Method for Visual Design

To validate the performance of the face attribute style transfer method for visual design, a large-scale Celebfaces Attributes dataset (CelebA) was used in the study [27, 32]. This dataset is a large-scale facial attribute dataset broadly applied in computer vision research, and contains a total of 202599 facial images, covering 40 attribute annotations such as gender, age, and hairstyle. The study selected 20000 images from the data and preprocessed them, including facial keypoint pose correction, cropping of facial regions, scaling, and standardization. Subsequently, the study divided the data into training and testing sets at a ratio of 7:3. The study used the Windows 10 operating system with an Intel Core i7-11700K central processor, which has a clock speed of 3.60GHz and a turbo frequency of up to 5.00GHz. In terms of comparative models, the study selected a combination model A that combines the original starGANv2, CycleGAN, conditional generative adversarial network, and StyleShot, and a combination model B that com-

Table 1
Experimental results of ablation of facial attribute style transfer method.

starGANv2	Dense connection module	Random noise module	Spectrum normalization network	Evaluation metrics		
				FID	Time consuming	AGC
√	√	√	√	51.27	68ms	98.56%
√	√	√	×	55.68	72ms	95.14%
√	√	×	×	57.46	83ms	93.37%
√	×	×	×	60.32	108ms	89.16%

bines Transformer and residual network. In terms of parameter settings, the learning rate of the models was 0.001, the number of iterations was 300, and the λ_q, λ_r and λ_s of face attribute style transfer method are 10, 0.5 and 10, respectively. In terms of evaluation indicators, the study selected FID, Attribute Generation Accuracy (AGC) and time consumption. The ablation experiment results of the facial attribute style transfer method are presented in Table 1.

It could be observed from Table 1 that when all three improved modules were incorporated into starGANv2, the FID reached 51.27, the processing time was 68 ms, and the AGC attained 98.56%, indicating the best performance. After the spectral normalization network was removed, the FID rose to 55.68, the AGC dropped by 3.42%, and the processing time increased by 4 ms, demonstrating that this network could enhance the quality and attribute accuracy of generated images. When the random noise module was further removed, the FID increased to 57.46, the AGC decreased to 93.37%, and the processing time rose by 15 ms, suggesting that the random noise module had a positive impact on the diversity and efficiency of generated images. When only the basic model was retained, the FID reached 60.32, the AGC was merely 89.16%, and the processing time was 108 ms, which also verified the synergistic effect of the three improved modules in enhancing the quality, accuracy, and efficiency of image generation.

To further confirm the capability of the research-designed face attribute style transfer method, the research collected a total of 2,000 facial image datasets required for visual design applications, preprocessed them, and then applied different models for style transfer. A comparison of FID and processing time for various models in practical applications is shown in Table 2.

From Table 2, in practical applications, the research-designed face attribute style transfer method demonstrated the best performance. Specifically,

Table 2
Comparison of FID and time consumption of different models in practical applications.

Method	FID					Time consuming/ms				
	Experiments					Experiments				
	1	2	3	4	5	1	2	3	4	5
starGANv2	63.56	63.82	63.31	63.69	63.56	112	113	111	118	112
CycleGAN	68.27	68.53	68.01	68.41	68.29	145	146	142	147	145
Combination model A	72.15	72.41	71.92	72.3	72.18	163	164	162	161	163
Combination model B	65.89	66.15	65.63	66.02	65.9	125	126	124	127	125
Facial attribute style transfer meth-od	52.13	51.92	52.31	51.87	52.05	71	70	72	69	74

