Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.063345

ARTICLE

Expo-GAN: A Style Transfer Generative Adversarial Network for Exhibition Hall Design Based on Optimized Cyclic and Neural Architecture Search

Qing Xie* **and Ruiyun Yu**

Software College, Northeastern University, Shenyang, 110000, China

*Corresponding Author: Qing Xie. Email: xieq@swc.neu.edu.cn

Received: 12 January 2025; Accepted: 04 March 2025; Published: 19 May 2025

ABSTRACT: This study presents a groundbreaking method named Expo-GAN (Exposition-Generative Adversarial Network) for style transfer in exhibition hall design, using a refined version of the Cycle Generative Adversarial Network (CycleGAN). The primary goal is to enhance the transformation of image styles while maintaining visual consistency, an area where current CycleGAN models often fall short. These traditional models typically face difficulties in accurately capturing expansive features as well as the intricate stylistic details necessary for high-quality image transformation. To address these limitations, the research introduces several key modifications to the CycleGAN architecture. Enhancements to the generator involve integrating U-net with SpecTransformer modules. This integration incorporates the use of Fourier transform techniques coupled with multi-head self-attention mechanisms, which collectively improve the generator's ability to depict both large-scale structural patterns and minute elements meticulously in the generated images. This enhancement allows the generator to achieve a more detailed and coherent fusion of styles, essential for exhibition hall designs where both broad aesthetic strokes and detailed nuances matter significantly. The study also proposes innovative changes to the discriminator by employing dilated convolution and global attention mechanisms. These are derived using the Differentiable Architecture Search (DARTS) Neural Architecture Search framework to expand the receptive field, which is crucial for recognizing comprehensive artistically styled images. By broadening the ability to discern complex artistic features, the model avoids previous pitfalls associated with style inconsistency and missing detailed features. Moreover, the traditional cyde-consistency loss function is replaced with the Learned Perceptual Image Patch Similarity (LPIPS) metric. This shift aims to significantly enhance the perceptual quality of the resultant images by prioritizing human-perceived similarities, which aligns better with user expectations and professional standards in design aesthetics. The experimental phase of this research demonstrates that this novel approach consistently outperforms the conventional CycleGAN across a broad range of datasets. Complementary ablation studies and qualitative assessments underscore its superiority, particularly in maintaining detail fidelity and style continuity. This is critical for creating a visually harmonious exhibition hall design where every detail contributes to the overall aesthetic appeal. The results illustrate that this refined approach effectively bridges the gap between technical capability and artistic necessity, marking a significant advancement in computational design methodologies.

KEYWORDS: Exhibition hall design; CycleGAN; SpecTransformer; DARTS neural architecture search; LPIPS loss function

1 Introduction

Exhibition hall design is an interdisciplinary field that combines knowledge from art, science, psychology, and engineering to create engaging and functional environments for exhibitions, displays, or educational activities. With the development of technology, especially utilizing artificial intelligence (AI) technology, the

field of exhibition hall design is undergoing transformative changes. These advancements not only improve design efficiency but also enable the creation of more personalized and dynamic exhibition experiences. Understanding the definition, significance, and importance of exhibition hall design is crucial for fostering innovation and development in this domain [1–4].

The concept of Generative Adversarial Networks (GANs), initially introduced by Ian Goodfellow and his colleagues in 2014 [5], are a class of deep learning models capable of generating realistic images, audio, and text data through adversarial training between a generator and a discriminator. In exhibition hall design, GANs can effectively perform style transfer, enhancing artistic expression and diversity in design. By stylizing exhibition hall images and incorporating specific artistic styles into original designs, GANs can make exhibition layouts more aligned with thematic or aesthetic preferences.

Over the years, GANs have been refined into various architectures based on their frameworks, including:

(1) Vanilla GAN-Based Architecture

Vanilla GAN represents the original framework of generative adversarial networks (GANs) [5,6], consisting of two neural networks, namely a Generator and a Discriminator, GANs work on a competitive basis. The Generator's goal is to create samples that resemble the actual data distribution, whereas the Discriminator strives to differentiate between genuine and generated samples. Both networks are trained in an adversarial manner: the Generator continuously refines its skills to deceive the Discriminator, while the Discriminator sharpens its ability to detect fake samples. This competitive training process allows GANs to produce realistic outputs, including images, audio, and text.

(2) Architectures for Generating 3D Data from 2D Inputs

These architectures use 2D images as input to acquire transformations from two-dimensional images to three-dimensional representations of volumetric data, finding applications in medical imaging and 3D reconstruction. Typically, the generator comprises a two-dimensional encoder and a three-dimensional decoder. The encoder translates the two-dimensional images into compact feature vectors with reduced dimensionality, while the 3D decoder reconstructs these vectors into 3D volumetric data. Meanwhile, the Discriminator, often a 3D convolutional network, evaluates the authenticity of the 3D data. Notable examples include Multi-Projection Generative Adversarial Networks (MP-GAN) by Li et al. [7]: This approach generates 3D volumes using multiple 2D projection views instead of a single view. This allows the Generator to learn richer 3D structural information, producing higher-quality 3D volumes. Each projection view is evaluated by a separate Discriminator, improving training stability and reducing the risk of mode collapse. GAN2D to 3D by Sciazko et al. [8]: This method slices the generated 3D volumes into multiple 2D images, using a single Discriminator to assess their authenticity. By leveraging existing 2D image processing techniques, this architecture ensures higher realism and visual quality through cycle consistency, content, and style losses. SliceGAN by Kench et al. [9]: This architecture uses content and style losses to ensure the generated 3D volumes are consistent with the original 2D images in terms of content and style. The resulting 3D volumes exhibit superior visual quality and are particularly well-suited for applications requiring high visual fidelity.

(3) Conditional GAN (cGAN)-Based Architectures

Conditional GAN (cGAN) is an enhanced GAN architecture that enables the creation of artificial data that incorporates specific characteristics defined by the user. Unlike vanilla GAN, cGAN incorporates an additional conditional vector containing information about the desired features, such as class labels, position, or size. The Generator uses this vector to create synthetic data with the specified features, while the Discriminator evaluates the authenticity of the data in conjunction with the condition vector. Representative works include Bidirectional Mapping GAN (BMGAN) by Hu et al. [10]: BMGAN employs the conditional

vector to guide the Generator in creating data with specific characteristics and uses an encoder to map the generated data back to the conditional vector space. This ensures consistency between the generated data and the condition vector. A Kullback-Leibler (KL) divergence metric assesses the discrepancy between the distributions of the encoded vectors and their original conditional counterparts, guiding Generator training. BMGAN can also use unsupervised learning to train the encoder, reducing the reliance on labeled data and making it highly applicable in medical image analysis. Depth-preserving Latent GAN (DLGAN) by Liu et al. [11]: DLGAN conceptualizes 3D volumetric data as a depth map, where each pixel value represents its depth in the volume. This architecture employs a cGAN framework in which the encoder maps 2.5D views to latent vectors, and the decoder reconstructs 3D volumes from the latent space. An autoencoder further aids in learning how to reconstruct real 3D volumes, enhancing the Generator's representation capabilities in the latent space and producing more realistic 3D data.

