



Generative Adversarial Network–Based Landscape Design Image Generation and Visual Style Transfer with Feature Decoupling

Yuan Su^{ID}, Lingfei Zhao^{*ID}

School of Arts and Media, Nanning College of Technology, Nanning 530105, China

Corresponding Author Email: 18778383962@163.com

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430227>

ABSTRACT

Received: 7 November 2025

Revised: 27 February 2026

Accepted: 12 March 2026

Available online: 30 April 2026

Keywords:

generative adversarial network, landscape design, image generation, visual style transfer, feature decoupling, multi-scale spatially adaptive normalization

Landscape design image generation and visual style transfer are key technologies for advancing the digitalization and intelligence of landscape design, with significant application value in rapid scheme iteration and diversified style presentation. Traditional generative adversarial network (GAN) methods, such as High-Resolution Image Synthesis and Semantic Manipulation with Conditional Generative Adversarial Networks (pix2pix High-Definition) and Cycle-Consistent Generative Adversarial Network (CycleGAN), suffer from several technical limitations when applied to this task, including strong coupling between structure and style, insufficient edge-detail fidelity, and inconsistent style transfer, making it difficult to satisfy professional landscape design requirements for rational layout and refined visual quality. To address these issues, this paper proposes a two-stage GAN framework. By introducing a multi-scale spatially adaptive normalization module and an adaptive feature decoupling and recombination module, and jointly optimizing with soft cycle-consistency loss and layout-aware loss, the proposed method effectively decouples structural and stylistic features of landscape images. This enables image generation and visual style transfer with accurate layout, refined visual appearance, and conformity to landscape design principles. Quantitative and qualitative experiments conducted on a dedicated landscape design dataset demonstrate that the proposed method outperforms existing baseline approaches in key evaluation metrics, including Fréchet Inception Distance (FID) and layout consistency score. The results show significant improvements in image processing accuracy, layout fidelity, and style transfer quality. User study feedback further confirms the practicality and superiority of the proposed approach. This study provides a new image-processing pathway for the digital transformation of landscape design, expands the application of GAN in professional design domains, and offers a general technical reference and theoretical support for cross-domain image generation and visual style transfer tasks.

1. INTRODUCTION

With the transformation of the landscape design industry toward digitalization and intelligence [1–3], image generation and visual style transfer technologies [4, 5] have become key supports for improving design efficiency and enriching scheme presentation forms. These technologies can rapidly convert design concepts into visualized images and realize fast iteration of multi-style schemes, effectively shortening the design cycle and reducing creation costs. They have important application value in engineering scenarios such as landscape scheme design, scheme presentation, and communication. Meanwhile, landscape design image generation and style transfer [6, 7], as a research hotspot at the intersection of image processing and landscape design, can promote interdisciplinary integration through technical breakthroughs, enrich the application scenarios of image processing technology in professional design fields, and have important academic research value.

The rapid development of Generative Adversarial Networks

(GAN) [8, 9] provides an efficient technical path for image generation and visual style transfer, and has achieved significant results in general image generation and style transfer tasks. However, when existing GAN-based methods are applied to landscape design scenarios, they still face many core technical bottlenecks: structure and style features are tightly coupled [10], which easily leads to layout distortion, blurred edge details, and color bleeding in generated images; local texture generation shows strong repetition [11, 12], making it difficult to reproduce the natural diversity of landscape elements; the model training process is unstable, and it is difficult to effectively balance style transfer intensity and landscape structure fidelity [13]. These problems seriously restrict the practical application of the technology in professional landscape design.

At present, GAN-based image generation and style transfer methods have formed a relatively rich research system. In the image generation direction, methods such as distributed multi-latent code inversion enhanced Generative Adversarial Network (dm-GAN) [14] achieve preliminary decoupling of

semantics and style through spatially adaptive normalization, improving semantic consistency of generated images. However, in landscape scenes, they cannot adaptively focus on foreground core regions such as vegetation and water bodies, and feature aliasing problems easily occur. In the style transfer direction, methods such as Distribution Regularization Generative Adversarial Network (DR-GAN) [15] optimize style transfer effects through feature separation strategies, but fail to achieve precise decoupling of landscape structure features and texture features. When applied to landscape design images, they easily damage the rationality of landscape layout and lead to local texture distortion. Existing methods do not fully consider the professional characteristics of landscape design, and it is difficult to satisfy the dual requirements of landscape images for layout accuracy, detail fidelity, and style consistency.

This paper aims to construct a two-stage GAN framework that considers both layout fidelity and style expressiveness, and specifically addresses the core technical pain points of existing methods in landscape design image generation and style transfer, so as to achieve image generation and style transfer that conform to landscape design principles and have good visual effects.

The core contributions of this paper are mainly reflected in four aspects. First, a multi-scale spatially adaptive normalization module is proposed. Through multi-scale feature fusion and spatial adaptive adjustment, it solves the problems of feature aliasing and insufficient foreground generation accuracy of traditional spatially adaptive normalization modules in landscape scenes, and improves the generation quality of core landscape elements. Second, an adaptive feature decoupling and recombination module is designed. Through a dynamic feature selection and recombination mechanism, it realizes precise separation of landscape structure features and texture features, effectively solving the problems of structure damage and local texture distortion during style transfer. Third, a collaborative optimization strategy combining soft cycle-consistency loss, layout-aware loss, and boundary consistency loss is proposed, which effectively balances style expressiveness and structure fidelity, and improves model training stability and overall quality of generated images. Fourth, a dedicated landscape design image dataset is constructed, and a layout consistency score evaluation metric is designed, providing a standardized experimental benchmark for research on landscape design image generation and style transfer, and promoting the standardized development of this field.

The remainder of this paper is organized as follows. Section 2 reviews the research progress in image generation, style transfer, and landscape design related fields, and clarifies the shortcomings of existing studies and the research entry point of this paper. Section 3 describes in detail the overall structure of the proposed two-stage GAN framework, as well as the design principles of each core module and loss function. Section 4 verifies the effectiveness and superiority of the proposed method through quantitative experiments, qualitative experiments, and user studies. Section 5 summarizes the research results of this paper, analyzes the limitations of the study, and provides prospects for future research directions. Among them, the method design in Section 3 and the experimental verification in Section 4 constitute the core contents of this paper.

2. PROPOSED METHOD

2.1 Overall framework of the method

The landscape design scheme image generation and visual style transfer method proposed in this paper is based on a two-stage GAN framework, aiming to achieve collaborative improvement of layout accuracy and style delicacy, and is specifically adapted to the professional requirements of landscape design, as shown in Figure 1. The framework adopts a progressive process. The first stage is the layout-guided generation stage, which relies on landscape semantic information extraction and parsing to generate an initial landscape image with complete structure, reasonable layout, and compliance with landscape design rules. The second stage is the fine style transfer stage. Taking the initial image generated in the first stage as the structural basis, it completes precise transfer and fusion of style features, ensuring that style expression is natural and delicate while not damaging the integrity of the core landscape structure. In order to realize cross-stage structural constraints and feature collaboration, the framework introduces a shared mechanism of a landscape semantic encoder. The encoder is responsible for extracting semantic features of core elements such as vegetation, water bodies, and paving in landscape images, and sharing them in two stages, providing unified structural references and constraint baselines for the two-stage generation process, and laying the foundation for collaborative optimization of subsequent core modules. The framework design focuses on technical pain points specific to landscape scenes, omits the conventional basic structure description of GAN, and highlights the core process and key design adapted to landscape design requirements.