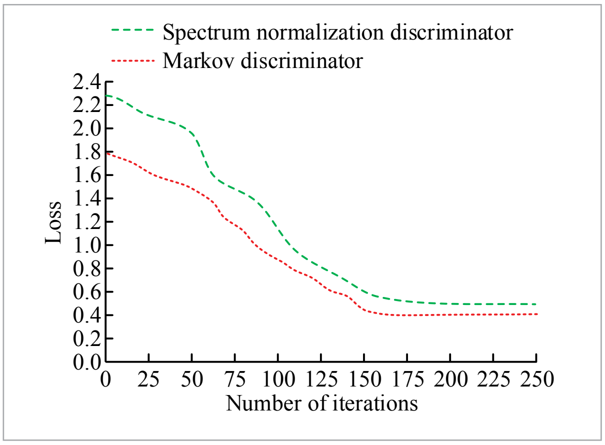
regarding the FID metric, the average value of the research-designed method was 52.06, while those of starGANv2, CycleGAN, Combined Model A, and Combined Model B were 63.59, 68.30, 72.19, and 65.92, respectively. Additionally, in terms of processing time, the values for the research-designed method remained stable within the range of 69-74 ms, with an average of approximately 71.2 ms, whereas the average processing times for the other four comparative models were 113.2 ms, 145.0 ms, 162.6 ms, and 125.4 ms, respectively. It was evident that the research-designed method incurred shorter processing times. In summary, the research-designed face attribute style transfer method exhibited superior performance. This could be attributed to the fact that, through the densely connected module, the method constructed a more efficient feature transfer pathway, enabling it to more precisely capture key facial attributes and reduce feature loss and erroneous transfer. Meanwhile, the spectral normalization network stabilized discriminator training, allowing the generator to rapidly converge to a better solution and reducing iteration time.

3.2. Performance Validation of Painting Style Transfer Model for Visual Design

To validate the performance of the visual design-oriented painting style transfer model, the same operating coefficients and central processing unit were used in the study. On the dataset, the study also used the CelebA dataset and additionally introduced the Painter by Numbers style dataset. Painter by Numbers is a dataset for art style recognition and analysis, consisting of 30000 images and featuring three styles: oil painting, sketching, and cartoon. Therefore, the final dataset consisted of 50000 images, which were also preprocessed and divided into datasets, and the style images were additionally color standardized. In terms of comparative models, the stu dy selected starGANv2, Domain Constrained Transformation Network (DCTN), a combination model C that combines a Gram matrix-based style transfer model and a multi-modal unsupervised image to image conversion model, and a combination model D that combines a dual representation image conversion model and a style transfer Transformer network. In terms of model parameter settings, the learning rate was 0.0001, the number of iterations was 250, and the λ_q, λ_r and λ_u of the painting style transfer model are 0.5, 10 and 5, respectively. The changes in the loss values of the two discriminators in the dual discriminator during the training process are shown in Figure 8.

Figure 8

Changes in loss values of two discriminators during training in a dual discriminator.



From Figure 8, in the dynamic interaction between the spectrum normalization discriminator and the Markov discriminator, in the early stage of training (the first 50 rounds), the loss value of the spectrum normalization discriminator was higher than that of the Markov discriminator. At this time, the global style feature was not yet stable, and the loss feedback of the spectrum normalization discriminator was relatively higher, giving priority to guiding the generator to establish a global style framework. As the training progressed (50-150 rounds), the loss value of the spectrum normalization discriminator rapidly decreased to around 0.6, while the loss value of the Markov discriminator slowly decreased to around 0.5. The gap between the two gradually narrows, and the generator began to focus on local detail optimization on the basis of global style stability. In the later stage of training (150-250 rounds), the loss values of the two tended to stabilize and the changing trends were highly synchronized. This indicated that the global style and local details formed a collaborative optimization state, and the generator could simultaneously meet the requirements of global style consistency and local detail fidelity. This dynamic interaction process achieved a training logic of "global first, local second, and then collaborative optimization" through complementary changes in loss values. It not only ensured the overall style accuracy of the

Table 3

Experimental results of ablation the transfer model of painting style.

starGANv2	Distributed shift convolution	Attention mechanism	Dual discriminator	Evaluation metrics		
				FID	Time consuming	Memory usage
√	√	√	√	28.59	120ms	10.32%
√	√	√	×	32.18	115ms	13.66%
√	√	×	×	36.75	108ms	15.27%
√	×	×	×	41.21	95ms	22.58%