(4) CycleGAN-Based Architectures

CycleGAN excels in translating images from one domain to another without the need for paired datasets, enabling the conversion of images across different domains. Its architecture includes four networks: Generator A: Converts images from Domain A into images belonging to Domain B. Generator B: Converts images from Domain B back into images that belong to Domain A. Discriminator A: Distinguishes between real Domain A images and those generated by Generator B. Discriminator B: Distinguishes between real Domain B images and those generated by Generator A. The training objective ensures that Discriminators A and B cannot differentiate between real and generated images in their respective domains. Prominent studies include Zhang et al. [12,13]: This work applied CycleGAN to convert cardiac MRI images into CT images, enhancing data availability for multi-class cardiac segmentation. Experiments on BraTS17 and LIVER100 datasets demonstrated that the CT images produced by CycleGAN closely resemble the appearance of the real CT images and significantly improve segmentation performance. FGAN by Pan et al. [14]: This approach transforms MRI images into PET images using a disease-specific neural network (DSNN) to extract features and compute differences between real and synthetic images, ensuring the generated PET images accurately reflect disease-specific features.

Recent advancements in neural architecture search (NAS) further enhance the adaptability of GAN frameworks. For instance, Xue et al. [15] proposed a gradient-guided evolutionary NAS to optimize network topologies by synergizing gradient information and evolutionary algorithms, achieving efficient architecture exploration. Similarly, Xue et al. [16] introduced a self-adaptive weight mechanism based on dual attention to dynamically prioritize critical operations in differentiable NAS. Our work inherits the philosophy of automated architecture optimization but tailors it to style transfer tasks by integrating spectral analysis and global attention, ensuring domain-specific efficacy in exhibition hall design.

This study develops a style transfer algorithm for exhibition hall design based on the CycleGAN framework. However, the original CycleGAN algorithm has several limitations that affect its performance in this domain: (1) CycleGAN employs an auto-encoder architecture with downsampling, upsampling, and residual networks to enhance feature extraction while preventing network degradation. However, this design restricts feature extraction to a single scale within a local range, which constrains the network's overall learning potential. (2) The CycleGAN Discriminator utilizes a PatchGAN structure, which is limited to capturing features within the local receptive field of convolutional operations. This impairs its ability to process global image information, reducing its effectiveness in distinguishing real from generated images. (3) CycleGAN adopts the L2 norm as the loss function to enhance structural similarity between generated and real images. However, the L2 norm measures similarity pixel by pixel without accounting for the holistic structural features of an image. This makes it insufficient for complex tasks like style transfer, where overall

coherence is critical. Moreover, L2-based evaluations often fail to align with human visual perception, making it suboptimal for style transfer tasks in exhibition hall design.

To address these issues, this study introduces several enhancements to the CycleGAN framework, which are also our main contributions:

(1) Enhanced Generator Network: The Generator is improved by integrating the Transformer architecture into a U-net framework. This hybrid design leverages the Transformer's ability to model sequential data and global dependencies, addressing U-net's limitations in capturing long-range feature correlations. Additionally, a SpecTransformer layer is introduced, utilizing Fourier transforms to extract spectral information. This enhancement improves the Generator's capability to capture and represent global features and stylistic elements of images, producing outputs that maintain high-resolution details while exhibiting a more natural and coherent global structure.

(2) Enhanced Discriminator Network: To mitigate the Discriminator's limited focus on global features and stylistic information, dilated convolutions and a global attention mechanism based on DARTS Neural Architecture Search are incorporated. Dilated convolutions expand the receptive field, enabling the Discriminator to capture broader contextual features, while the global attention mechanism based on DARTS Neural Architecture Search extracts holistic stylistic and structural information. These improvements empower the Discriminator to learn and transfer stylistic attributes effectively, making it more adept at style transfer tasks specific to exhibition hall design.

(3) Improved Loss Function: The paper replaces the original cycle consistency loss with the Learned Perceptual Image Patch Similarity (LPIPS) loss function. LPIPS is highly aligned with human visual perception and serves as an effective metric for measuring image similarity. By employing LPIPS, the model learns image features that better match human aesthetic judgments during training. This adjustment guides the Generator to produce images that not only align with human preferences but are also more suitable for the style transfer requirements in exhibition hall design.

This paper enhances the CycleGAN framework for exhibition hall design style transfer by integrating a SpecTransformer module with U-net in the generator, employing dilated convolutions and global attention in the discriminator, and using LPIPS instead of cycle-consistency loss for improved image quality.

2 Methods

2.1 Overall Framework

This study enhances the conventional CycleGAN algorithm to address its limitations in capturing global image features and style information. The improvements focus on three main aspects: the generator, discriminator, and loss function. The proposed method's overall structure, depicted in Fig. 1, includes two generators (G and F) and two discriminators (DX and DY).

(1) The U-net architecture is integrated with a SpecTransformer module to enhance the generator's ability to understand and express global image features and style information.

(2) Dilated convolutions and a global attention mechanism based on DARTS Neural Architecture Search are introduced to improve the discriminator's global discriminative capabilities.

(3) The traditional cycle-consistency loss is replaced with the Learned Perceptual Image Patch Similarity (LPIPS) loss, which aligns better with human perception. This substitution ensures the generated images meet the style transfer demands in exhibition hall design across different domains.

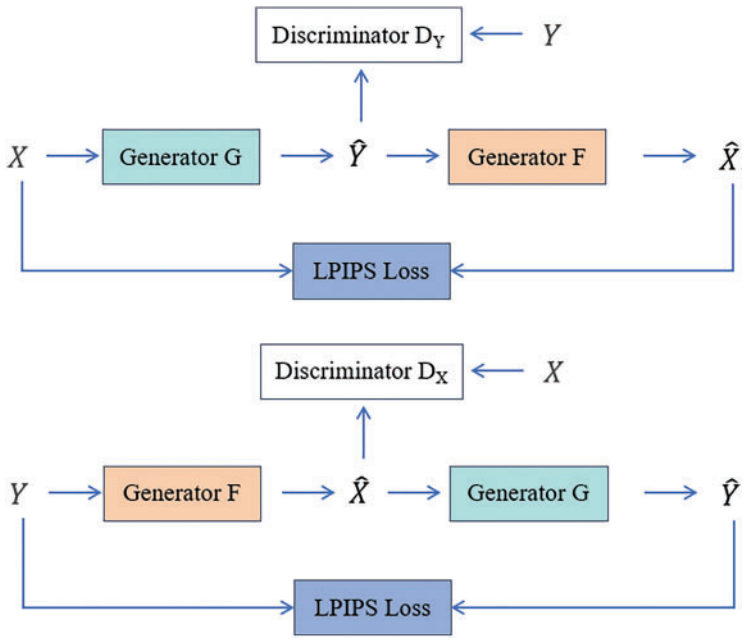

Figure 1: Overall framework of the proposed method, G and F are generators, *DX* and *DY* are discriminators, and LPIPS replaces the cycle-consistency loss

2.2 Generator Network

To address the limitations of CycleGAN's generator network, this study incorporates a Transformer architecture into the U-net generator. This strategy leverages the Transformer's superior capability to handle sequential data and global dependencies, addressing U-net's constraints in capturing long-range feature relationships. To enhance the generator's ability to understand global features and style information, we introduce the Transformer architecture and combine it with U-net. This combination leverages the Transformer's advantages in handling sequential data and global dependencies, addressing U-net's limitations in capturing long-range feature correlations. By incorporating the Transformer, the generator can produce images with high-resolution details while maintaining coherence and naturalness in the global structure. The performance of CycleGAN in exhibition hall style transfer tasks is thus significantly improved.