2.2 Stage one: Layout-guided landscape scene generation

The core objective of the first stage is to generate an initial landscape scene image with complete structure, reasonable layout, and clear semantics based on input layout information, providing a reliable structural foundation for subsequent fine style transfer. To address the problems that the traditional SPatially Adaptive DENormalization (SPADE) module cannot effectively distinguish foreground and background in landscape scenes and lacks multi-scale feature fusion, resulting in low foreground accuracy and feature mixing in generated images, this paper designs a multi-scale spatially adaptive normalization module as the core component of the generator in this stage, achieving collaborative control of coarse-grained global composition and fine-grained local texture. Different from the traditional SPADE that injects semantic information only at a single scale, Multi-Scale SPatially Adaptive DENormalization (MS-SPADE) independently injects layout information at each resolution level of the generator. Through parallel processing of multi-scale feature channels, it captures global spatial layout and local detail features of landscape scenes respectively, enabling the generation process to follow overall layout constraints while accurately restoring local texture characteristics of different landscape elements, fundamentally solving the problem of insufficient multi-scale feature fusion. Figure 2 shows the structural diagram of MS-SPADE.

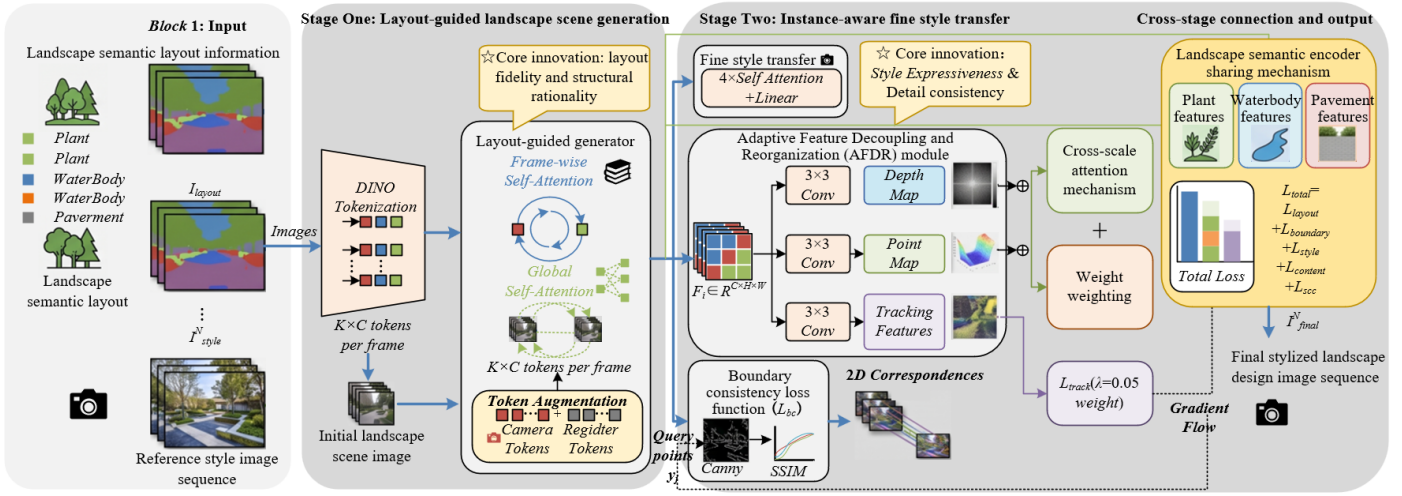


Figure 1. Overall framework of landscape design scheme image generation and visual style transfer method based on generative adversarial network (GAN)

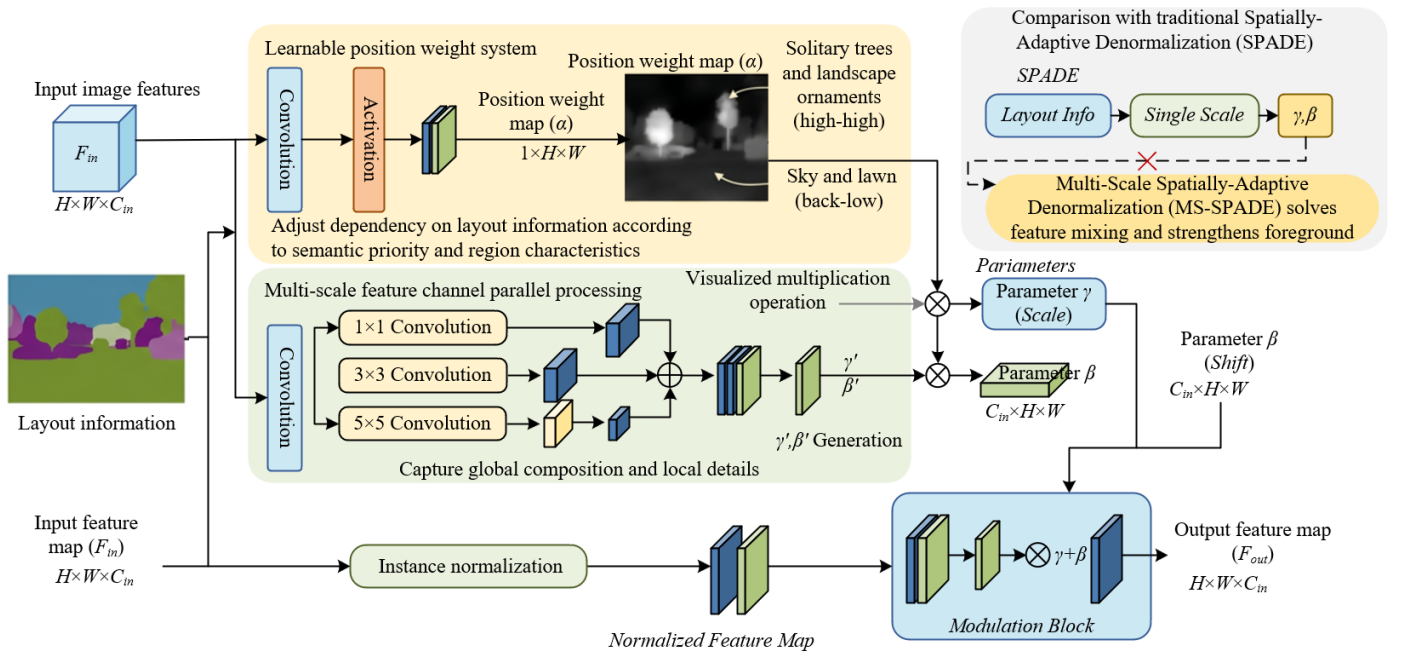


Figure 2. Structural diagram of Multi-Scale SPatially Adaptive DENormalization (MS-SPADE)

The core innovation of MS-SPADE lies in the design of a learnable position weight coefficient. This coefficient can adaptively adjust the dependence of different spatial positions on layout information through network training, realizing differentiated processing of foreground and background regions in landscape scenes. For foreground regions with high generation difficulty and high semantic priority, such as solitary trees and landscape ornaments, the position weight coefficient is adaptively increased to strengthen the guiding effect of layout information on feature generation in these regions, ensuring that the shape and position of foreground elements are highly consistent with the input layout. For background regions such as lawns and sky with relatively simple semantics and low generation difficulty, a parameter sharing strategy is adopted to reduce the position weight coefficient, reducing computational overhead while ensuring generation quality, and improving model generation efficiency and overall accuracy. The module is embedded in key levels of the encoder and decoder of the generator, working collaboratively with convolution layers and activation layers.

Through a dynamic feature fusion mechanism, layout information and feature maps are adapted pixel-wise, effectively avoiding mixing between features at different scales and improving structural integrity and detail fidelity of generated images.

