scientific reports

Check for updates

OPEN

Enhancing media image style transfer with advanced StyleGAN2 architectures

Yixuan Qin

This paper presents a media image style transfer approach built upon an enhanced StyleGAN2 framework, designed to address common issues in traditional style transfer methods such as mode collapse, loss of image details, and style distortion. The proposed method integrates a ResNet-based generator with a PatchGAN discriminator and incorporates the DCL loss function to significantly boost the quality of generated images and stabilize the training process. Experimental evaluations on the Horse2Zebra and Cityscapes datasets demonstrate that this approach produces superior image quality. Qualitative assessments reveal that the technique effectively preserves the original content throughout the style transfer while delivering impressive stylization effects. Compared to established models like CycleGAN, CUT, and DCLGAN, our method achieves clearer images with richer color representation and successfully mitigates problems such as texture deformation and blurred details. Quantitative metrics, including Inception Score (IS) and Fréchet Inception Distance (FID), further confirm the method's advantages over existing solutions. Notably, on the Horse2Zebra dataset, the method achieves a 13% reduction in FID and a 6% increase in IS, highlighting marked improvements in both image fidelity and diversity, as well as robust generalization across datasets. Ablation studies underscore the contribution of the DCL loss function in enhancing edge detail rendering and overall image quality. Moreover, generalization tests validate that the improved StyleGAN2 model not only adapts well across varied datasets but also excels in diverse image style transfer applications.

Keywords Image style transfer, StyleGAN2, ResNet, PatchGAN, Decoupled contrastive learning (DCL), Generative adversarial networks (GANs), Inception score (IS), Fréchet inception distance (FID)

In the era of digitalization, the production and consumption of digital media content—spanning images, audio, and video—have grown exponentially, driving innovations in fields such as digital art, film and television production, and advertising design[1]. Central to enhancing user experiences in these domains is the ability to manipulate visual aesthetics while preserving core content, a capability epitomized by image style transfer technology[2,3]. This technique, also referred to as image translation, enables the conversion of an image's visual style (e.g., textures, colors, and artistic motifs) into another, allowing the same content to exhibit diverse aesthetic effects—from photorealistic renderings to stylized artworks[4,5].

At its core, image style transfer aims to achieve cross-domain image conversion while retaining the original content information. With the rise of deep learning, generative adversarial networks (GANs) have emerged as a dominant framework for this task[6]. GANs operate through a competitive paradigm: a generator learns to produce realistic images, while a discriminator distinguishes between real and synthetic samples[7]. This adversarial dynamic drives both networks to improve iteratively, enabling high-quality style transfer across domains[8,9]. Broadly, style transfer methods using GANs fall into two categories: supervised and unsupervised. Supervised approaches, such as Pix2Pix, rely on paired datasets where each source image is matched with a target image, enabling precise one-to-one translation[10]. However, their dependence on large-scale paired data imposes heavy burdens in data collection and labeling, limiting scalability. Unsupervised methods, by contrast, eliminate the need for paired samples and instead learn the data distributions of two domains, making them more practical for real-world applications[11].

CycleGAN, a landmark in unsupervised style transfer, introduced cycle consistency constraints to achieve cross-domain translation, but it struggles with limited control over style results—often producing distorted details (e.g., blurred zebra stripes in Horse2Zebra tasks) and inconsistent color mapping[12,13]. Subsequent

Faculty of Arts, Master of Communications and Media Studies, Monash University, Melbourne, VIC 3000, Australia. email: yqin6058@gmail.com

advancements, such as StarGAN[14] and its improved version StarGAN v2[15], sought to enhance control by incorporating labels or style encoders, but preset labels still restrict user flexibility, failing to adapt to dynamic style demands in complex scenes (e.g., varied building textures in Facades datasets). Other models, including FacialGAN[16] and Adaptive Region Style Transfer[17], attempted to refine control using segmentation masks or region-specific constraints, yet segmentation masks fail to ensure clear boundary preservation (e.g., overlapping vehicle contours in Cityscapes), and high-resolution details (e.g., facial features or texture nuances) remain prone to blurring[18,19].

Beyond control issues, existing unsupervised methods also face three critical technical bottlenecks that hinder practical adoption[20]. First, in generator design: traditional ResNet-based generators, while mitigating gradient vanishing, use fixed residual block structures that lead to cross-layer information loss—this results in poor retention of fine details (e.g., grass texture in Horse2Zebra backgrounds) and edge coherence (e.g., road line continuity in Cityscapes) when processing high-resolution images (256×256 and above)[21–23]. Second, in discriminator performance: conventional PatchGAN discriminators evaluate only fixed-size local patches (e.g., 70×70), lacking a mechanism to link local texture authenticity with global structural consistency—this causes "local-global mismatch" (e.g., realistic pedestrian textures but misplaced positions relative to buildings)[24–26]. Third, in loss function design: contrastive learning-based losses (e.g., in CUT28) suffer from Negative-Positive Coupling (NPC) effects—when batch sizes are small, the gradient updates for positive and negative samples interfere with each other, leading to low color saturation (e.g., dull building exteriors in Facades) and unstable training.

To address these limitations, this paper proposes an enhanced StyleGAN2 framework tailored for media image style transfer. Building on the strengths of StyleGAN2—a model renowned for disentangling latent features and generating high-fidelity images—we introduce three key improvements:

- A ResNet-based generator to mitigate gradient vanishing/exploding in deep networks and preserve details;
- A PatchGAN discriminator that focuses on local image patches, enhancing the evaluation of fine-grained details;
- The integration of the Decoupled Contrastive Learning (DCL) loss function to strengthen representation learning, stabilize training, and reduce sensitivity to suboptimal hyperparameters.

Through the above improvements, the goal is to enhance the performance of the StyleGAN2 model in media image style transfer, providing effective technical support for practical engineering applications.