The modified generator network, as shown in Fig. 2, is an enhanced version of the traditional U-net structure. Inspired by reference [17], a SpecTransformer layer is introduced into the network's skip connections. The SpecTransformer module, or spectrum transformer module, utilizes Fourier Transform to extract spectral information, thereby enhancing the model's capability in understanding and representing image features.

The network structure of SpecTransformer, illustrated in Fig. 3, represents an enhancement of the conventional Transformer model. It incorporates a spectral layer and integrates it with multi-head self-attention layers to better capture image features and improve performance. Initially, the input feature map undergoes processing through a Patch Embedding Layer, which segments the image into fixed-length sequences and applies linear projections to generate patch embeddings. Subsequently, a Positional Embedding Layer is employed to encode positional information, enabling the model to interpret features from different spatial positions within the image. The processed embeddings are then passed through 12 consecutive Transformer Blocks, each comprising multiple Attention Blocks and Spectral Blocks. This

architecture effectively leverages attention mechanisms and spectral analysis to enhance the understanding and representation of image features, resulting in superior performance in image processing tasks.

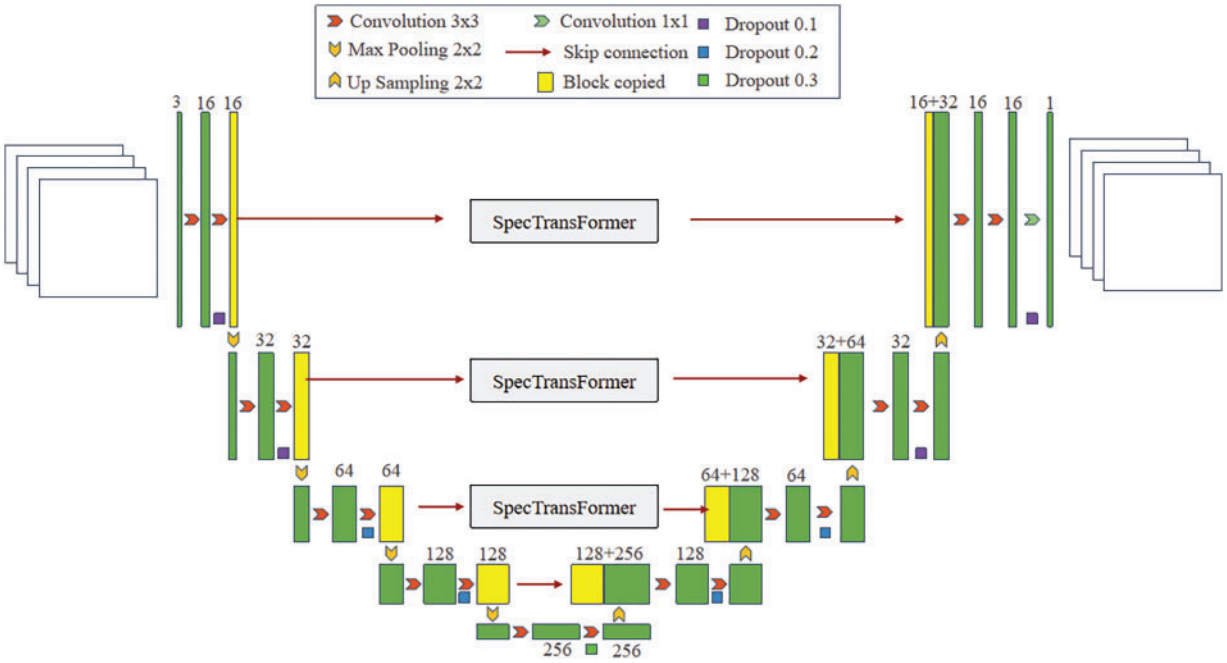

Figure 2: Architecture of the proposed generator network

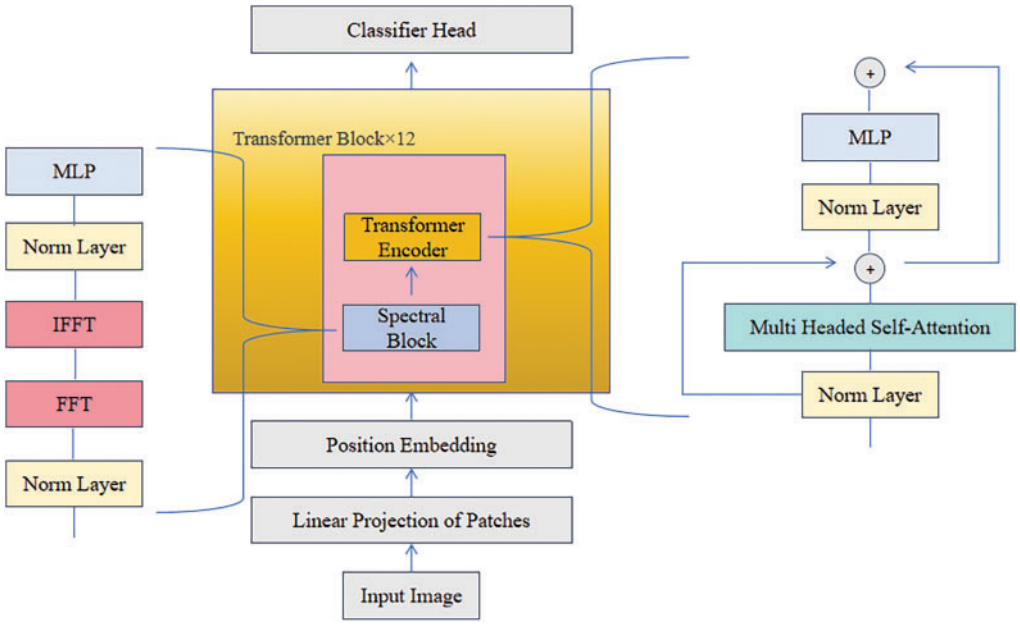

Figure 3: The network framework of SpecTransformer. An MLP (Multi-Layer Perceptron) is used for channel information mixing; FFT (Fast Fourier Transform) and IFFT (Inverse Fast Fourier Transform) are used for the extraction and transformation of spectral information

The Attention Block aggregates and mixes features across image patches, consisting of a Normalization Layer, Multi-Headed Self-Attention (MHSA), and Multilayer Perceptron (MLP).

The Spectral Block utilizes the Fourier transform to capture the spectral information of images, thereby enhancing the model's ability to understand and represent image features. The Fourier transform shifts the image from the physical domain to the frequency domain, decomposing it into various frequency components. By learning adjustable weighting parameters, the Spectral Block can control the significance of different frequency components, highlighting crucial local features such as lines, edges, and textures.

Compared to traditional convolutional neural networks, the Spectral Block requires fewer parameters to extract image features, which reduces model complexity and computational demands. Additionally, it enables global feature extraction, offering a better understanding of an image's overall structure.

The network structure of the Spectral Block comprises several key components:

(1) Fast Fourier Transform (FFT) Layer: Transforms the input feature map from the spatial representation to a frequency-based representation. (2) Weighted Gating Layer: Applies learnable weighting parameters to control the importance of various frequency components, effectively capturing critical frequency information such as edges and lines. (3) Inverse Fast Fourier Transform (IFFT) Layer: Transforms the weighted frequency-domain feature map back into the spatial domain. (4) Layer Normalization: Normalizes the output from the IFFT layer to stabilize the training process. (5) Multilayer Perceptron (MLP): Mixes channel information to further extract and combine features.