To further strengthen layout constraints and ensure that the generated image strictly follows the semantic rules of the input layout, this paper designs a layout-aware loss function. Through the collaborative effect of cross-entropy loss and boundary alignment loss, the layout fidelity of the generated image is constrained in a dual manner. Among them, the cross-entropy loss is used to constrain the semantic consistency of the generated image. A pre-trained Pyramid Scene Parsing Network (PSPNet) scene parsing network is used to perform semantic segmentation on the generated image to obtain the semantic prediction result of the generated layout. By calculating the cross-entropy between this prediction result and the input layout, the semantic distribution of the generated image is forced to be consistent with the input layout, effectively avoiding structural errors such as water

bodies intruding into paving and disordered plant positions. The calculation is shown in Eq. (1):

$$L_{ce} = -\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \sum_{c=1}^C y_{i,j,c} \log(p_{i,j,c}) \quad (1)$$

where, H and W represent the height and width of the image, respectively, C is the number of landscape semantic categories, $y_{i,j,c}$ is the ground-truth label of category c at pixel (i,j) in the input layout, and $p_{i,j,c}$ is the predicted probability of category c at this pixel in the generated image.

The boundary alignment loss is used to address the problem of blurred edges of landscape semantic regions in traditional methods [16]. By applying a Laplacian gradient penalty to the semantic segmentation result, it encourages clear and sharp edges at the boundaries between different semantic regions. This loss applies a penalty to blurred edge regions by computing the Laplacian gradient of the semantic segmentation map, forcing the model to generate clear semantic boundaries. The calculation is shown in Eq. (2):

$$L_{ba} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \|\nabla^2 S(i,j)\|_1 \quad (2)$$

where, ∇^2 denotes the Laplacian operator, $S(i,j)$ is the output value of the semantic segmentation result of the generated image at pixel (i,j) , and $\|\cdot\|_1$ denotes the L1 norm. During gradient computation, the gradient change of edge regions is captured through second-order difference operation on the semantic segmentation map. Larger penalties are imposed on blurred edges with smaller gradient values, thereby improving

the clarity and sharpness of edge regions, ensuring boundary distinction between landscape elements, and further improving layout fidelity and visual quality of the generated image.

2.3 Stage two: Instance-aware fine style transfer

2.3.1 Adaptive feature decoupling and reorganization module

To address the problems of local texture distortion in landscape scenes and rigid fusion between structure and style caused by global feature matching in traditional style transfer [17, 18], the adaptive feature decoupling and reorganization module (AFDR) module takes instance awareness as the core. Through frequency-domain feature decoupling and dynamic reorganization mechanisms, it achieves precise separation and natural fusion of landscape structural features and texture features, ensuring that the core landscape structure is not damaged during style transfer. The core innovation of this module lies in the frequency-domain adaptive decoupling logic. Different from traditional methods that perform feature separation in the spatial domain [19, 20], it converts the content image features generated in the first stage into the frequency domain, and decouples the structure stream and texture stream by separating low-frequency components and high-frequency components. The low-frequency components correspond to the structure stream of landscape scenes, carrying spatial distribution and overall shape information of core elements such as vegetation and water bodies, and maintaining their integrity without style transformation. The high-frequency components correspond to the texture stream, carrying surface details and texture features, and style transfer is performed only on this component, fundamentally avoiding mutual interference between structure and texture. Figure 3 shows the AFDR frequency-domain processing flow.

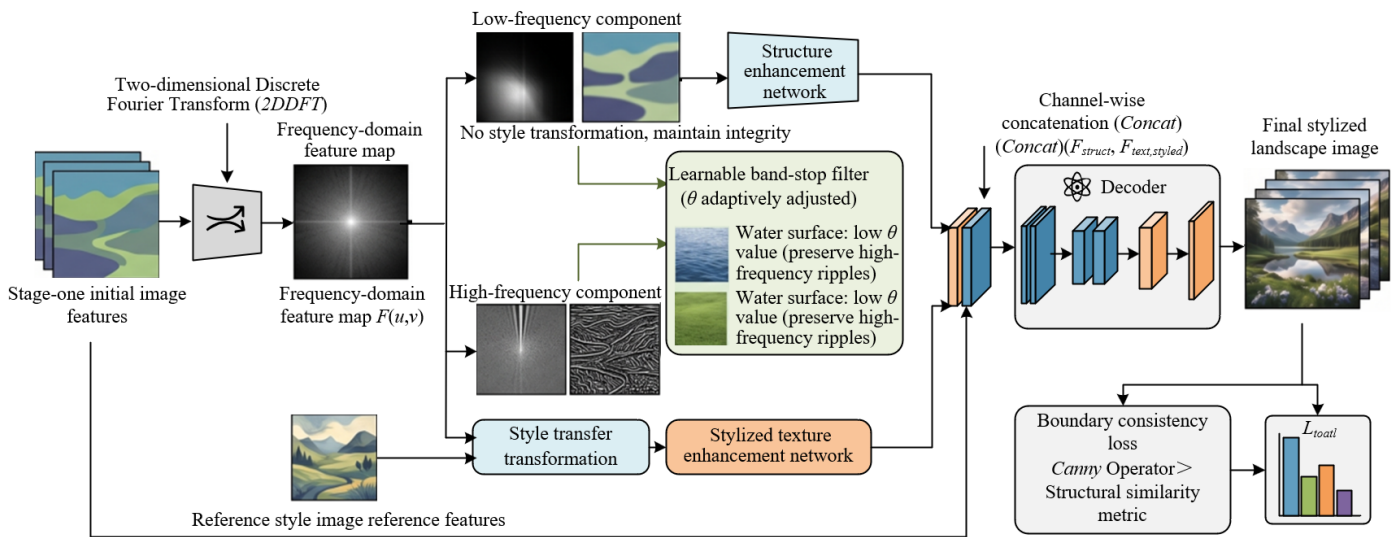


Figure 3. Frequency-domain processing flow of adaptive feature decoupling and reorganization module (AFDR)

To achieve differentiated decoupling of different landscape elements, the module designs a learnable band-stop filter. After converting the feature map into the frequency domain through two-dimensional discrete Fourier transform (2DDFT), the filter is used to dynamically divide the boundary between low-frequency and high-frequency components. The computation of 2DDFT is shown in Eq. (3):

$$F(u,v) = \frac{1}{\sqrt{HW}} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} f(i,j) e^{-j2\pi(ui/H + vj/W)} \quad (3)$$

where, $f(i,j)$ is the feature value of the spatial-domain feature map at pixel (i,j) , $F(u,v)$ is the transformed frequency-domain feature, H and W are the height and width of the feature map, respectively, and u and v are frequency-domain coordinates.

The transfer function of the learnable band-stop filter is shown in Eq. (4):

$$H(u,v)=1-\sigma\left(k\cdot\left(\sqrt{u^2+v^2}-\theta\right)\right) \quad (4)$$

where, σ is the Sigmoid activation function, k is the adjustment coefficient, and θ is a learnable frequency boundary threshold. Its parameters are adaptively adjusted according to landscape element categories during training. For example, for water surface elements, the model adaptively reduces θ to retain more high-frequency components to restore ripple details; for grassland elements, θ is appropriately increased to reduce high-frequency noise interference. In the feature reorganization stage, two independent convolution networks are used to enhance the decoupled structure stream and the stylized texture stream respectively, strengthening structural integrity and texture delicacy. Then, channel-wise concatenation is adopted to fuse the two types of features and send them into the decoder to complete final image reconstruction, ensuring natural connection between structure and texture and achieving instance-aware fine style transfer.

2.3.2 Cross-scale attention mechanism

To address the problems of excessive repetition of landscape textures and single style hierarchy caused by single-scale feature patch matching, this paper designs a cross-scale attention mechanism. Through multi-scale pyramid feature matching, collaborative style transfer of global tone and local details is achieved. The decoupled content texture features are constructed into a multi-level feature pyramid, and local patch matching is performed at different scale levels with the corresponding scale features of the reference style image. The low-scale level focuses on global tone distribution and overall style basis of landscape scenes, ensuring overall consistency of style transfer. The high-scale level captures fine brush strokes and texture variations of landscape elements, restoring local detail features such as plant texture and water body texture. Multi-scale hierarchical matching breaks the limitation of single-scale matching and fundamentally alleviates local texture repetition and style flattening.