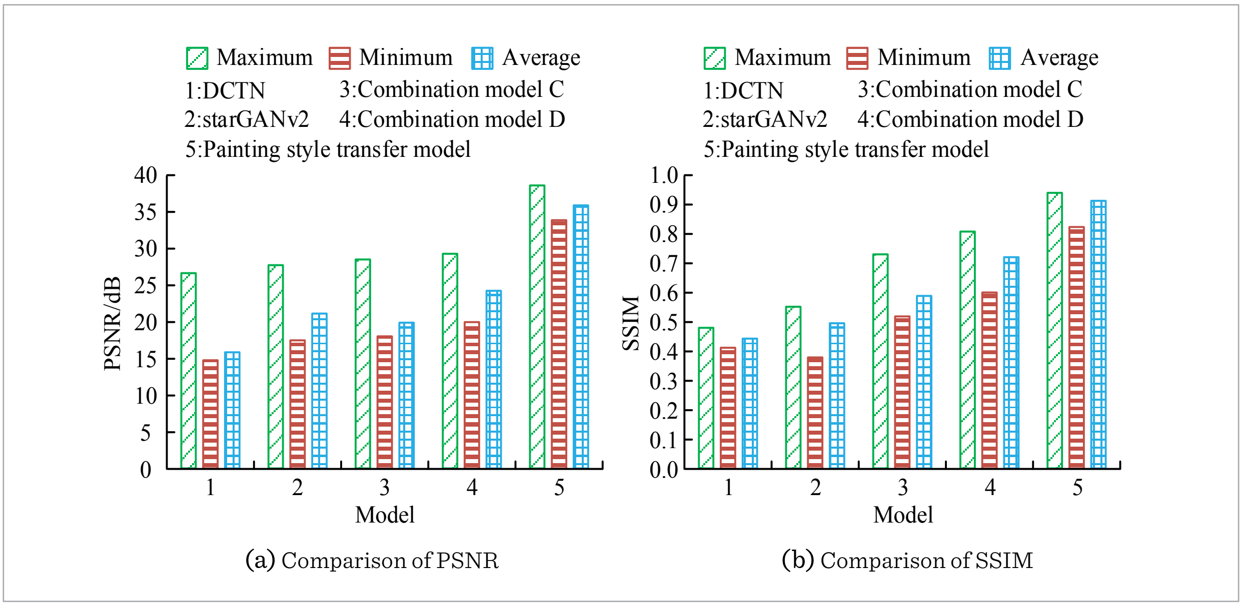
generated image, but also improved the delicacy of local details, ultimately significantly enhancing the comprehensive effect of painting style transfer. The results of the ablation experiment of the painting style transfer model are shown in Table 3.

From Table 3 when starGANv2 activated all the improved modules, namely the distribution-shifting convolution, the attention mechanism, and the dual-discriminator module, it achieved an FID of 28.59, a processing time of 120 ms, and a memory usage rate of 10.32%, demonstrating relatively good performance. When any one of these improved modules was removed, the model's FID, processing time, and memory usage rate all increased. For instance, when the dual-discriminator module was removed, the FID rose to 32.18, the processing time increased to 115 ms, and the memory usage rate climbed to 13.66%. This indicated that the collaborative operation of all modules enabled the model to generate images that more closely resembled the real style distribution while utilizing memory more efficiently. This could be attributed to the fact that the distribution-shifting convolution optimized the transfer of style features, the attention mechanism directed the model's focus to key style regions, and the dual-discriminator constrained the quality of the generated images.

A comparison of Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) for different models is shown in Figure 9.

Figure 9

Comparison of PSNR and SSIM among various models.



From Figure 9(a), in terms of PSNR, the painting style transfer model designed in the study still demonstrated better performance, followed by Combined Model D, Combined Model C, and the StarGAN model, with DCTN coming last. Regarding specific values, the maximum PSNR values of the five models were 38.68 dB, 29.39 dB, 28.11 dB, 27.85 dB, and 26.54 dB, respectively. According to Figure 9(b), in terms of SSIM scores, the painting style transfer model designed in the study also held an advantage, with its maximum SSIM reaching 0.948, while the maximum SSIM values of the other comparative models were all below 0.9. In summary, in terms of PSNR, the maximum value of the painting style transfer model designed in this study was 9.29dB, 10.57dB, 10.83dB, and 12.14dB higher than that of the combination model D, combination model C, StarGAN model, and DCTN. This may be because the painting style transfer model designed in this study could effectively retain the feature information of images through distribution-shifted convolution, thereby resulting in generated images with minimal differences from the original images at the pixel level. Moreover, the attention mechanism it employed could focus on preserving key information during the style transfer process. To further validate the performance of the research design painting style transfer model, the study introduced the current optimal technology, which included AdaIN, Style-Attentional Network (SANet), Mamba-GAN (MauGAN), and added the original StarGANv2. The SSIM and time consumption comparison of different models in practical applications are shown in Table 4.