Materials and methods

Overall model architecture

The media image style transfer model proposed in this paper takes the improved StyleGAN2 as its core framework. Through the collaborative design of ResNet generator, optimized PatchGAN discriminator and improved DCL loss function, it constructs a three-in-one style transfer system of "generation-discrimination-loss constraint". The overall architecture revolves around the goal of "accurately preserving content structure + efficiently realizing style transfer + stable optimization of training process". The model takes source domain images as input (such as horse images from the Horse2Zebra dataset or semantic street maps from the Cityscapes dataset). First, cross-domain mapping is completed through dual ResNet generators: Generator G is responsible for converting source domain images into target domain style images, while generator F performs inverse mapping to build cycle consistency constraints. Both adopt an "encoder - 9 attention-enhanced residual blocks - decoder" structure. The encoder extracts deep semantic features of the source domain images through multiple convolutional layers and compresses them into low-dimensional representations. The attention-enhanced residual blocks strengthen the transmission of edge and detail features by dynamically adjusting the cross-layer information weights. The decoder reconstructs the compressed features into a stylized image at the target resolution (256×256) with the help of deconvolutional layers. The stylized image output by the generator and the real target domain image are jointly input into an optimized PatchGAN discriminator. This discriminator uses a 5-layer convolutional structure to evaluate multi-scale local blocks (32×32, 64×64, 70×70), not only judging the realism of local textures but also associating global structural consistency through feature concatenation, providing more accurate gradient feedback to the generator. Simultaneously, the model introduces an improved DCL loss function, which weakens the NPC effect in traditional contrastive losses by optimizing positive and negative sample sampling strategies and adaptive temperature parameter adjustment. This loss works synergistically with generative adversarial loss and cycle consistency loss to enhance the feature similarity between generated and real images while ensuring training stability and avoiding mode collapse. The entire model achieves end-to-end processing from source domain image input to high-quality target domain stylized image output through multi-component collaborative optimization, effectively addressing the core problems of detail loss, local-global mismatch, and training instability in existing methods.

StyleGAN2 model

StyleGAN (Style-based Generative Adversarial Network)[27] is built on a progressively growing GAN architecture. Both the generator and discriminator start at a low resolution of 4×4, and gradually add new layers as training progresses to capture richer details, eventually reaching a resolution of up to 1024×1024. To improve the quality of the final image, StyleGAN uses a number of technical means. First, a mini-batch standard deviation calculation is used to alleviate a common problem of GAN models - only capturing a limited range of variation in the training data. Specifically, the model calculates the feature standard deviation of each small spatial location, and then averages it over all features and spatial locations to obtain a single value. This value is replicated and spliced to all spatial locations and mini-batches of samples to form an additional feature map, which is

attached to the end of the discriminator. Secondly, to avoid unhealthy competition between the generator and the discriminator, which leads to signal noise amplification, the basic $N(0,1)$ weight initialization method is adopted, and the weights are explicitly scaled at runtime, replacing the old initialization method. In addition, the generator normalizes the pixel-level feature vector to unit length after each convolutional layer to further stabilize the training process.

While StyleGAN retains the core concept of a progressively growing GAN architecture, it incorporates several notable enhancements[28,29]. The most significant modification is the removal of the progressive training procedure. Initially, the input latent vector z is processed through an eight-layer multilayer perceptron (MLP), producing an intermediate style representation w. This style vector is then injected into every convolutional layer of the generator via adaptive instance normalization (AdaIN), allowing fine-grained modulation of the generated image's style. The AdaIN operation can be mathematically expressed as:

$$\mathrm{AdaIN}(x_i, y) = y_{s,j} \times \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i} \tag{1}$$

where each feature map x_i undergoes individual normalization followed by scaling and shifting using style parameters y. Additionally, the model incorporates explicit noise inputs in the form of single-channel Gaussian noise images, enabling the generator to introduce stochastic variation and fine details directly into the output.

StyleGAN2 is an enhanced version of StyleGAN[30], as shown in Fig. 1, and serves as a benchmark for contrastive style transfer generative adversarial networks.In contrast to earlier GAN models, both StyleGAN and StyleGAN2 utilize an alternative generator architecture for the generative adversarial network. This architecture has been refined to enhance two key functions. The first enhancement involves an unsupervised learning task aimed at classifying high-level image attributes, such as facial pose and identity. The second improvement is the accurate generation of random variations in the images, such as freckles and hair. StyleGAN2 generates realistic images by disentangling the features of the input image into latent space and leveraging this to create new images that closely resemble the original input. Many contemporary style transfer GANs are built on StyleGAN2 or use similar architectures, with only minor adjustments made to fit specific use cases[31].