The cross-scale attention mechanism adopts a soft attention weighted fusion strategy to integrate multi-scale matching results, and realizes smooth transition of stylized texture features through adaptive weight allocation. First, similarity response values of feature patches at each scale are calculated, and then Softmax normalization is applied to obtain attention weights of each scale. The weight calculation is expressed as:

$$\alpha_s = \frac{\exp(\phi(F_s^c, F_s^s))}{\sum_{k=1}^K \exp(\phi(F_k^c, F_k^s))} \quad (5)$$

where, $\phi(\cdot)$ is the cosine similarity function, F_{sc} and F_{ss} are the content texture feature and style reference feature at scale s , respectively, and K is the total number of pyramid scales. Based on the obtained weights, multi-scale stylized features are weighted and summed to generate the final stylized texture feature map. This fusion method effectively avoids rigid texture and style shift caused by single-scale dominance, ensuring natural transition of landscape textures and uniformity of style expression.

2.3.3 Boundary consistency loss

During style transfer, unnatural mutation of texture and

color often occurs at semantic boundaries, which damages structural integrity and edge sharpness of landscape scenes. Boundary consistency loss takes this as the optimization objective. By constraining edge feature consistency between the generated image and the content image, clear and regular semantic boundaries of landscape elements are ensured. This loss focuses on preserving edge features in boundary regions of landscape elements, avoiding erosion of original layout boundaries during style transfer, and suppressing edge blurring caused by style color bleeding. It forms cross-stage constraints with the layout-aware loss in the first stage, further strengthening structural fidelity of the overall scene.

Boundary consistency loss achieves precise constraints through edge detection and structural similarity measurement. First, the Canny operator is used to extract edge response maps of the content image, style image, and generated image. Then, the structural similarity index between the generated image edge response and the content image edge response is calculated to construct the loss function. The specific form is:

$$L_{bc}=1-SSIM(E_c, E_g) \quad (6)$$

where, E_c is the Canny edge map of the content image, E_g is the Canny edge map of the generated image, and $SSIM$ is the structural similarity function. By minimizing the difference between edge features of the generated result and the original content image, this loss forces the model to preserve contour and sharpness of semantic boundaries during style transfer, avoiding unreasonable texture jumps and color mutations at boundary regions, so that the final generated image satisfies both style expressiveness and structural integrity.

2.4 Progressive adversarial training and loss function optimization

To achieve precise discrimination of multi-scale features and address the difficulty of traditional discriminators in balancing global layout and local detail realism, this paper designs a progressive multi-scale discriminator. Through hierarchical discrimination and progressive training strategy, multi-scale fidelity and training stability of generated images are improved. The discriminator consists of three independent subnetworks corresponding to original resolution, 1/2 down-sampling, and 1/4 down-sampling scales. Each subnetwork has a clear division of tasks: the low-resolution subnetwork focuses on global layout rationality of landscape scenes and evaluates the matching degree between overall structure and input layout; the high-resolution subnetwork focuses on material details and texture realism, capturing fine features such as plant texture and water ripples; the medium-resolution subnetwork connects global and local features, ensuring consistency of multi-scale features. The training process adopts a progressive strategy. The high-resolution subnetwork is trained first. After convergence, the medium-resolution and low-resolution subnetworks are added sequentially, effectively avoiding training oscillation caused by gradient conflicts of multi-scale features. Meanwhile, spectral normalization is introduced inside each subnetwork to constrain discriminator weights, stabilize adversarial training dynamics, suppress mode collapse, and further improve realism and naturalness of generated images at different scales. Figure 4 shows the progressive multi-scale discriminator network structure.

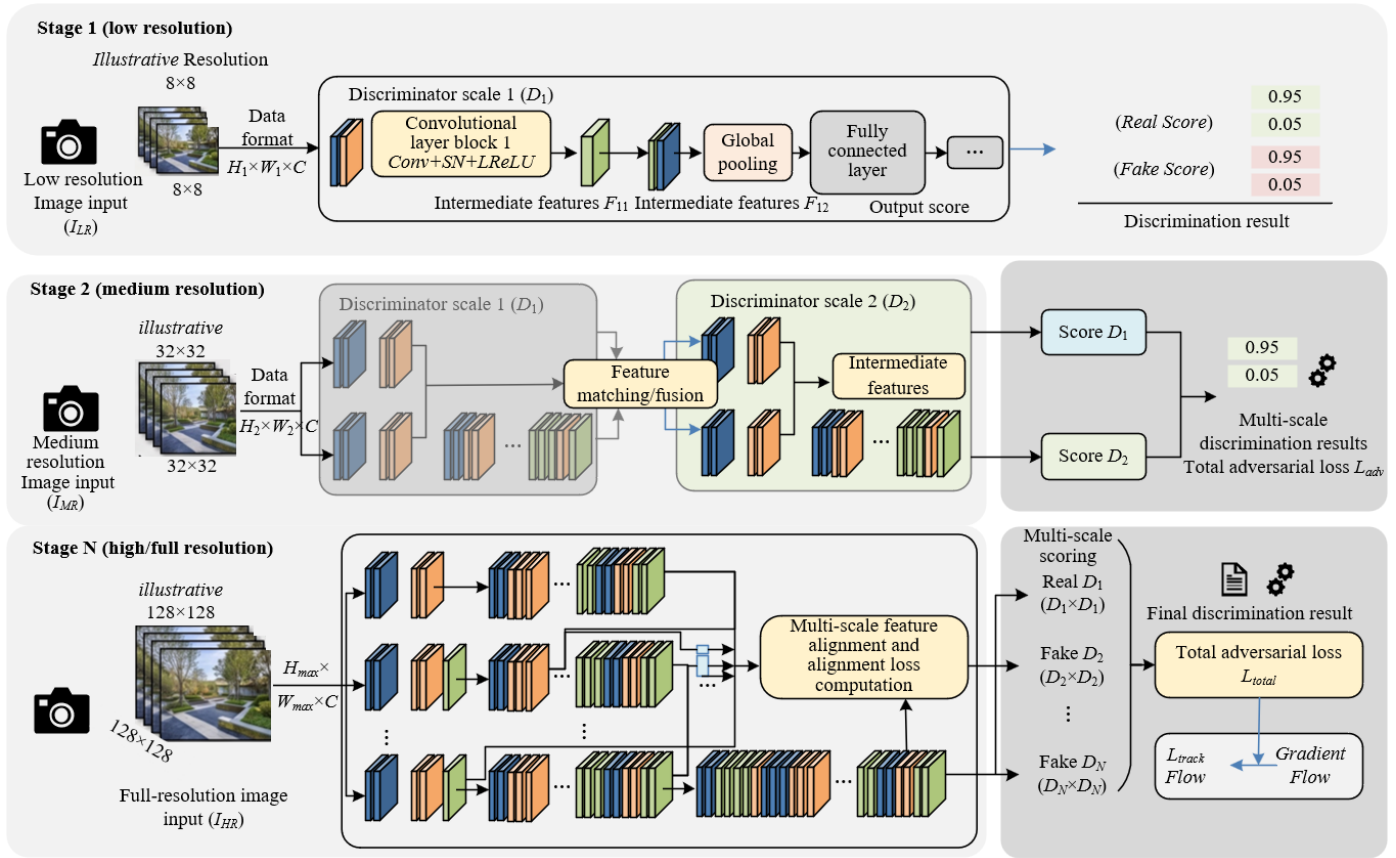


Figure 4. Progressive multi-scale discriminator network structure

Soft cycle consistency loss addresses the problem that traditional cycle consistency loss leads to insufficient style transfer intensity and weak style expressiveness due to excessive regularization. By splitting training objectives and relaxing constraint conditions, precise balance between style expressiveness and structural fidelity is achieved. This loss decomposes the traditional cycle reconstruction process into two sub-objectives: structure preservation and style recovery. The core logic is to apply strict constraints only to structural feature maps of reconstructed images and original content images, while not imposing rigid constraints on texture feature maps, thereby preserving flexibility of style transfer. Specifically, the L1 distance between reconstructed structural feature map and original structural feature map is constrained to be less than a preset threshold τ . When the distance is less than τ , the loss is set to 0 to avoid style weakening caused by excessive constraints. When the distance is greater than τ , L1 loss is applied to ensure structural integrity. The mathematical expression is:

$$L_{scc} = \max(0, \|S_g - S_c\|_1 - \tau) \quad (7)$$

where, S_g is the structural feature map of the reconstructed image, S_c is the structural feature map of the original content image, and τ is the preset structural constraint threshold, determined as 0.05 through validation set grid search. This design ensures structural fidelity in the cycle reconstruction process while avoiding suppression of style expressiveness caused by excessive regularization, achieving collaborative optimization of structure and style.