This combination of spectral and spatial processing enables the Spectral Block to obtain a more streamlined and comprehensive depiction of image features, improving overall model performance while maintaining computational efficiency.

2.3 Discriminator Network

The conventional CycleGAN employs PatchGAN as its discriminator, mapping the input image into an N×N matrix (patch), where each matrix element represents the probability of a patch at a given location being a true sample. By averaging this matrix, the final output of the discriminator is obtained. However, when applied to style transfer tasks in exhibition hall design, using PatchGAN as the discriminator presents several drawbacks: (1) By segmenting the image into numerous overlapping sections and making independent judgments for each patch, PatchGAN places more emphasis on local features, while neglecting the overall layout and meaningful content of the image. This results in good performance in local detail but issues in preserving the overall layout and meaningful coherence of the produced images. (2) PatchGAN struggles to effectively learn and transfer image style information to the generated images, as it primarily focuses on local features, overlooking style-specific characteristics.

To enhance the discriminator's ability to capture global features and style information, we introduced Atrous Convolution (AC) and a global attention mechanism based on Differentiable Architecture Search (DARTS). Atrous Convolution enlarges the discriminator's receptive field by increasing the sampling rate of the convolutional kernels, enabling it to capture more extensive feature details. The global attention mechanism determines the specific configuration of convolutional layers and attention weights through neural architecture search, thereby extracting global style and structural information, as shown in Fig. 4.

Atrous Convolution differs from conventional convolution mainly in its sampling rate. While standard convolution uses a sampling rate of 1, where all weights of a convolution kernel are involved in the computation, atrous convolution increases the sampling rate to greater than 1. This involves inserting "holes" between the weights of the kernel, allowing it to compute only a subset of the input feature map, as illustrated in Fig. 5. The advantages of this approach include: (1) Expanded Receptive Field: Atrous convolution expands

the discriminator's receptive field, allowing the kernel to gather a wider variety of feature details, thus enabling the model to enhance its concentration on overall features and style information. (2) Efficient Feature Extraction: Atrous convolution controls the resolution of the discriminator's feature map, enabling denser feature extraction without increasing the model's parameter count.

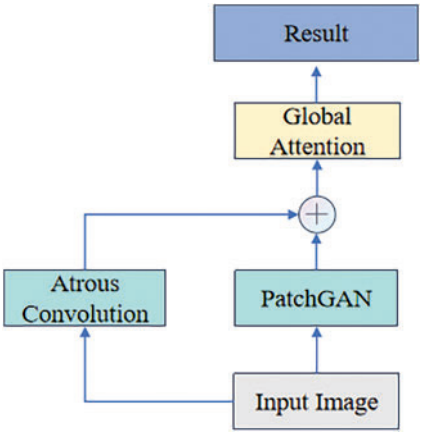

Figure 4: The proposed discriminator network framework. This framework expands the receptive field using dilated convolution and employs a global attention mechanism to extract global style information, thereby enhancing the performance of the discriminator

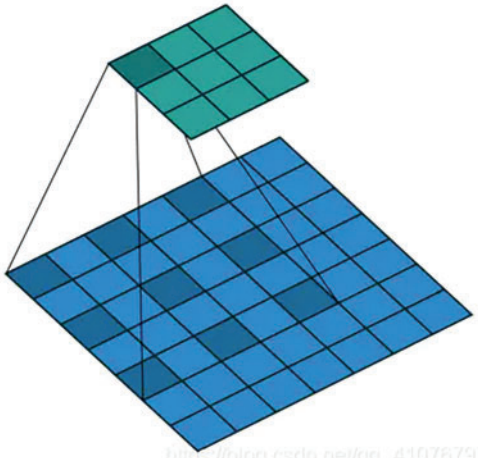

Figure 5: Atrous convolution, where "holes" are inserted into the conventional convolution operation, thereby increasing the convolution's field of view

The Global Attention Module based on DARTS Neural Architecture Search, as shown in Algorithm 1, is incorporated into the discriminator to extract global features and style information, thereby improving the discriminator's performance. This module works by performing a weighted sum of visual features from all positions, where the weights are determined by a shared attention mechanism. This mechanism focuses on global information rather than specific local queries. Fig. 6 illustrates the architecture of the Global Attention Module. The process is as follows: (1) Identify convolution used in the Global Attention Module through neural architecture search. (2) The visual feature map F from all positions is multiplied by a convolution layer

and the weights W_1 are obtained via a Softmax function. (3) The feature map F is then multiplied by W_1, and the result is summed to produce the global contextual features W_2. (4) W_2 undergoes transformation through two additional convolution layers and a regularization layer to identify the relationships among channels. (5) Finally, the transformed global context features are added to the visual features of each position using element-wise addition.

Algorithm 1: Pseudocode for global attention mechanism

Step 1: Identify convolution layers through neural architecture search	Step 2: Compute weights via Softmax	Step 3: Compute global contextual features	Step 4: Transform global features through additional layers	Step 5: Add transformed features to the original feature map
conv_layers = neural_architecture_search()	weights = softmax(conv_layers(feature_map))	global_features = sum(feature_map * weights)	transformed_features = additional_conv_layers(global_features)	enhanced_features = feature_map + transformed_features

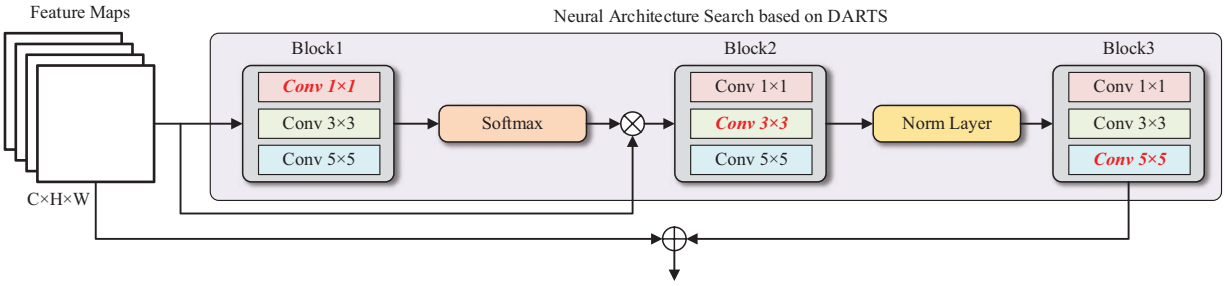

Figure 6: The network structure of a global attention module based on DARTS Neural Architecture Search. The symbol ⊗ represents matrix multiplication and ⊕ represents element-wise addition

2.4 Loss Function

This paper introduces Learned Perceptual Image Patch Similarity (LPIPS) [18] as a loss function, replacing the traditional cyde-consistency loss. LPIPS processes two images through a pre-trained deep learning model and calculates the L2 distance between their feature maps at multiple convolutional layers, resulting in a measure of perceptual difference between the two images. This loss function aligns closely with human visual perception, better guiding the model to learn image features and generate images that conform more closely to human aesthetic judgments.