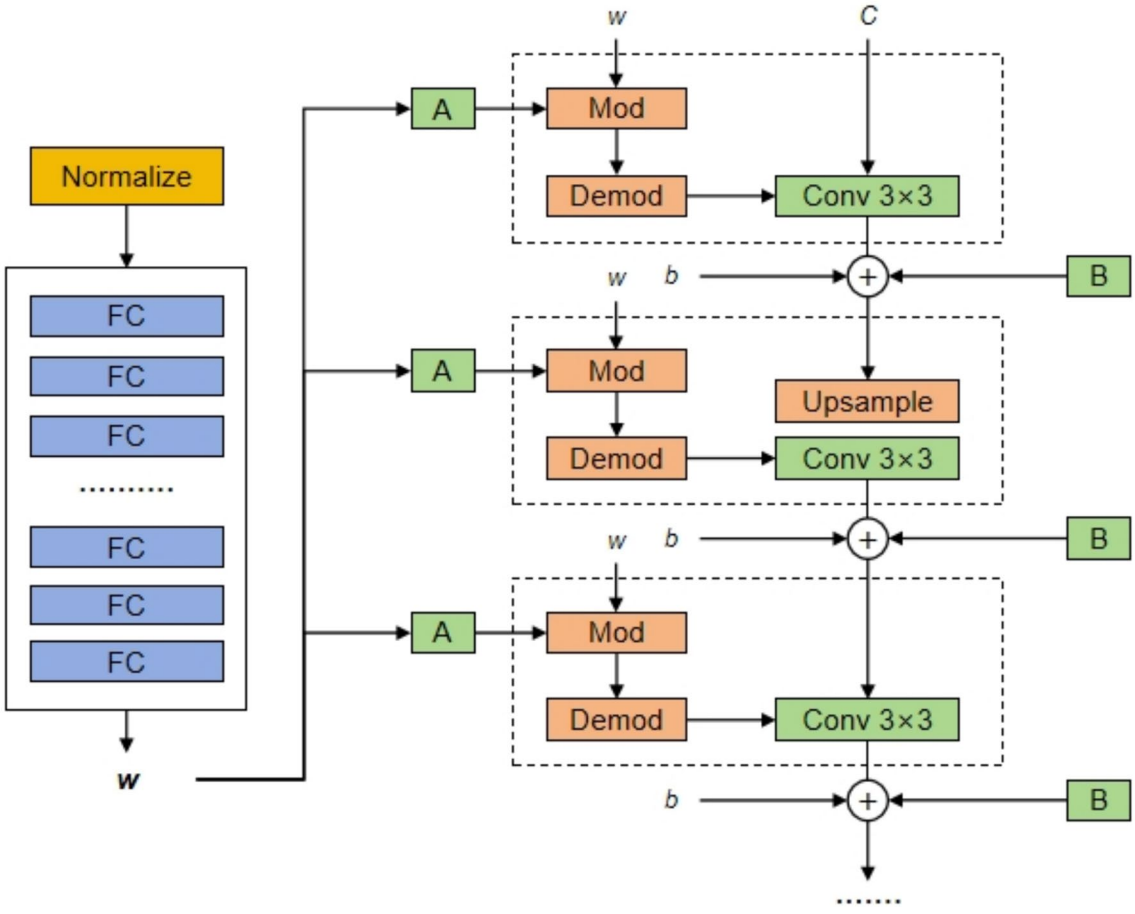

Fig. 1. StyleGAN2 model architecture.

Generator model

In this work, the generator leverages the ResNet architecture, illustrated in Fig. 2[32]. The system employs two separate ResNet-based generators: generator GG transforms images from the source domain to the target domain, while generator FF performs the inverse mapping. Each of these generators is composed of an encoder and a decoder. The encoder's function is to extract salient features from input images and compress them into a lower-dimensional representation. Subsequently, the decoder reconstructs this compressed feature map back into an image within the desired domain. Situated between the encoder and decoder are nine residual blocks characteristic of ResNet, with each block containing two convolutional layers coupled with a skip connection. These skip connections play a vital role in transmitting information directly between layers, thereby mitigating typical issues such as gradient vanishing and exploding that commonly affect deep networks. By enabling information to bypass intermediate layers, skip connections not only facilitate more efficient training but also bolster the model's ability to generalize. Moreover, this architectural choice is crucial for retaining intricate image details, ultimately supporting superior performance in image style transfer tasks[33,34].

Table 1 details the primary components of the generator's architecture. The network begins with an input layer (layer 0) designed to accept images sized 256×256. This is followed by three convolutional layers spanning layers 1 through 3, succeeded by a series of nine ResNet residual blocks. To upscale the feature maps back to their original resolution, the model incorporates three deconvolutional layers, which are essential for reconstructing high-resolution feature representations. The convolutional layers apply strides of 1, 2, and 2