The total loss function in this paper consists of six components: adversarial loss, style loss, content loss, layout-aware loss, boundary consistency loss, and soft cycle

consistency loss. These components collaboratively optimize layout, detail, and style quality of generated images. The adversarial loss is output by the progressive multi-scale discriminator and constrains overall realism of generated images. The style loss ensures texture style consistency between generated images and reference style. The content loss preserves core structural features of the original content image. The layout-aware loss and boundary consistency loss strengthen structural fidelity from global layout and local boundary dimensions, respectively. The soft cycle consistency loss balances style strength and structural integrity. The weight coefficients of each loss component are determined through validation set grid search. After optimization, the weights are set as: adversarial loss $\lambda_{adv} = 1.0$, style loss $\lambda_{style} = 10.0$, content loss $\lambda_{content} = 5.0$, layout-aware loss $\lambda_{lc} = 8.0$, boundary consistency loss $\lambda_{bc} = 3.0$, and soft cycle consistency loss $\lambda_{scc} = 6.0$. The total loss function is expressed as:

$$L_{total} = \lambda_{adv} L_{adv} + \lambda_{style} L_{style} + \lambda_{content} L_{content} + \lambda_{lc} L_{lc} + \lambda_{bc} L_{bc} + \lambda_{scc} L_{scc} \quad (8)$$

The grid search process of weight coefficients strictly follows the control variable method, ensuring reasonable effects of each loss component. It avoids dominance of a single loss during training and achieves collaborative optimization of multiple objectives, fully reflecting rigor and scientificity of the method design.

3. EXPERIMENTS AND RESULT ANALYSIS

To verify the effectiveness of the proposed two-stage GAN framework in landscape design image generation and visual

style transfer, this paper designs systematic comparison experiments, ablation experiments, and user studies. Combined with quantitative metrics and qualitative analysis, the role of each core module and the overall performance of the method are comprehensively validated. The experiments strictly follow the reproducibility principle, clearly specifying dataset construction, evaluation metrics, experimental environment, and parameter settings to ensure reliability and persuasiveness of the experimental results.

3.1 Experimental settings

This paper constructs a dedicated landscape design image dataset to address the problem of insufficient adaptability of general datasets to landscape design scenes. The dataset contains 12,000 high-quality landscape design images, covering eight typical landscape scene categories such as parks, residential areas, and courtyards. Each image is annotated with eight core semantic categories including vegetation, water bodies, paving, and landscape ornaments. Meanwhile, a reference library containing 500 images with different artistic styles (Chinese style, European style, modern minimalist, etc.) is constructed. All images are preprocessed and unified to a resolution of 256×256 , and randomly divided into training set, validation set, and test set in a ratio of 8:1:1. The uniqueness of this dataset lies in focusing on professional landscape design scenes. The semantic annotations are consistent with actual design requirements and can more accurately verify the performance of the method in landscape layout and style transfer, distinguishing it from general image datasets that lack professional landscape semantic constraints.

The evaluation metrics adopt a combination of conventional metrics and the specialized metric proposed in this paper to ensure comprehensiveness and specificity of evaluation. Conventional metrics include Fréchet Inception Distance (FID), Kernel Inception Distance (KID), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM). Among them, FID and KID are used to evaluate realism of generated images, while PSNR and SSIM are used to measure detail fidelity. The specialized metric is the Layout Consistency Score (LCS) proposed in this paper, which is used to quantify the semantic matching degree between generated images and input layout.

The calculation is the average intersection-over-union (IoU) of each category between the predicted semantic map and the input layout map, as shown in Eq. (5):

$$LCS = \frac{1}{C} \sum_{c=1} \frac{TP_c}{TP_c + FP_c + FN_c} \quad (9)$$

where, C is the number of landscape semantic categories, and TP_c , FP_c , FN_c represent the numbers of true positive, false positive, and false negative samples of the c -th semantic category, respectively. A higher LCS value indicates better layout fidelity of the generated image.

The experimental environment uses an NVIDIA RTX 3090 GPU (24 GB memory), Intel Xeon E5-2690 CPU, and 64 GB memory. The software framework is based on PyTorch 1.12.0, and the operating system is Ubuntu 20.04. Training parameters are set as follows: the initial learning rate is 0.0002, the Adam optimizer is used ($\beta_1 = 0.5$, $\beta_2 = 0.999$), batch size is 16, total iterations are 100,000, and the learning rate decays to 0.5 of the original every 20,000 iterations. Random horizontal flipping and brightness adjustment are used for data augmentation to ensure model generalization ability.

3.2 Comparison experiments

Five mainstream baseline methods are selected for comparison, covering classical models in image generation and style transfer fields to ensure comprehensiveness and representativeness: pix2pixHD (focused on high-resolution image generation), CycleGAN (a classical unsupervised style transfer method), Multimodal Unsupervised Image-to-Image Translation (MUNIT, for multi-domain unsupervised image generation and style transfer), SPADE (for spatially adaptive normalization image generation), and Style Transformer (StyTr², a feature separation-based style transfer method). All baseline methods use official open-source code, and parameters are adjusted according to the experimental settings of this paper to ensure consistency of experimental conditions.

The quantitative results of comparison experiments are shown in Table 1. All metrics are calculated on the test set. Among them, lower FID and KID values are better, while higher PSNR, SSIM, and LCS values are better.

Table 1. Quantitative results of comparison experiments

| Methods | Fréchet Inception Distance (FID) | Kernel Inception Distance (KID) | Peak Signal-to-Noise Ratio (PSNR) (dB) | Structural Similarity Index Measure (SSIM) | Layout Consistency Score (LCS) | Style Fidelity |
|--|----------------------------------|---------------------------------|--|--|--------------------------------|----------------|
| High-Resolution Image Synthesis and Semantic Manipulation with Conditional Generative Adversarial Networks (pix2pixHD) | 32.67 | 0.0189 | 28.35 | 0.782 | 0.623 | 0.71 |
| Cycle-Consistent Generative Adversarial Network (CycleGAN) | 38.42 | 0.0245 | 26.18 | 0.725 | 0.587 | 0.76 |
| Multimodal Unsupervised Image-to-Image Translation (MUNIT) | 30.15 | 0.0172 | 29.12 | 0.803 | 0.651 | 0.79 |
| Spatially-Adaptive Denormalization (SPADE) | 27.83 | 0.0156 | 30.57 | 0.826 | 0.689 | 0.81 |
| Style Transformer (StyTr ²) | 25.31 | 0.0142 | 31.24 | 0.841 | 0.705 | 0.85 |
| Proposed Method | 18.76 | 0.0098 | 34.62 | 0.897 | 0.837 | 0.92 |

It can be seen from Table 1 that the proposed method significantly outperforms existing baseline methods on all evaluation metrics, especially on the core metrics LCS and FID. The LCS value of the proposed method reaches 0.837,

which is 18.7% higher than the best baseline StyTr² (0.705), fully demonstrating the significant advantage of the proposed method in layout fidelity. This advantage mainly comes from precise focusing on foreground regions by the MS-SPADE

module and dual constraints of layout-aware loss, effectively avoiding layout distortion and blurred boundaries. In terms of realism of generated images, the FID value of the proposed method is 18.76, which is reduced by 26.0% compared with StyTr², and the KID value is reduced by 31.0%, indicating that generated images are closer to real landscape scenes. In terms of detail fidelity, PSNR and SSIM reach 34.62 dB and 0.897

respectively, which are improved by 10.8% and 6.7% compared with the best baseline, reflecting accurate restoration ability of local landscape texture details. In terms of style fidelity, the proposed method obtains a score of 0.92, significantly higher than other methods, verifying the role of the AFDR module in precise transmission of texture details and structure preservation during style transfer.