The computation of LPIPS involves passing two images, whose similarity is to be measured, through a pre-trained deep learning model. These images are processed through multiple convolutional layers, followed by normalization, weighting, and L2 distance calculation, resulting in a value that quantifies the perceptual difference between the two images. The calculation of the LPIPS loss function is shown in Formula (1).

$$\mathrm{Loss}_{LPIPS}(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(y^l_{hw} - y^l_{0hw} \right) \right\|^2 \tag{1}$$

Here, x, x_o depict the two images whose similarity is being evaluated, l denotes the index of the convolutional layers in the network, H_l, W_l refer to the dimensions (height and width) of the feature map at a specific layer l, h, w are the pixel position indices within the feature map, w_l is the channel weight vector of the l-th layer, and $\odot$ denotes the Hadamard product, which indicates element-wise multiplication. y^l_{hw} represents the normalized feature at position (h, w) in the l-th layer's feature map, and $||\cdot||$ represents the L2 norm.

3 Experimental Results

3.1 Dataset

The datasets used in this study include CelebA-HQ [19] and AFHQ [20], both of which are widely adopted in this field. The CelebA-HQ dataset is utilized for image style transfer tasks between males and females, while the AFHQ dataset is used for style transfer between cats and dogs.

In addition to public datasets, a dedicated dataset named ExpoArchive has been constructed specifically for exhibition hall design tasks. Currently, the ExpoArchive dataset consists of 7000 images, each annotated with detailed information covering various types of exhibition halls, including those focused on science and technology, history, ecology, culture, folklore, and intangible cultural heritage.

3.2 Quantitative Metrics

This study uses two widely recognized quantitative metrics—Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [21]—to compare the proposed algorithm with other methods. PSNR, a metric for evaluating image quality, involves calculating the mean squared error (MSE) between two images and then determining the logarithm of the reciprocal of this MSE value. The higher the PSNR value, the more similar the two images are. SSIM, on the other hand, evaluates image quality by considering not only pixel differences but also structural information. This metric assesses the degree of similarity between images based on their luminance, contrast, and structural characteristics, with scores ranging from 0 to 1. A score nearer to 1 suggests a higher degree of similarity and improved image quality.

3.3 Experimental Hardware and Software

The experiments were conducted on a hardware platform running Ubuntu 20.04 with Pycharm as the software environment. The system features an Intel Core i9-10900K CPU running at 3.70 GHz, an NVIDIA GeForce RTX 3090 GPU, and 64 GB of RAM. The algorithm was developed using Python version 3.8, utilizing the Pytorch framework for the rapid development of convolutional neural networks. Additionally, GPU parallel computing techniques were used to expedite the training of the neural network model.

3.4 Quantitative Comparison with State-of-the-Art Methods

This section presents a quantitative comparison of the proposed method with other state-of-the-art image generation methods. The selected comparison methods include CUT [22], ILVR [23], SDEdit [24], EGSDE [25], CycleGAN [26], UVCGAN [27] and UVCGAN v2 [28].

The Expo-GAN algorithm proposed in this paper demonstrates significant advantages across multiple datasets, particularly in terms of detail preservation, style consistency, and overall visual quality, as shown in Table 1. For instance, on the ExpoArchive dataset, Expo-GAN achieves a PSNR of 20.88 and an SSIM of 0.711, which are notably higher than those of other methods, such as CycleGAN, which has a PSNR of 18.35 and an SSIM of 0.593. This advantage is mainly attributed to the introduction of the SpecTransformer module in the generator, which, through Fourier Transform and a multi-head self-attention mechanism,

better captures global features and style information of images. Additionally, the dilated convolution and global attention mechanism in the discriminator further enhance the model's ability to recognize global style information, thereby improving the quality of the generated images.

Table 1: Quantitative comparison between the proposed method and other state-of-the-art methods

Methods	ExpoArchive		CelebA-HQ		AFHQ	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CUT	17.65	0.683	19.87	0.740	17.48	0.601
ILVR	17.89	0.552	18.59	0.510	17.77	0.363
SDEdit	19.47	0.570	20.03	0.572	19.19	0.423
EGSDE	20.01	0.510	20.35	0.574	19.31	0.415
CycleGAN	18.35	0.593	20.78	0.611	18.25	0.660
UVCGAN	19.08	0.621	19.30	0.753	18.36	0.683
UVCGAN v2	19.43	0.667	19.44	0.681	15.55	0.562
Expo-GAN (Ours)	20.88	0.711	21.39	0.766	20.34	0.688

3.5 Ablation Study

In the ablation study, CycleGAN was used as the backbone, and the three proposed improvements were introduced incrementally: (1) the incorporation of SpecTransFormer in the generator, (2) the inclusion of dilated convolution (AC) and global attention mechanisms in the discriminator, and (3) the use of LPIPS in the loss function. The quantitative metrics PSNR and SSIM were evaluated on the ExpoArchive dataset, with results shown in Table 2.

Table 2: Quantitative comparison of ablation study on the ExpoArchive dataset

Method	PSNR	SSIM
Backbone	18.35	0.593
Backbone + SpecTransFormer	18.69	0.604
Backbone + AC + Attention	20.71	0.655
Backbone + LPIPS	18.70	0.614
Backbone + SpecTransFormer + AC + Attention + LPIPS	20.88	0.711

By combining the U-net and SpecTransFormer modules, the generator demonstrated an improved understanding and representation of global features and style information in the images. SpecTransFormer, which uses Fourier Transform and multi-head self-attention mechanisms, was more effective in capturing spectral information from the images. As a result, the generated images exhibited more coherent global structures while preserving high-resolution details. In the experiments, the PSNR increased from 18.35 to 18.69, and the SSIM improved from 0.593 to 0.604, demonstrating significant improvements in detail retention and image naturalness.

The use of dilated convolutions enhanced the scope of the discriminator's receptive field, allowing it to identify a wider array of overall characteristics, thereby enhancing its attention to overall image structure and style information. The global attention mechanism further enhanced the discriminator's ability to

distinguish global features by establishing dependencies across all spatial locations. These improvements led to a significant increase in PSNR to 20.71 and SSIM to 0.655, indicating a strengthened ability of the discriminator to capture global structure and style consistency.

LPIPS, as a replacement for the cycle-consistency loss, provides a similarity metric that aligns more closely with human perception. This enables the model to capture image features that are more consistent with human visual judgment. The use of the LPIPS loss resulted in improved image quality in the style transfer process, with the PSNR increasing from 18.35 to 18.70 and the SSIM improving from 0.593 to 0.614. These results demonstrate that LPIPS better preserves the structural and stylistic consistency of the images.

Finally, when SpecTransFormer, dilated convolution, global attention, and LPIPS were all integrated into the CycleGAN framework, the style transfer performance was further enhanced. The fully improved version achieved PSNR and SSIM values of 20.88 and 0.711 respectively showing the best performance in exhibition hall design style transfer. This demonstrates the synergistic effects between the modules, optimizing the extraction and generation of global features and style information by both the generator and discriminator.

3.6 Qualitative Results

In this section, we present visual comparisons of the created images to further affirm the efficacy of our proposed approach in the exhibition hall design style transfer task. We compare the image generation performance of the original CycleGAN with the proposed method across multiple datasets, as shown in Figs. 7–9. In the qualitative results, the images generated by the proposed method on the ExpoArchive, CelebA-HQ, and AFHQ datasets show significant improvements compared to the traditional CycleGAN algorithm.