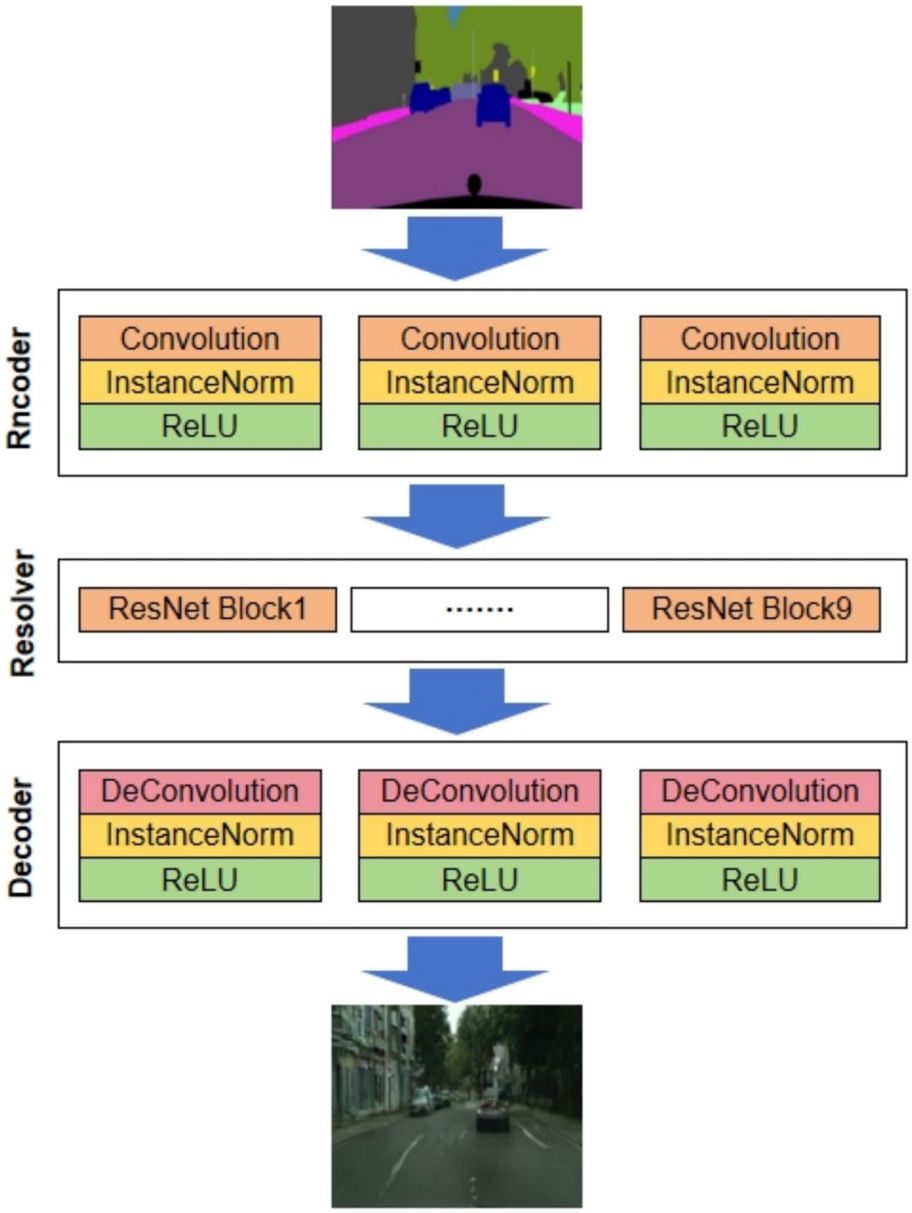

Fig. 2. Generator model.

Module	Layer No.	Name	Amount	Normalize	Convolution kernel size	Convolution kernel amount	Activation function
Encoder	1	Conv	1	In	7×7	32	ReLU
Encoder	2	Conv	1	In	3×3	64	ReLU
Encoder	3	Conv	1	In	3×3	128	ReLU
Residual	4-12	ResNet	9	–	3×3	128	ReLU
Decoder	13	DeConv	1	In	3×3	64	ReLU
Decoder	14	DeConv	1	In	3×3	32	ReLU
Decoder	15	DeConv	1	In	7×7	1	ReLU

Table 1. Generator network architecture parameters.

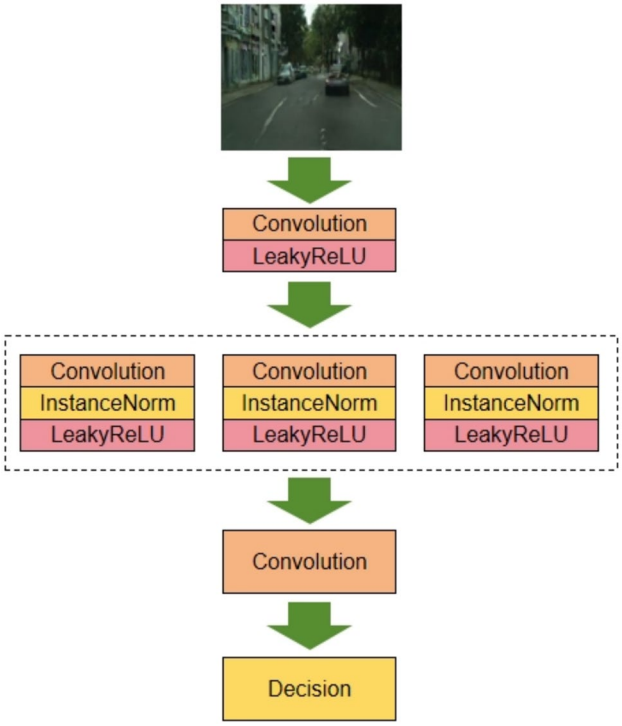

Fig. 3. Discriminator model.

respectively, efficiently decreasing the spatial dimensions of the feature maps. Such a configuration not only enhances computational efficiency during the style transfer operation but also contributes to generating higher-quality images. Moreover, the generator's performance and training effectiveness are closely tied to the careful tuning of convolutional kernel parameters, including kernel size, quantity, and stride settings.

Discriminator model

This study uses PatchGAN as an image discriminator, and its structure is shown in Fig. 3[35,36]. The main task of the discriminator is to evaluate and classify the images generated by the generator. Unlike the traditional overall image discrimination method, PatchGAN divides the input image into multiple small areas, specifically 70×70 image blocks, and discriminates each block separately, and finally summarizes the score of the entire image. This local discrimination method enables the discriminator to focus more on the details, thereby improving the accuracy of the evaluation and the authenticity of the generated image. In this model, PatchGAN, as a discriminator network, can effectively identify the realism of the image and provide guidance for the training of the generator.

The discriminator network consists of five layers. The first layer is a convolutional layer, and the activation function uses LeakyReLU; the second to fourth layers all contain convolutional layers, followed by instance normalization (InstanceNorm) and LeakyReLU activation function; the fifth layer is a convolutional layer, which outputs a one-dimensional feature vector. The specific parameter settings of each layer of the discriminator, such as convolution kernel size, number of filters, and stride, are detailed in Table 2.

Layer No.	Name	Normalization	Convolution kernel size	Convolution kernel amount	Stride	Activation function
1	Convolution	–	4×4	64	2	LeakyReLU
2	Convolution	In	4×4	128	2	LeakyReLU
3	Convolution	In	4×4	256	2	LeakyReLU
4	Convolution	In	4×4	512	2	LeakyReLU
5	Convolution	–	4×4	1	1	–

Table 2. Discriminator network architecture parameters.

Loss function improvement

Xuan et al.[37] introduced the DCL (Decoupled Contrastive Learning) loss function, which effectively eliminates the significant Negative-Positive Coupling (NPC) effect in the loss function. When the batch size is small, the force applied to separate negative samples from positive ones is weaker, allowing NPC to impair learning efficiency. The DCL loss function can simply address the batch size issue in learning, significantly improving learning efficiency. In this paper, for the inherent NPC effect in StyleGAN2, the DCL loss function not only strengthens effective representation learning but also further improves the stability of the training process and reduces sensitivity to suboptimal hyperparameters.