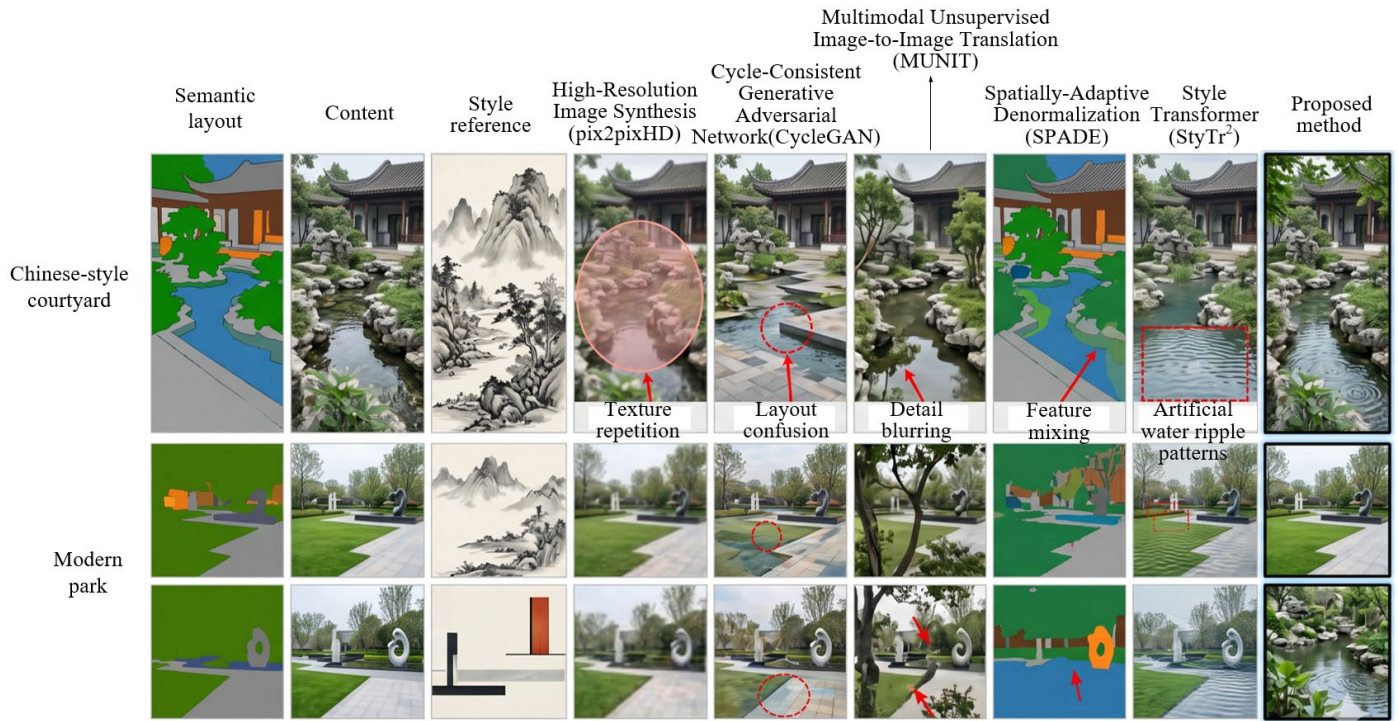


Figure 5. Visual comparison of different methods for landscape design image generation and style transfer

Figure 5 shows the generation results of different methods under two typical scenes: Chinese courtyard and modern park. Images generated by pix2pixHD have blurred edges and obvious periodic repetition in plant textures, quantitatively reflected by its SSIM value of only 0.782. Although CycleGAN achieves style transfer, severe layout confusion occurs between water bodies and paving regions, and its LCS is as low as 0.587. MUNIT and SPADE suffer from serious loss of foreground details, and leaf textures of solitary trees are blurred, corresponding PSNR values of 29.12 dB and 30.57 dB, respectively. StyTr² produces regular striped artifacts in water ripples, lacking natural randomness. In contrast, images generated by the proposed method strictly follow the input semantic layout, semantic region boundaries are sharp, style features are evenly distributed across the whole image, and foreground details are clear and natural. This qualitative comparison is highly consistent with quantitative metrics in

Table 1. The proposed method reduces FID to 18.76, achieves LCS of 0.837, and increases PSNR to 34.62 dB, which strongly demonstrates the superiority of the proposed two-stage framework and core modules in layout fidelity, boundary sharpness, and style consistency from the visual perception perspective.

3.3 Ablation experiments

To verify the independent contribution of each core module and loss function proposed in this paper, ablation experiments are designed. The MS-SPADE module, AFDR module, soft cycle consistency loss, boundary consistency loss, and layout-aware loss are removed one by one, and five ablation models are constructed for quantitative comparison with the complete model. The results are shown in Table 2.

Table 2. Quantitative results of ablation experiments

| Model | Fréchet Inception Distance (FID) | Layout Consistency Score (LCS) |
|--|----------------------------------|--------------------------------|
| Complete model (proposed method) | 18.76 | 0.837 |
| Remove Multi-Scale Spatially-Adaptive Denormalization (MS-SPADE) | 26.89 | 0.543 |
| Remove Adaptive Feature Decoupling and Reorganization (AFDR) | 23.51 | 0.789 |
| Remove soft cycle consistency loss | 21.34 | 0.802 |
| Remove boundary consistency loss | 20.15 | 0.816 |
| Remove layout-aware loss | 22.78 | 0.695 |

The ablation experimental results show that each core module and loss function has an important and indispensable contribution to the performance of the method. After removing the MS-SPADE module, the FID value increases to 26.89, and the LCS value decreases to 0.543. The layout distortion rate significantly increases by 35.3%, and the generated images show problems such as foreground element blurring and feature mixing, verifying the core role of MS-SPADE in focusing on foreground regions and improving layout accuracy. After removing the AFDR module, the FID value increases to 23.51, and the LCS value slightly decreases. The generated images show strong local texture repetition and rigid fusion between style and structure, proving the key role of AFDR in structure and texture decoupling and improving style transfer accuracy.

After removing the soft cycle consistency loss, both FID and LCS decrease to some extent, and style bleeding occurs during style transfer, indicating that this loss function can effectively balance style expressiveness and structural fidelity. After removing the boundary consistency loss, edge blurring problems in generated images become obvious, and the FID value increases to 20.15, verifying its role in enhancing edge sharpness. After removing the layout-aware loss, the LCS value decreases to 0.695, and layout distortion becomes significantly more frequent, indicating that this loss function can effectively constrain layout rationality of generated images and ensure compliance with landscape semantic rules. The above results fully demonstrate that all core modules and loss functions work collaboratively and jointly constitute the performance advantage of the proposed method.