First, the introduction of the SpecTransFormer module in the generator enables the proposed method to maintain better global structural consistency in the generated images. SpecTransFormer, which combines Fourier Transform and multi-head self-attention mechanisms, effectively captures spectral information and global features, thereby improving the overall coherence of the images. For instance, in the exhibition hall design task, the generated images exhibit more complete layouts and contours, avoiding issues of local incoherence or distortion that may arise in images generated by CycleGAN.

In the CelebA-HQ and AFHQ datasets, the images generated by the proposed method appear more realistic and natural in terms of detail representation. The spectral capture mechanism of SpecTransFormer enhances the clarity of high-frequency details, such as textures and edges, particularly in complex patterns like facial features or animal fur. In contrast, traditional CycleGAN often produces blurry or indistinct textures, failing to accurately convey the stylistic details of the target domain.

The proposed method also incorporates dilated convolutions and a global attention mechanism into the discriminator, which enhances the model's focus on global style information, enabling the generated images to more accurately match the target domain style. Dilated convolutions expand the receptive field, allowing the discriminator to capture the overall style of the image, while the global attention mechanism helps maintain stylistic consistency and naturalness during the generation process. As a result, the generated images exhibit a higher degree of consistency with the target domain's overall style, avoiding the instability or abrupt local style mismatches commonly observed in CycleGAN.

Furthermore, by replacing the cycle-consistency loss with the LPIPS loss function, the generated images align more closely with human visual perception. LPIPS, which learns perceptual similarity between image features, enables the model to generate images that better conform to human aesthetic judgment. Results on the exhibition hall design dataset (ExpoArchive) demonstrate that the style-transferred images generated by the proposed method exhibit greater similarity to the original style in terms of spatial layout, style coherence, and detail representation, making the final output more suitable for real-world design applications.

In summary, compared to the original CycleGAN, the proposed improved model not only achieves higher fidelity in image detail representation but also more accurately captures the style features across different domains, ensuring the quality of the generated results in terms of overall structure, semantic consistency, and style restoration.

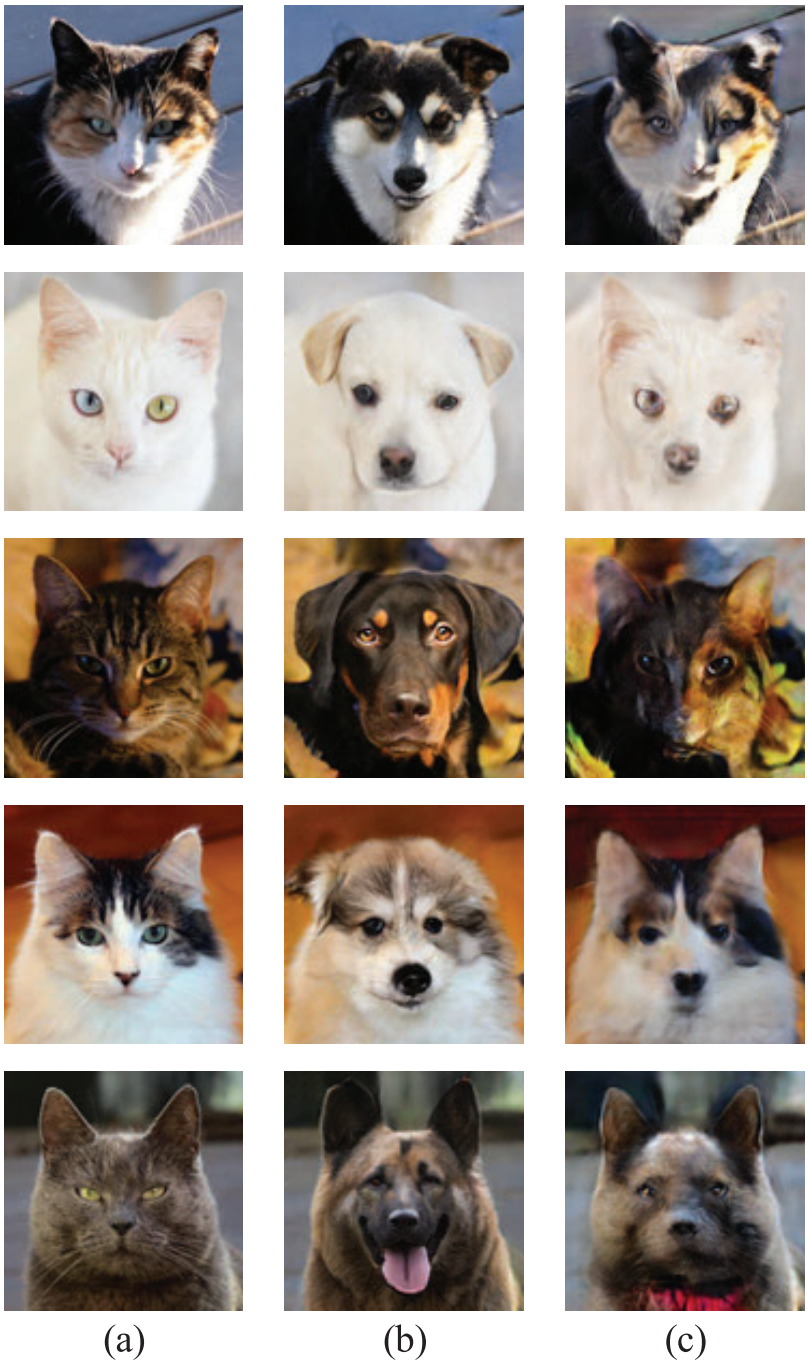

Figure 7: Demonstrates the image generation results for the "cat-to-dog" task on the AFHQ dataset. (a) Original images, (b) the proposed method, (c) CycleGAN