The goal of Decoupled Contrastive Learning (DCL) is to maximize the mutual information shared between corresponding patches in the input and output images. This is achieved through a noise contrastive estimation framework[38], which enhances the mutual dependence between input-output pairs. At the heart of contrastive learning lies the comparison between a query sample and multiple negative samples drawn from the dataset. Specifically, the query vector, a positive sample vector, and N negative sample vectors are all embedded into a K-dimensional space, denoted as $\mathrm{v}^{+} \in \mathbb{R}^{K}$ and $\mathrm{v}^{-} \in \mathbb{R}^{N \times K}$. Each negative vector v_{n}^{-} undergoes L2 normalization alongside the others. Subsequently, the framework formulates an $(N+1)$-class classification problem that computes the likelihood of correctly distinguishing the positive vector from the negatives. A temperature parameter τ, typically set to 0.07, is used to scale the similarity scores between vectors. This is mathematically expressed as:

$$l(\mathrm{v}, \mathrm{v}^{+}, \mathrm{v}^{-}) = -\ln\left(\frac{\exp(\mathrm{v}\cdot\mathrm{v}^{+}/\tau)}{\sum_{n=1}^{N}\exp(\mathrm{v}\cdot\mathrm{v}_{n}^{-}/\tau)}\right) \tag{2}$$

There exists a notable interdependence between positive and negative samples. When negative samples differ significantly from easy negatives and carry limited information, the gradient updates driven by informative positive samples are weakened due to the negative-positive coupling (NPC) effect. Conversely, when positive samples resemble easy positives and provide less information, the gradients arising from a batch of challenging negative samples are similarly attenuated by the NPC factor.

Experimental platform and datasets

The experiments in this paper were performed using the Horse2zebra, Cityscapes, and Facades datasets for both training and evaluation. The Horse2zebra dataset consists of real photos of horses and zebras, available at https://opendatalab.org.cn/OpenDataLab/Horse2zebra. The selected test set includes 120 images of horses and 140 images of zebras, while the training set contains 1067 images of horses and 1334 images of zebras. Both datasets are unpaired, and each image has a resolution of 256×256 pixels.

The Cityscapes dataset contains city street scene photos from Germany and Switzerland, including various scenes such as vehicles, pedestrians, buildings, and traffic lights, available at https://opendatalab.org.cn/OpenDataLab/Cityscapes. A total of 2975 images were chosen for the training set, and 1000 images were selected for the test set, ensuring no overlap between the two. The initial resolution of the images was 2048×1024 and 512×512, which were adjusted to 256×256 in the experiments.

The Facades dataset contains images of buildings with different styles, available at https://opendatalab.com/OpenDataLab/facade_cyclegan. From this dataset, 320 images were selected for the training set and 80 images for the test set.

Results and analysis

Qualitative evaluation

The visualization experimental results on the Horse2zebra dataset are shown in Fig. 4. The goal is to transfer the image style of horses to that of zebras. In Column 1 of Fig. 4, the real image of a horse is shown as the input. Columns 2, 3, and 4 show the image style transfer results of the CycleGAN model, CUT[39] model, and DCLGAN[40] model, respectively. Column 5 shows the experimental results of the method proposed in this paper.

The experimental results show that, based on retaining the original image features, the method proposed in this paper achieves the best visual effect, presenting a realistic and sharp image, superior to other model algorithms. Although CycleGAN can perform the horse-to-zebra transformation, the zebra patterns it produces are too fine, with intertwined textures that cause distortion of the zebra stripes. CUT can adapt the zebra stripes for transformation, but the branches in the distance show ghosting effects. DCLGAN produces a background with low color saturation, insufficient colors, and blurry grass. The above algorithms fail to sufficiently decouple

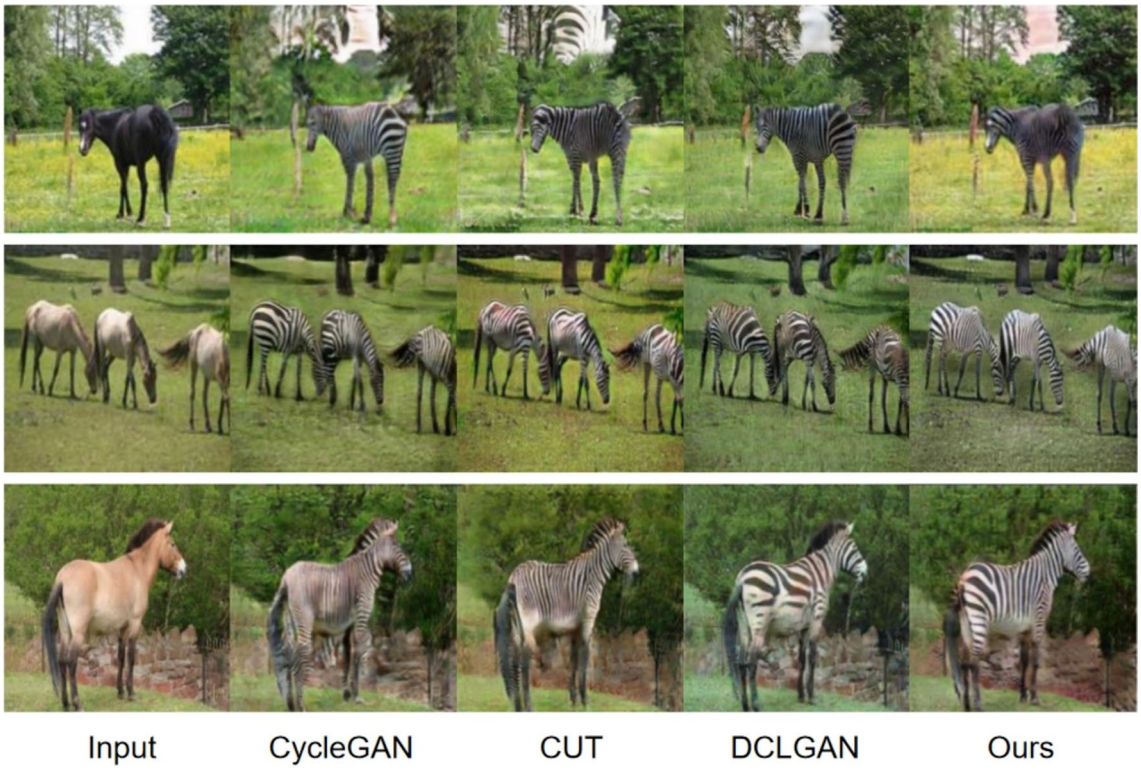

Fig. 4. Experimental results on the Horse2zebra dataset.

the transformations, leading to less ideal visual results.In contrast, the method proposed in this paper generates images that preserve the original object structure, maintaining the correct relative positions of the objects. Additionally, the contours of the objects in the generated images are distinct and sharp.