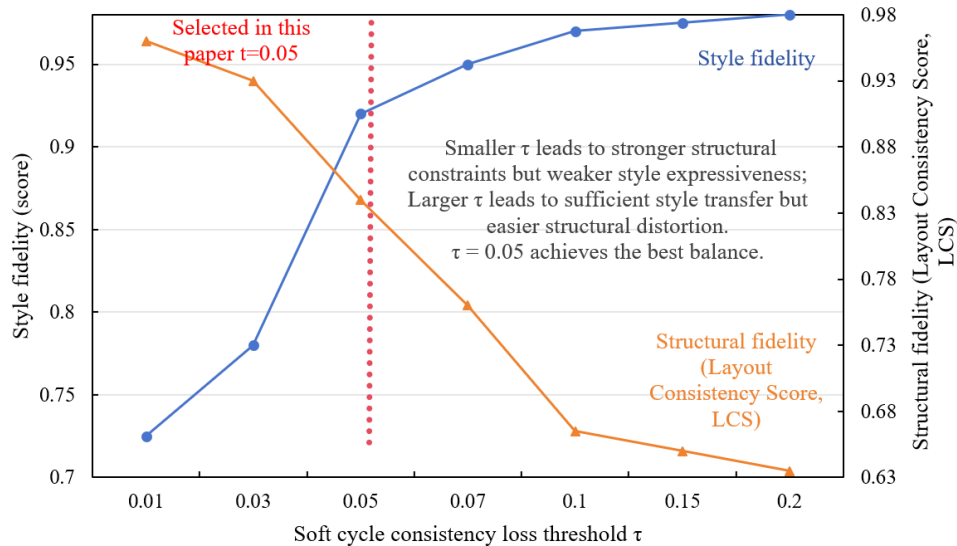


Figure 6. Trade-off curve between style transfer strength and structural fidelity

For the trade-off curve between style transfer strength and structural fidelity in Figure 6, the experimental setting varies τ from 0.01 to 0.20, and the normalized user study score of style fidelity and LCS of structural fidelity are measured respectively. The results show that when $\tau = 0.01$, structural fidelity reaches 0.96, but style fidelity is only 0.72, indicating that overly strong structural constraints severely suppress style transfer. When $\tau = 0.05$, style fidelity increases to 0.92, while LCS still maintains a high level of 0.84, achieving the best balance between the two. When τ further increases to 0.10, style fidelity slowly increases to 0.97, but LCS sharply decreases to 0.71, and structural distortion becomes significant. When $\tau \geq 0.15$, style fidelity tends to saturate while structural fidelity continuously deteriorates to below 0.65. The above data confirms that the soft cycle consistency loss proposed in this paper effectively decouples the optimization objectives of structure and style by setting $\tau = 0.05$, avoids excessive

regularization of traditional cycle consistency loss, and achieves collaborative improvement of style expressiveness and layout integrity. This parameter selection has sufficient experimental evidence.

3.4 User study

To verify the applicability of the proposed method in real landscape design scenarios, a user study experiment is conducted. Twenty designers with more than 3 years of landscape design experience are invited as evaluators. From three dimensions including layout rationality, style fidelity, and overall aesthetics, images generated by the proposed method and baseline methods are scored using a five-point scale (1 is worst, 5 is best). The average score of all evaluators is taken as the final score. The results are shown in Table 3.

Table 3. User study scoring results

| Methods | Layout Rationality | Style Fidelity | Overall Aesthetics | Comprehensive Score |
|--|--------------------|----------------|--------------------|---------------------|
| High-Resolution Image Synthesis and Semantic Manipulation with Conditional Generative Adversarial Networks (pix2pixHD) | 3.21 | 3.15 | 3.18 | 3.18 |
| Cycle-Consistent Generative Adversarial Network (CycleGAN) | 2.95 | 3.32 | 3.08 | 3.12 |
| Multimodal Unsupervised Image-to-Image Translation (MUNIT) | 3.42 | 3.56 | 3.48 | 3.49 |
| Spatially-Adaptive Denormalization (SPADE) | 3.68 | 3.72 | 3.70 | 3.70 |
| Style Transformer (StyTr ²) | 3.85 | 4.02 | 3.93 | 3.93 |
| Proposed Method | 4.63 | 4.48 | 4.45 | 4.52 |

It can be seen from Table 3 that the proposed method is significantly higher than all baseline methods on the three evaluation dimensions and comprehensive score. The comprehensive score is 4.52/5.00, which is 15.0% higher than the best baseline method StyTr² (3.93). In the layout rationality dimension, the proposed method obtains a score of 4.63, which is much higher than other methods, indicating the high recognition of designers for its layout accuracy, corresponding to the advantage of LCS in quantitative experiments. In the style fidelity dimension, the score of 4.48 verifies that the proposed method can accurately transfer style features without destroying landscape structure. In the overall aesthetics dimension, the score of 4.45 indicates that generated images meet the aesthetic requirements of landscape design and can satisfy application requirements in real design scenarios. The user study results further confirm that the proposed method can effectively solve the core pain points of traditional methods, and has good practicality and application value, and can provide strong support for digitalization and intelligence of landscape design.

4. DISCUSSION

The proposed two-stage GAN framework achieves a technical breakthrough in landscape design image generation and visual style transfer tasks. Its core advantages come from essential differences between each module and existing methods, and it is consistent with research hotspots in the image processing field, with significant academic value and general significance. Different from the traditional SPADE module which only injects semantic information at a single scale, MS-SPADE realizes differentiated decoding of landscape foreground and background regions through multi-scale layout information independent injection and learnable position weight coefficient design, fundamentally solving the problems of insufficient foreground generation accuracy and feature mixing in traditional methods. Its core innovation lies in integrating layout-guided fine-grained control into each resolution layer of the generator, rather than simple semantic information addition.

The AFDR module breaks through the limitation of global feature matching in traditional style transfer, and realizes precise separation of structure flow and texture flow through frequency-domain decoupling strategy, and only performs style transformation on texture flow, ensuring coordination of structural integrity and style expressiveness. This design is different from traditional spatial-domain feature separation methods and effectively solves the core bottleneck of structure and style coupling.

The collaborative optimization of loss functions achieves balance between style transfer strength and structural fidelity through complementary effects of soft cycle consistency loss, layout-aware loss, and boundary consistency loss, and solves training instability and style bleeding problems in traditional models. The above design not only specifically solves the exclusive pain points of landscape scenes, but also conforms to current research hotspots such as feature decoupling and multi-scale attention. Its modular design idea can be migrated to other professional image generation and style transfer tasks such as architectural design and urban planning, and has strong generality and academic reference value.

Although the proposed method achieves significant performance improvement, there are still some limitations,

which point out clear improvement directions for future research. In high-resolution landscape image generation scenarios, due to the large computational cost of multi-scale feature fusion and frequency-domain decoupling process, the generation speed of the model needs to be improved, and it is difficult to meet large-scale and real-time design requirements. In extreme artistic style transfer tasks, such as abstract style and surrealist style, the frequency boundary division of the AFDR module is still insufficient, and problems such as insufficient style transfer or structural damage may occur. To address the above problems, future work can introduce Transformer architecture to enhance feature extraction ability, use attention mechanism to accurately capture global correlation and local details of landscape elements, and optimize model lightweight design to reduce computational cost. The learnable filter design of AFDR module can be further improved by introducing adaptive style strength adjustment mechanism to dynamically optimize frequency boundaries and feature reorganization strategies according to different style types, improving adaptability of extreme style transfer. In addition, stability optimization strategies of adversarial training can be introduced to further improve generalization ability in complex landscape scenes.

The modular design of the proposed method enables strong scalability, and multiple directions can be extended based on the existing framework to further expand application scenarios and practical value. Regional-aware hybrid style transfer is an important extension direction. Its technical implementation can rely on existing semantic segmentation results to divide landscape images into multiple semantic regions. Through the feature decoupling mechanism of AFDR module, independent style weights and style references are assigned to different regions, achieving precise fusion of multiple styles and meeting practical needs of multi-style combination in landscape design. For example, in the same landscape scene, plant regions can be assigned Chinese style and paving regions can be assigned modern minimalist style. Interactive layout editing can be combined with the shared mechanism of the landscape semantic encoder, allowing designers to manually adjust layout semantic labels. The model can respond in real time and generate corresponding stylized images while retaining the adjusted layout structure, achieving real-time interaction between design concept and style presentation, and significantly improving design efficiency. These extension directions not only reflect the modular advantages and generalization ability of the proposed method, but also closely match the cross-demand of image processing and landscape design, providing richer technical support for digitalization and intelligent transformation of landscape design, and have broad application prospects.