From Table 4, in terms of SSIM indicators, the research-design painting style transfer model performed the best, with SSIM values above 0.947, far higher than the original starGANv2 (0.866-0.882), AdaIN (0.843-0.856), SANet (0.886-0.893), and MauGAN (0.895-0.901) models. This indicated that the images generated by the model had stronger structural consistency with the target style. In terms of time consumption, AdaIN had the fastest inference speed, but SSIM performed the worst. The time interval for studying and designing a painting style transfer model was 119-123ms, which was slightly higher than AdaIN but significantly lower than the original starGANv2, SANet, and MauGAN. In summary, the research-design painting style transfer model achieved a good balance between SSIM (style transfer effect) and time consumption (inference efficiency). Compared with the current optimal technical model, it not only ensured higher quality of style transfer, but also had higher operational efficiency and overall performance.

4. Conclusion

To address the issue of improving the generation effect of facial images in visual design, this study uses starGANv2 as the baseline model and designs facial attribute style transfer methods and painting style transfer models through module improvement and structural optimization. The core scientific discovery is that the feature decoupling and dual module collaborative mechanism can effectively break

Table 4

Comparison of SSIM and time consumption of different models in practical applications.

Method	SSIM					Time consuming/ms				
	Experiments					Experiments				
	1	2	3	4	5	1	2	3	4	5
starGANv2	0.882	0.875	0.868	0.871	0.866	156	158	154	159	157
AdaIN	0.856	0.849	0.853	0.847	0.843	78	80	79	77	81
SANet	0.893	0.887	0.891	0.889	0.887	132	134	131	133	135
MauGAN	0.901	0.896	0.899	0.900	0.895	148	150	147	149	151
Painting style transfer model	0.948	0.951	0.947	0.949	0.950	120	122	119	121	123

through the bottleneck of balancing "attribute accuracy style purity inference efficiency" in StarGANv2. The combination of dense connections and spectral normalization enhances the preservation of facial structural features, the channel/spatial attention mechanism achieves dimensional capture of style elements, and the dual discriminator constructs a "global local" closed-loop optimization logic. This discovery provides a reusable technical paradigm for modular design of cross task visual generation.

The results showed that in the face attribute transfer task, the improved method achieved a FID of 51.27 when all modules were enabled in the ablation experiment on the CelebA dataset, which was nearly 9.05 lower than the 60.32 obtained by retaining only the starGANv2 base model. This indicated that the improvements made by the research were effective. The average FID of this method in practical applications was 52.06, much lower than the original starGANv2's 63.59. In the painting style transfer task, the improved model performed better in the ablation experiment with a full module startup corresponding FID of 28.59, a time of 120ms. Furthermore, the model's PSNR and SSIM metrics outperformed those of the baseline models. Both the facial attribute style transfer method and the painting style transfer model designed for research had good performance. The research significance of this achievement lies in providing an integrated solution of "attribute controllability style fidelity efficient deployment" for digital art creation, film and television character design and other scenarios, reducing the technical threshold of AI-assisted visual design, and verifying the effectiveness of dynamic constraint mechanisms in feature decoupling.

However, there are obvious limitations to the research: firstly, the dataset style coverage was limited and did not cover niche styles such as minimalism and ukiyo-e, resulting in insufficient generalization of the model in extreme style differences scenarios. Future research can construct a multi-source dataset that integrates WikiArt and PainterbyNumbers, incorporating over 50 cross genre art style samples, expanding training data with federated learning techniques, and introducing a dynamic domain adaptation module to optimize the style transfer generalization ability of StarGANv2. Secondly, there is a lack of precise semantic control, making it difficult to adapt styles to specific facial features. Future research can combine ControlNet control network with facial parser to achieve style transfer in facial features such as eyebrows and lips, while integrating Contrastive Language Image Pre training text guidance module to support local style customization under natural language instructions. The third issue is that the model parameter count and computational cost were relatively high, making it difficult to meet real-time application requirements when deployed on low computing power devices such as mobile devices. In future research, knowledge distillation and quantitative perception training techniques can be used, combined with lightweight feature extractors such as MobileNet to replace traditional dense layers, compressing the model parameter count by more than 60% and adapting to real-time design scenarios on the end side.