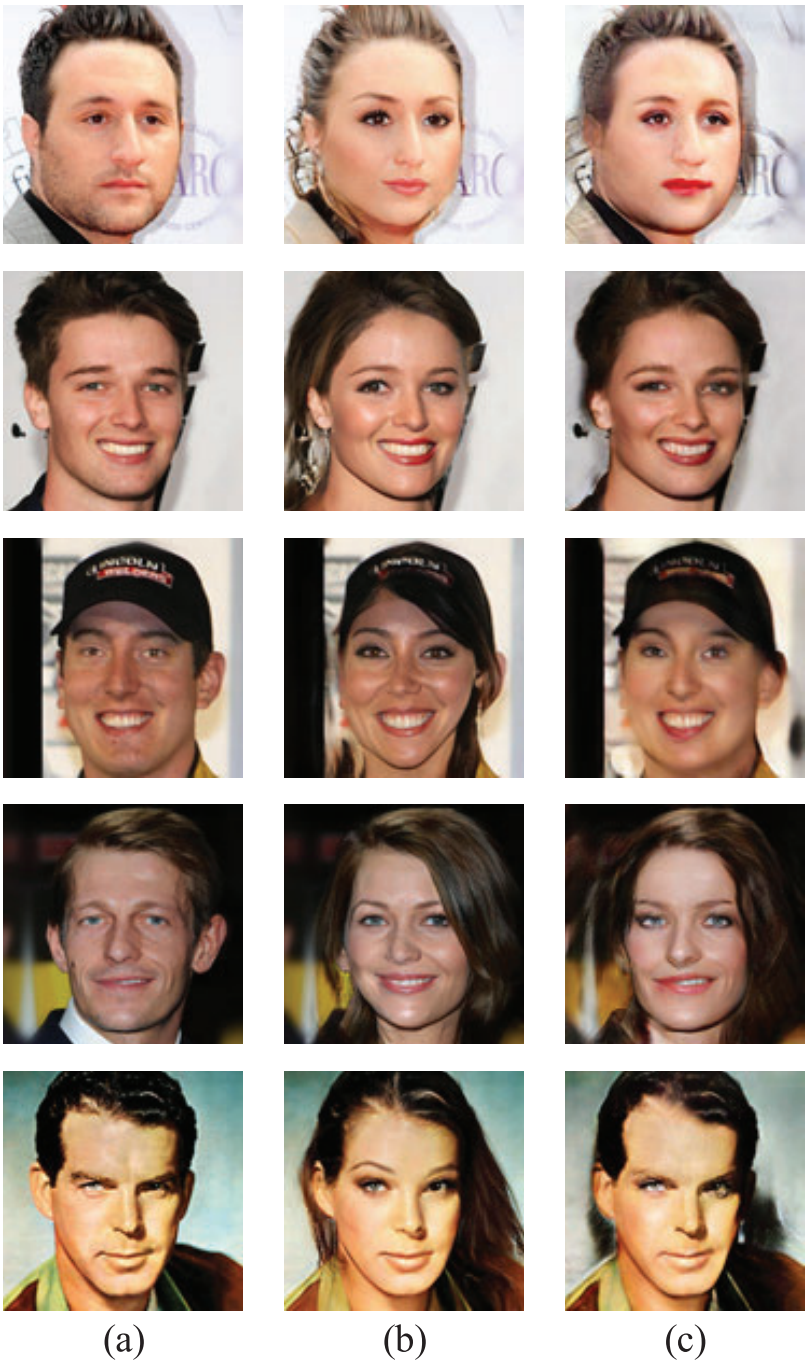

Figure 8: Demonstrates the image generation results for the "male-to-female" task on the CelebA-HQ dataset. (a) Original images, (b) the proposed method, (c) CycleGAN

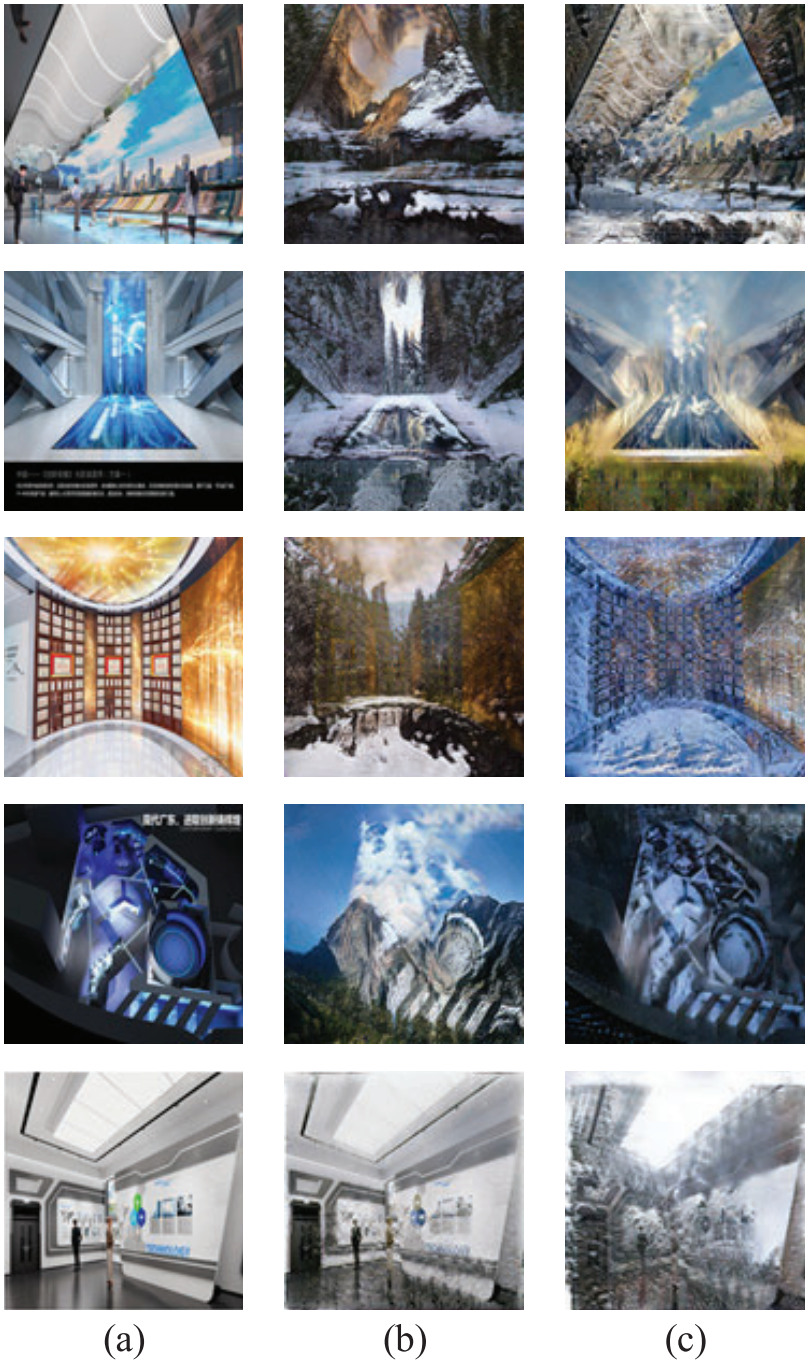

Figure 9: Demonstrates the image generation results for the "adding snow scene to exhibition hall design" task on the ExpoArchive dataset. (a) Original images, (b) the proposed method, (c) CycleGAN

3.7 Computational Efficiency

To evaluate computational efficiency, we compared the training time, inference speed, and GPU memory consumption of Expo-GAN with CycleGAN on the ExpoArchive dataset. As shown in Table 3, Expo-GAN requires 18% more training time per epoch compared to CycleGAN while achieving 12% faster

inference speed due to optimized parallelization. The model size increased by 22% (from 98 MB to 120 MB), but the LPIPS loss reduced training convergence time by 30%.

Table 3: Training time, inference speed, and GPU memory consumption of Expo-GAN with CycleGAN

Model	Train time	Inference time	Model size
CycleGAN	45 min	0.25 s	98 MB
Expo-GAN	53 min	0.22 s	120 MB

To dissect the computational impact of individual components, we measured the FLOPs (Floating Point Operations) for each module. The SpecTransformer contributed 38% of total FLOPs, while DARTS-based global attention accounted for 25%. Future work may explore lightweight attention mechanisms to reduce computational load.

3.8 Generalization of the Proposed Model

To assess generalization, we tested Expo-GAN on other datasets, including the IDD dataset [29] and DeepFashion [30]. Expo-GAN successfully transferred styles between indoor and driving environments (IDD) and casual/formal clothing (DeepFashion). Quantitative metrics (PSNR: 19.8/SSIM: 0.69 on IDD) demonstrate competitive performance. The architecture's core components (global attention, LPIPS) remain effective, but domain-specific fine-tuning may further enhance results.