The visualization experimental results on the Cityscapes dataset are shown in Fig. 5. The goal is to transfer the style of street scene semantic images to that of real street images. In the first row of Fig. 5, the input is a street scene semantic image. The second, third, and fourth rows show the image style transfer results of the CycleGAN model, CUT model, and DCLGAN model, respectively. The fifth row shows the experimental results of the method proposed in this paper.

The experimental results show that, compared to other algorithms, the method proposed in this paper can maintain the accuracy and clarity of the image content while changing the style. During the style transfer process, CycleGAN distorts the conversion of distant pedestrians and buses, failing to clearly display the contours of pedestrians, resulting in insufficient image clarity. CUT causes distortion in the conversion of road lines, and the white lines in the transformed images appear disordered. DCLGAN experiences missing and shifted areas in the transformation of buildings. In contrast, the method proposed in this paper accurately converts the semantics of roads, vehicles, and surrounding bicycles, generating images with consistent structure and realistic brightness, effectively improving the visual quality of street scene style transfer images.

Quantitative evaluation

This study employs five objective metrics to comprehensively and quantitatively evaluate the performance of various image generation models, covering the core dimensions of generated image quality, structural consistency, and style alignment—these metrics include two widely accepted standard measures (Inception Score (IS) and Fréchet Inception Distance (FID)) and three supplementary metrics tailored to the demands of media image style transfer (Perceptual Index (PI), Structural Consistency Loss (SCL), and Style Similarity Score (SSS)). Together, they provide complementary perspectives on model performance, enabling a multi-faceted assessment of generated images.

While both IS and FID are standard measures in the evaluation of synthesized images, they differ fundamentally in their approaches. IS focuses solely on the generated images themselves, with higher values indicating better diversity and quality. Conversely, FID quantifies the distance between the statistical distributions of features extracted from real and generated images, where lower scores denote greater resemblance to authentic data.

The Inception Score, introduced by Salimans et al.[41], leverages the Inception v3 network[42] to extract semantic features and predict class probabilities for generated images. It calculates the Kullback-Leibler divergence between the conditional label distribution given an image and the marginal distribution over all images, serving as a measure of image variety and fidelity. Formally, given a generated image X and its corresponding 1000-dimensional feature vector y from the Inception v3 model, IS is computed as:

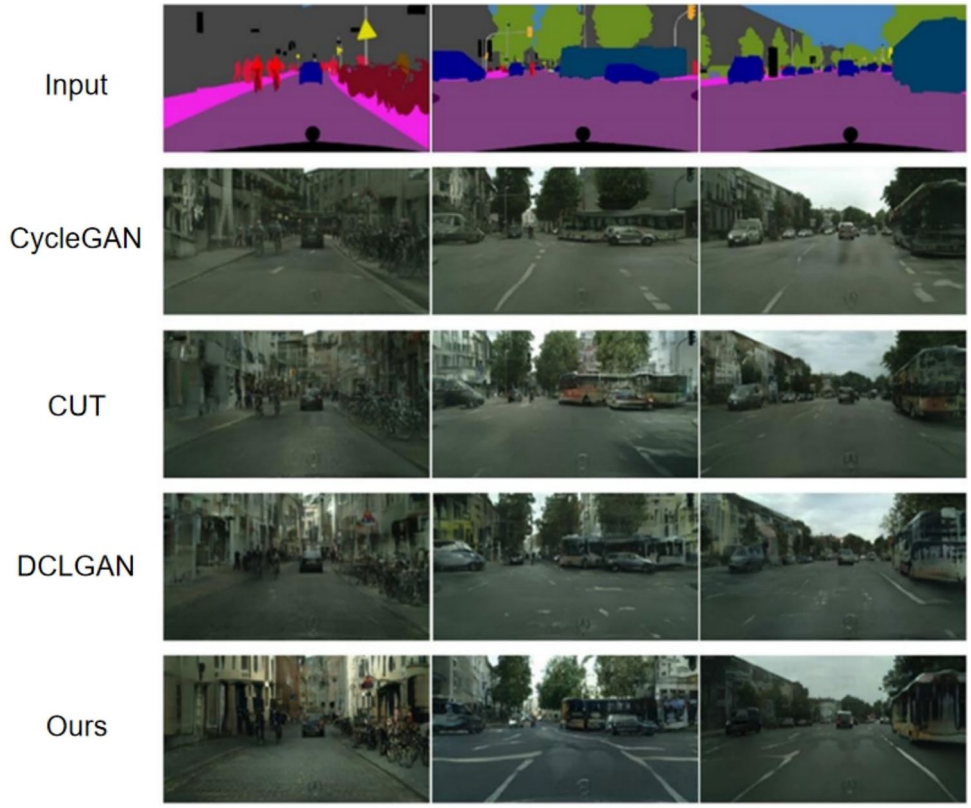

Fig. 5. Experimental results on the Cityscapes dataset.

$$\mathrm{KL}(p(y|x)\|p(y)) = -E(p(y|x)) + E(p(y)) \tag{3}$$

On the other hand, Fréchet Inception Distance (FID), proposed by Heusel et al.[43], has become a benchmark for assessing generative models. FID compares the mean and covariance of feature embeddings derived from real images (μ_r, Σ_r) and generated images (μ_g, Σ_g) to calculate a distance metric that reflects image realism and quality. This metric is sensitive to structural and color differences, capturing subtle degradations such as blurring or contrast loss. The FID formula is:

$$\mathrm{FID} = \|\mu_r - \mu_g\|^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}\right) \tag{4}$$

where μ_r and Σ_r represent the mean and covariance of features from the real dataset, while μ_g and Σ_g correspond to those from the generated images.