References

1. Cai, W., Jiang, B., Jiang, X., Yang, Z. Infra-red Image Generation with Unpaired Train-ing Samples. Optics and Precision Engineer-ing, 2023, 31(24), 3651-3661. https://doi.org/10.37188/OPE.20233124.3651
2. Chen, J. The Application of Artificial Intel-li-gence in Comic Creation: Taking the Exag-ger-ation Method of Face Images as an Ex-ample. Discover Artificial Intelligence, 2025, 5(1), 1-15. https://doi.org/10.1007/s44163-025-00563-8
3. Groumpos, P. P. A Critical Historic Over-view of Artificial Intelligence: Issues, Chal-lenges, Opportunities, and Threats. Artifi-cial Intelligence Advances, 2023, 1(4), 197-213. https://doi.org/10.47852/bonviewAIA3202689
4. Ji, Z., Bei, J., Liu, R., Ren, T. Dual-Channel Vision Transformer-Based Image Style Transfer. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(7), 2488-2497. DOI: 10.13700/j.bh.1001-5965.2023.0392.

5. Lee, S. H., Kim, S., Byeon, W., Oh, G., In, S., Park, H. Audio-Guided Implicit Neural Representation for Local Image Stylization. Computational Visual Media, 2024, 10(6), 1185-1204. https://doi.org/10.1007/s41095-024-0413-5
6. Li, H., Wang, L., Liu, J. A Review of Deep Learning-Based Image Style Transfer Re-search. Imaging Science Journal, 2025, 73(4), 504-526. https://doi.org/10.1080/13682199.2024.2418216
7. Li, Q., Wu, M., Chen, D. PhotoGAN: A Novel Style Transfer Model for Digital Pho-tographs. Computers, Materials and Con-tinua, 2025, 83(6), 4477-4494. https://doi.org/10.32604/cmc.2025.062969
8. Li, X., Pu, Y. Y., Zhao, Z. P., Xu, D., Qian, W. H. Content Semantics and Style Features Match Consistent Artistic Style Transfer. Journal of Graphics, 2023, 44(4), 699-709. DOI: 10.11996/JG.j.2095-302X.2023040699.
9. Li, Y., Zhu, J., Liu, X., Chen, J., Su, X. Anime Image Style Transfer Algorithm Based on Improved Generative Adversarial Networks. Journal of Beijing University of Posts and Telecommunications, 2024, 47(4), 117-123. DOI: 10.13190/j.jbupt.2023-144.
10. Lian, L., Tian, Q., Tan, R., Zhang, X. Re-search Progress of Image Style Transfer Based on Neural Network. Computer Engi-neering and Applications, 2024, 60(9), 30-47. DOI: 10.3778/j.issn.1002-8331.2309-0204.
11. Lu, D., Wu, Y., Acar, A., Yao, X., Wu, J. Y., Kavoussi, N. ASSIST-U: A System for Seg-mentation and Image Style Transfer for Ureteroscopy. Healthcare Technology Let-ters, 2024, 11(2), 40-47. https://doi.org/10.1049/htl2.12065
12. Lu, W., Wang, J., Jin, X., Jiang, X., Zhao, H. FACEMUG: A Multimodal Generative and Fusion Framework for Local Facial Editing. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(9), 5130-5145. https://doi.org/10.1109/TVCG.2024.3434386
13. Ma, Z., Lin, T., Li, X., Li, F., He, D., Ding, R. Dual-Affinity Style Embedding Network for Semantic-Aligned Image Style Transfer. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(10), 7404-7417. https://doi.org/10.1109/TNNLS.2022.3143356
14. Ponamaryov, V. V., Kitov, V. V. Automatic Image Style Transfer Using an Augmented Style Set. Programming and Computer Software, 2024, 50(3), 231-237. https://doi.org/10.1134/S0361768824700038
15. Qin, C., Dong, T. L., Yao, H. Constructive Data Hiding Based on Texture Synthesis and Recognition with Image Style Transfer. Journal of Software, 2023, 34(12), 5773-5786. DOI: 10.13328/j.cnki.jos.006752.
16. Rao, S., Qian, W., Zhang, J. Image Style Transfer via Curved Stroke Rendering. Journal of Beijing University of Technology, 2023, 28(12), 3825-3837. https://doi.org/10.11834/jig.221150
17. Su, Z., Zhao, S., Zhang, H., Li, P., Lu, Y. Digital Printing Image Generation Method Based on Style Transfer. Textile Research Journal, 2023, 93(23), 5211-5223. https://doi.org/10.1177/00405175231195367
18. Sun, Y., Liu, J., Zhang, R. Large Capacity Generative Image Steganography via Image Style Transfer and Feature-Wise Deep Fu-sion. Applied Intelligence, 2023, 53(23), 28675-28693. https://doi.org/10.1007/s10489-023-04993-8
19. Sun, Y., Xie, X., Li, Z., Zhao, H. Image Style Transfer with Saliency Constrained and SIFT Feature Fusion. The Visual Computer, 2025, 41(7), 4915-4930. https://doi.org/10.1007/s00371-024-03698-4
20. Tan, R., Tian, Q., Lian, L., Zhang, X. Image Arbitrary Style Transfer via Super-Resolution Reconstruction. Computer Engi-neering and Applications, 2024, 60(15), 170-179. DOI: 10.3778/j.issn.1002-8331.2305-0042.
21. Wang, J. Garment Image Style Transfer Based on Deep Learning. Journal of Intelli-gent and Fuzzy Systems, 2023, 44(3), 3973-3986. https://doi.org/10.3233/JIFS-220761
22. Wang, S., Hou, Q., Li, J., Liu, J. TSID-Net: A Two-Stage Single Image Dehazing Frame-work with Style Transfer and Contrastive Knowledge Transfer. The Visual Computer, 2025, 41(3), 1921-1938. https://doi.org/10.1007/s00371-024-03511-2