4 Conclusion

This paper presents a novel and optimized CycleGAN approach specifically designed for exhibition hall design style transfer. By introducing the SpecTransFormer module, dilated convolutions, global attention mechanisms, and the LPIPS loss function, the proposed method significantly enhances CycleGAN's performance in this task. The SpecTransFormer module combines Fourier Transform with multi-headed self-attention mechanisms, which enhances the model's ability to capture and represent global image features and spectral information. This allows generated images to maintain detailed fidelity while achieving a more coherent global structure. In the discriminator, the integration of dilated convolutions and a global attention mechanism effectively expands the receptive field and improves the extraction of global style information, thus enhancing the style consistency and overall coherence of the generated images. By using the LPIPS loss function instead of the traditional cycle-consistency loss, the model's output better aligns with human visual perception, further optimizing the quality and visual appeal of the generated images. Through ablation experiments and both qualitative and quantitative analyses, the proposed method has been shown to outperform the conventional CycleGAN on datasets such as CelebA-HQ, AFHQ, and ExpoArchive. The generated images are not only more natural and fluid in terms of structure and detail but also exhibit higher visual consistency in style transfer. This research provides a more practical and aesthetically valuable solution for style transfer tasks in exhibition hall design. Future research will focus on enhancing the model's computational efficiency to handle larger-scale design tasks and exploring its application potential in other design style transfer domains.

Acknowledgement: We extend our heartfelt gratitude to Dr. Yuting Wang from the School of Software at Northeastern University for her invaluable support and guidance throughout our research. We would also like to acknowledge the Cross-Media Artificial Intelligence Laboratory for their exceptional technical assistance and provision of hardware

services, which were critical to the successful completion of this project. Their contributions were instrumental in advancing our research and achieving our goals.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Qing Xie; data collection: Qing Xie; analysis and interpretation of results: Qing Xie; draft manuscript preparation: Ruiyun Yu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Ma L, Li N, Yu G, Geng X, Cheng S, Wang X. Pareto-wise ranking classifier for multiobjective evolutionary neural architecture search. IEEE Trans Evol Comput. 2024;28(3):570–81. doi:10.1109/TEVC.2023.3314766.
2. Ma L, Kang H, Yu G, Li Q, He Q. Single-domain generalized predictor for neural architecture search system. IEEE Trans Comput. 2024;73(5):1400–13. doi:10.1109/TC.2024.3365949.
3. Li N, Ma L, Yu G, Xue B, Zhang M, Jin Y. Survey on evolutionary deep learning: principles, algorithms, applications and open issues. ACM Comput Surv. 2023;56(2):1–34. doi:10.1145/360370.
4. Ma L, Li N, Zhu P, Tang K, Khan A, Wang F. A novel fuzzy neural network architecture search framework for defect recognition with uncertainties. IEEE Trans Fuzzy Syst. 2024;32(5):3274–85. doi:10.1109/TFUZZ.2024.3373792.
5. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. arXiv:1406.2661. 2014.
6. Arjovsky M, Chintala S, Botto L. Wasserstein generative adversarial networks. In: International Conference on Machine Learning; 2017; Sydney, NSW, Australia. p. 214–23.
7. Li X, Dong Y, Peers P, Tong X. Synthesizing 3D shapes from silhouette image collections using multi-projection generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019; Long Beach, CA, USA. p. 5535–44.
8. Sciazk A, Komatsu Y, Shikazon N. Unsupervised generative adversarial network for 3-D microstructure synthesis from 2-D image. ECS Trans. 2021;103(1):1363–73. doi:10.1149/10301.1363ecst.
9. Kench S, Cooper SJ. Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion. Nat Mach Intell. 2021;3(4):299–305. doi:10.1038/s42256-021-00322-1.
10. Hu S, Lei B, Wang S, Wang Y, Feng Z, Shen Y. Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. IEEE Trans Med Imaging. 2022;41(1):145–57. doi:10.1109/TMI.2021.3107013.
11. Liu C, Kong D, Wang S, Li J, Yin B. DLGAN: depth-preserving latent generative adversarial network for 3D reconstruction. IEEE Trans Multimed. 2021;23:2843–56. doi:10.1109/TMM.2020.3017924.
12. Zhan Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018; Salt Lake City, UT, USA. p. 9242–51.
13. Zhang Z, Yang L, Zheng Y. Multimodal medical volumes translation and segmentation with generative adversarial network. In: Handbook of medical image computing and computer assisted intervention. Academic Press; 2019. p. 183–204.
14. Pan Y, Liu M, Lian C, Xia Y, Shen D. Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages. In: Medical image computing and computer assisted intervention. vol. 11766. Cham: Springer; 2019. p. 137–45.

15. Xue Y, Han X, Neri F, Qin J, Pelusi D. A gradient-guided evolutionary neural architecture search. IEEE Trans Neural Networks Learn Syst. 2024;36(3):4345–57. doi:10.1109/TNNLS.2024.3371432.
16. Xue Y, Han X, Wang Z. Self-adaptive weight self-adaptive weight based on dual-attention for differentiable neural architecture search. IEEE Trans Ind Inf. 2024;20(4):6394–403. doi:10.1109/TII.2023.3348843.
17. Patro BN, Namboodiri VP, Agneeswaran VS. SpectFormer: frequency and attention is what you need in a vision transformer. arXiv:2304.06446. 2023.
18. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018; Salt Lake City, UT, USA. p. 586–95.
19. Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. arXiv:1710.10196. 2017.
20. Choi Y, Uh Y, Yoo J, Ha J. StarGAN v2: diverse image synthesis for multiple domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020; Seattle, WA, USA. p. 8188–97.
21. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12. doi:10.1109/TIP.2003.819861.
22. Park T, Efros AA, Zhang R, Zhu JY. Contrastive learning for unpaired image-to-image translation. In: Computer Vision-ECCV 2020: 16th European Conference; 2020 Aug 23–28; Glasgow, UK. p. 319–45.
23. Choi J, Kim S, Jeong Y, Gwon Y, Yoon S. ILVR: conditioning method for denoising diffusion probabilistic models. arXiv:2108.02938. 2021.
24. Meng C, He Y, Song Y, Song J, Wu J, Zhu JY, et al. SDEdit: guided image synthesis and editing with stochastic differential equations. In: International Conference on Learning; 2022; Virtual Event.
25. Zhao M, Bao F, Li C, Zhu J. EGSDE: unpaired image-to-image translation via energyguided stochastic differential equations. arXiv:2207.06635. 2022.
26. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycleconsistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017; Venice, Italy. p. 2223–32.
27. Torbunov D, Huang Y, Yu H, Huang J, Yoo S, Lin M, et al. UVCGAN: UNet vision transformer cycle-consistent GAN for unpaired image-to-image translation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2023; Waikoloa, HI, USA. p. 702–12.
28. Torbunov D, Huang Y, Tseng HH, Yu H, Huang J, Yoo S, et al. UVCGAN v2: an improved cycle-consistent GAN for unpaired image-to-image translation. arXiv:2303.16280. 2023.
29. Interior Design Dataset (IDD). [cited 2025 Feb 25]. Available from: https://idd.insaan.iiit.ac.in.
30. Liu Z, Luo P, Qiu S, Wang X, Tang X. DeepFashion: powering robust clothes recognition and retrieval with rich annotations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, NV, USA. p. 1096–104.