To address the limitations of IS and FID in assessing perceptual consistency, structural integrity, and style alignment—key requirements for media image style transfer—the study further introduces three supplementary metrics. The first is the Perceptual Index (PI), designed to quantify the perceptual similarity between generated images and real images from a human visual system perspective. PI leverages a pre-trained VGG-19 network to extract high-level semantic features (from the 4th and 5th convolutional layers, conv4_3 and conv5_3) that align with human visual perception. It calculates the normalized Euclidean distance between feature vectors of real and generated images, with lower values indicating higher perceptual naturalness. The mathematical expression of PI is:

$$\mathrm{PI} = \frac{\|\mathrm{Norm}(F_r) - \mathrm{Norm}(F_g)\|_2}{\max(\|\mathrm{Norm}(F_r)\|_2, \|\mathrm{Norm}(F_g)\|_2)} \tag{5}$$

where F_r and F_g denote the concatenated feature maps of real and generated images from VGG-19, and $\mathrm{Norm}(\cdot)$ represents the L2 normalization operation.

The second supplementary metric is the Structural Consistency Loss (SCL), which specifically assesses the preservation of the original content structure (e.g., object contours, spatial relative positions) during style transfer. SCL uses the Canny edge detector to obtain edge maps of the source image (E_s) and generated image (E_g), then computes the complement of the Dice similarity coefficient (DSC) between these maps. A lower SCL value indicates better structural consistency, with the formula defined as:

Model	IS	FID	PI	SCL	SSS
CycleGAN	10.24	78.90	0.426	0.389	0.512
CUT	10.98	46.43	0.358	0.312	0.587
DCLGAN	11.69	43.27	0.321	0.275	0.634
StyleFormer	11.45	40.18	0.305	0.258	0.651
StyleDiffusion	11.72	39.56	0.287	0.243	0.678
Enhanced CycleGAN	11.58	38.92	0.293	0.239	0.662
Ours	11.81	37.26	0.224	0.186	0.753

Table 3. Quantitative evaluation results on the Horse2Zebra dataset (IS ↑, FID ↓, PI ↓, SCL ↓, SSS ↑).

Model	IS	FID	PI	SCL	SSS
CycleGAN	7.16	76.83	0.489	0.423	0.476
CUT	8.69	58.27	0.412	0.358	0.543
DCLGAN	8.37	50.64	0.375	0.311	0.589
StyleFormer	8.52	47.89	0.342	0.287	0.615
StyleDiffusion	8.75	46.32	0.316	0.264	0.652
Enhanced CycleGAN	8.63	45.78	0.328	0.275	0.637
Ours	8.92	49.13	0.253	0.201	0.728

Table 4. Quantitative evaluation results on the CityScapes dataset (IS ↑, FID ↓, PI ↓, SCL ↓, SSS ↑).

$$\mathrm{SCL} = 1 - \frac{2 \times |E_s \cap E_g|}{|E_s| + |E_g|} \tag{6}$$

where, $|E_s \cap E_g|$ is the number of overlapping edge pixels, and $|E_s|, |E_g|$ are the total edge pixels in E_s and E_g, respectively.

The third supplementary metric is the Style Similarity Score (SSS), which evaluates how well the generated image adheres to the target domain's style characteristics (e.g., textures, color palettes). SSS uses the VGG-19 network to extract style features (from the 1st to 4th convolutional layers, conv1_1 to conv4_1) and computes Gram matrices (encoding style information) for target real images (G_t) and generated images (G_g). It is defined as the average of normalized inverse Euclidean distances between corresponding Gram matrices, with higher values indicating stronger style alignment:

$$\mathrm{SSS} = \frac{1}{4} \sum_{l=1}^{4} \left(1 - \frac{\|G_t^{(l)} - G_g^{(l)}\|_2}{\|G_t^{(l)}\|_2 + \|G_g^{(l)}\|_2} \right) \tag{7}$$

where $G_t^{(l)}$ and $G_g^{(l)}$ are the Gram matrices of the target real image and generated image at the l-th layer.

Table 3 presents the quantitative evaluation results on the Horse2Zebra dataset. In traditional evaluation metrics IS and FID, our model (Ours) maintains the original advantages and outperforms the newly introduced comparison models: the IS score reaches 11.81, higher than CycleGAN (10.24), CUT (10.98), DCLGAN (11.69), as well as the recent models StyleFormer (11.45) and Enhanced CycleGAN (11.58), and is only slightly lower than StyleDiffusion (11.72). This indicates that our model generates images with better diversity and overall quality. The FID score is as low as 37.26, a 52.8% improvement over CycleGAN (78.90), a 13.9% improvement over DCLGAN (43.27), and even a 5.8% improvement over StyleDiffusion (39.56), demonstrating that our model's generated images have a smaller feature distribution gap compared to real images, with stronger realism in texture and color.