23. Wang, X., Mao, Y., Xu, B., Wang, Z. Con-tent Structure Preserved Image Style Trans-fer Meth-od. Computer Engineering and Applications, 2023, 59(6), 146-154. DOI: 10.3778/j.issn.1002-8331.2110-0107.
24. Wu, C. Color Analysis of Cloud Brocade Pattern by Image Style Transfer. HighTech and Innovation Journal, 2023, 4(4), 779-786. https://doi.org/10.28991/HIJ-2023-04-04-07
25. Wu, Y., Liu, H., Kong, X., Deng, M., Li, Y. Simulation and Evaluation of GAN-Based Implementation of Infrared Texture Genera-tion. Journal of Engineering Science and Technology Review, 2024, 17(1), 122-130. https://doi.org/10.25103/jestr.171.16
26. Xie, Q., Yu, R. Image Style Transfer for Ex-hibition Hall Design Based on Multi-Modal Semantic-Enhanced Algorithm. Computers, Materials and Continua, 2025, 84(1), 1123-1144. https://doi.org/10.32604/cmc.2025.062712
27. Xu, L., Yuan, Q., Sun, Y., Gao, Q. Image Neural Style Transfer Combining Global and Local Optimization. The Visual Com-puter, 2024, 40(12), 8397-8411. https://doi.org/10.1007/s00371-023-03244-8
28. Yue, G., Li, M., Zheng, H. Image Style Transfer Based on Generative Adversarial Network and Feature Transformation for Modern Home Design. Journal of Applied Science and Engineering, 2025, 28(2), 257-263. DOI: 10.6180/jase.202502_28(2).0005.
29. Zhao, Y. Application of CycleGAN-Based Image Style Transfer Algorithm in Visual Communication Design. Journal of Compu-tational Methods in Sciences and Engineer-ing, 2025, 25(4), 3152-3164. https://doi.org/10.1177/14727978251318804
30. Zhou, Z., Tang, F., Zhang, Y., Deussen, O., Cao, J., Dong, W. A Comprehensive Evalu-ation of Arbitrary Image Style Transfer Methods. IEEE Transactions on Visualiza-tion and Computer Graphics, 2025, 31(9), 5668-5686. https://doi.org/10.1109/TVCG.2024.3466964
31. Zhu, Z., Mao, Y., Cai, K., Liu, W., Pu, D., Du, Y. Image Style Transfer Method for In-dustrial Inspection. Computer Engineering and Applications, 2023, 59(18), 234-241. DOI: 10.3778/j.issn.1002-8331.2205-0583.
32. Zou, W., Mutti, M. A. I. B. Research on Im-age Generation Style Transfer and Recon-struction Loss Reduction Based on Deep Learning Framework. Journal of Intelligent Systems and Internet of Things, 2025, 16(2), 117-122. https://doi.org/10.54216/JISIoT.160209