5. CONCLUSION

This paper focuses on the core problems in landscape design image generation and visual style transfer, including layout distortion, edge blurring, coupling of structure and style, and training instability. A systematic study is carried out, and a two-stage GAN framework is proposed that balances layout fidelity and style expressiveness. Around three core dimensions of feature normalization, feature decoupling, and loss function, seven key innovations are designed, including multi-scale spatial adaptive normalization module, adaptive feature decoupling and reorganization module, and

collaborative optimization strategies such as soft cycle consistency loss, layout-aware loss, and boundary consistency loss. A complete technical process from layout-guided generation to fine style transfer is constructed. Experimental results show that the proposed method significantly outperforms existing mainstream baseline methods on all evaluation metrics on a dedicated landscape dataset. The layout consistency score is improved by 18.7% compared with the best baseline. The generated images show outstanding performance in layout rationality, edge sharpness, and style consistency. User study further verifies the applicability and usability of the method in real landscape design scenarios, effectively solving the application bottleneck of traditional methods in landscape scenes.

The research contribution of this paper is mainly reflected in two levels: technical level and application level. At the technical level, the proposed multi-scale spatial adaptive normalization module and adaptive feature decoupling and reorganization module break through the limitations of traditional methods in feature processing and decoupling. Their design ideas and technical details can provide important reference for other image generation and style transfer tasks in the image processing field, and enrich the application mode of GAN in professional domains. At the application level, the proposed method provides an efficient and accurate technical tool for digitalization and intelligent transformation of landscape design, which can significantly shorten design cycles and enrich design expression forms, and provides strong support for technological upgrading of the landscape design industry. Future work will focus on the limitations of the method, further optimize the model structure, introduce Transformer architecture to enhance feature extraction capability, adopt lightweight model design to improve high-resolution image generation speed, and optimize the feature decoupling module to improve adaptability of extreme style transfer, promoting the deep application and development of this technology in landscape design and related interdisciplinary fields.

REFERENCES

- [1] Lai, Y., Chou, T., Huang, L. (2025). A design-fit approach to architecture, engineering, and construction digitalization: Leveraging big data and real-scene imaging in landscape projects. *IT Professional*, 27(6): 81-86. <https://doi.org/10.1109/mitp.2025.3611525>
- [2] Ji, M., Lu, J., Zhang, X. (2022). Construction of a landscape design and greenery maintenance scheduling system based on multimodal intelligent computing and deep neural networks. *Computational Intelligence and Neuroscience*, 2022: 1-11. <https://doi.org/10.1155/2022/8307398>
- [3] Ma, B., Dong, Y., Liu, H., Cao, Z. (2021). Soft multimedia assisted new energy productive landscape design based on environmental analysis and edge-driven artificial intelligence. *Soft Computing*, 26(23): 12957-12967. <https://doi.org/10.1007/s00500-021-06155-9>
- [4] Azizi, Z., Kuo, C.C.J. (2022). Pager: Progressive attribute-guided extendable robust image generation. *arXiv preprint arXiv:2206.00162*. <https://doi.org/10.48550/arXiv.2206.00162>
- [5] Lee, T., Lee, D., Kang, M. (2025). PointT2I: LLM-based text-to-image generation via keypoints. *Neurocomputing*, 668: 132363. <https://doi.org/10.1016/j.neucom.2025.132363>
- [6] Chen, Y., Xie, Y. (2023). Neural network based multi-dimensional and nonlinear landscape design. *Journal of Computational Methods in Sciences and Engineering*, 23(3): 1279-1293. <https://doi.org/10.3233/jcm-226724>
- [7] Li, B., Sharma, A. (2022). Application of interactive Genetic Algorithm in landscape planning and design. *Informatica*, 46(3): 365-372. <https://doi.org/10.31449/inf.v46i3.4049>
- [8] Zhou, H., Zheng, H., Liu, Q., Liu, J., Wang, Y. (2021). Linear electromagnetic inverse scattering via generative adversarial networks. *International Journal of Microwave and Wireless Technologies*, 14(9): 1168-1176. <https://doi.org/10.1017/s1759078721001331>
- [9] Jin, G., Zhang, Y., Lu, K. (2019). Deep hashing based on VAE-GAN for efficient similarity retrieval. *Chinese Journal of Electronics*, 28(6): 1191-1197. <https://doi.org/10.1049/cje.2019.08.001>
- [10] Zhao, X., Chen, W., Xie, W., Shen, L. (2023). Style attention based global-local aware GAN for personalized facial caricature generation. *Frontiers in Neuroscience*, 17: 1136416. <https://doi.org/10.3389/fnins.2023.1136416>
- [11] Kong, F., Pu, Y., Lee, I., Nie, R., Zhao, Z., Xu, D., Qian, W., Liang, H. (2023). Unpaired artistic portrait style transfer via asymmetric double-stream GAN. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9): 5427-5439. <https://doi.org/10.1109/tnnls.2023.3263846>
- [12] Gui, X., Zhang, Y., Li, L. (2025). Image style migration based on cycleGAN with same mapping loss, 23-32. *International Journal of Robotics and Automation*, 40(1): 23-32. <https://doi.org/10.2316/j.2025.206-1048>
- [13] Liu, C., Gu, J., Yao, L., Zhang, Y. (2024). Research on embroidery style migration model based on texture cycle GAN. *International Journal of Clothing Science and Technology*, 37(1): 138-153. <https://doi.org/10.1108/ijcst-04-2023-0062>
- [14] Jiao, J., Xiao, X., Li, Z. (2023). Dm-GAN: Distributed multi-latent code inversion enhanced GAN for fast and accurate breast X-ray image automatic generation. *Mathematical Biosciences and Engineering*, 20(11): 19485-19503. <https://doi.org/10.3934/mbe.2023863>
- [15] Tan, H., Liu, X., Yin, B., Li, X. (2022). DR-GAN: Distribution regularization for text-to-image generation. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12): 10309-10323. <https://doi.org/10.1109/tnnls.2022.3165573>
- [16] Rashid, K.I., Yang, C., Huang, C. (2025). EPDPM-SinGAN: Enhancing urban street semantic segmentation with region-wise GANs feature. *Expert Systems with Applications*, 285: 128053. <https://doi.org/10.1016/j.eswa.2025.128053>
- [17] Hong, S., Shen, J., Lü, G., Liu, X., Mao, Y., Sun, N., Tang, L. (2023). Aesthetic style transferring method based on deep neural network between Chinese landscape painting and classical private garden's virtual scenario. *International Journal of Digital Earth*, 16(1): 1491-1509. <https://doi.org/10.1080/17538947.2023.2202422>
- [18] Pan, J., Li, F.W.B., Yang, B., Nan, F. (2025). CLPFusion: A latent diffusion model framework for realistic Chinese landscape painting style transfer. *Computer Animation*

- and Virtual Worlds, 36(3): e70053. <https://doi.org/10.1002/cav.70053>
- [19] Chen, Y. (2024). Application of style transfer algorithm in artistic design expression of terrain environment. International Journal of Advanced Computer Science and Applications, 15(1): 1094-1103. <https://doi.org/10.14569/ijacsa.2024.01501108>
- [20] Zhang, Q., Wang, S., Cui, D. (2024). Feature consistency-based style transfer for landscape images using dual-channel attention. IEEE Access, 12: 164018-164027. <https://doi.org/10.1109/access.2024.3485063>