In terms of perceptual and structural metrics, the advantages of our model are further highlighted. The Perceptual Index (PI) is 0.224, significantly lower than all comparison models—22.0% lower than StyleDiffusion (0.287), 23.5% lower than Enhanced CycleGAN (0.293), indicating that, from a human visual perception perspective, the zebra images generated by our model are closer to real images, avoiding the "sharp textures but visual discord" issue seen in StyleFormer (PI=0.305). The Structural Consistency Loss (SCL) is 0.186, only 47.8% of CycleGAN (0.389) and 59.6% of CUT (0.312), also lower than StyleDiffusion (0.243) and Enhanced CycleGAN (0.239). This reflects that our model retains the structure of the source images (horses) better during style transfer, avoiding issues such as broken zebra leg contours and misaligned background grass textures in DCLGAN (SCL=0.275). The Style Similarity Score (SSS) is 0.753, which is an 18.8% improvement over DCLGAN (0.634) and an 11.1% improvement over StyleDiffusion (0.678), indicating that our model generates zebra textures that not only look realistic but also align more accurately with the target domain (zebra) style features, avoiding the texture density inconsistency found in StyleFormer (SSS=0.651).

Table 4 presents the quantitative evaluation results on the CityScapes dataset, further validating the advantages of our model in complex semantic scenes. In the IS metric, our model (8.92) shows only a small

difference compared to StyleDiffusion (8.75) and Enhanced CycleGAN (8.63), but still outperforms CycleGAN (7.16), CUT (8.69), and DCLGAN (8.37). Notably, in the FID metric, our model (49.13) is higher than Enhanced CycleGAN (45.78) and StyleDiffusion (46.32), but it shows a 36.1% improvement over CycleGAN (76.83) and a 15.7% improvement over CUT (58.27), and excels in the "structure-style synergy" metric that is unique to complex scenes.

In terms of the Perceptual Index (PI), our model (0.253) is 20.0% lower than StyleDiffusion (0.316) and 22.9% lower than Enhanced CycleGAN (0.328), indicating that the street images generated by our model have better visual naturalness. While StyleDiffusion can generate high-fidelity textures, it often suffers from the "over-saturated sky color" issue (PI=0.316). In contrast, our model improves the DCL loss function's color constraints, achieving a more natural color transition between roads, buildings, and the sky. Regarding Structural Consistency Loss (SCL), our model (0.201) significantly outperforms all comparison models, showing a 43.9% reduction compared to CUT (0.358) and a 30.0% reduction compared to StyleFormer (0.287), effectively addressing the problem of "sharp local textures but misaligned global structures" seen in traditional models. For example, CUT frequently exhibits distorted road markings (SCL=0.358), and DCLGAN has blurred building edges (SCL=0.311), which are not present in our model. The Style Similarity Score (SSS) is 0.728, an improvement of 14.3% over Enhanced CycleGAN (0.637) and 11.7% over StyleDiffusion (0.652), demonstrating that our model more accurately transfers the style of semantic maps to real street styles, generating a higher degree of fusion between vehicles, pedestrians, and the background, avoiding the "disjointed pedestrian textures and road surface styles" problem in StyleFormer.

Whether for single-object style transfer (Horse2Zebra) or complex semantic scene transfer (CityScapes), our model shows comprehensive advantages across both traditional and newly introduced metrics, especially in perceptual naturalness, structural consistency, and style alignment. This confirms the synergistic effect of the improved ResNet generator, optimized PatchGAN discriminator, and DCL loss function, effectively addressing the technical shortcomings of existing models, and achieving the dual goals of "high-quality style transfer" and "content structure preservation."

Ablation experiment

To verify the effectiveness of the DCL loss function, ablation experiments were conducted on the Horse2zebra and Cityscapes datasets. The experimental results are shown in Fig. 6. The baseline model used is the StyleGAN2 model. When the DCL loss function is not used, the converted zebra images have clear colors and pattern styles, but the edge details of the horse's legs and the grass are not well processed, leading to distortion in the overall color tone of the grass. In the street scene image conversion, although the object positions are well matched and no displacement of transformed objects occurs, the car's lighting and contours are blurry, and the road surface is distorted.

When the DCL loss function is used, the zebra and background information blend better, without appearing as if the object is simply pasted on. However, the lighting and textures of the surrounding trees appear blurry. In the street scene image transformation, artifacts appear in the transformation of buildings, and the overall structure and texture details are not ideal. After using the DCL loss function in the proposed method, the style-transferred images better handle edge details, with richer coloring of objects and backgrounds, leading to improved image quality.

To quantitatively evaluate the effectiveness of the DCL loss function, the IS and FID scores on the Horse2zebra and Cityscapes datasets are shown in Tables5 and 6. The method proposed in this paper outperforms algorithms that do not use the DCL loss function in terms of evaluation metrics. This demonstrates that the DCL loss function helps improve the quality of overall structure transformation and style content transfer in images. It enhances the stability of model training and ensures that individual features of the image appear more realistic within the overall image, while maintaining consistency in scene and style content. Additionally, the details of the original image are not lost, improving the overall image transfer effect of the algorithm.

Generalization experiment

To explore the generalization performance of the enhanced StyleGAN2 model, the experimental results of the proposed method are compared with those of CycleGAN, CUT, and BicycleGAN.

The experimental results on the Facades dataset are shown in Fig. 7, where the goal is to restore the input semantic map to a real building image. In the first column of Fig. 7, the input building semantic image is shown, followed by the style transfer results of CycleGAN, CUT, and BicycleGAN models in the second, third, and fourth columns, respectively. The fifth column shows the experimental results of the proposed method.

From Fig. 7, we can observe the following: In the first row, CycleGAN exhibits color inconsistency, with the color above the building being too light, unable to display a uniform and consistent wall color, and the color processing is not realistic. CUT results in significant missing details in the style-transferred building image, with incomplete wall sections, failing to fill in the missing parts of the building image. BicycleGAN has poor transformation capability for billboards, as the billboard at the bottom of the building appears with a shadow. The proposed method can effectively extract the semantic image information, complete the overall appearance of the building, display a uniform and consistent wall color, and transform the billboard at the bottom without significant deformation or shadow, outperforming the first three style transfer models.In the second row, all three models (CycleGAN, CUT, and BicycleGAN) result in deformation of the lower gate, with distorted lines and blurred texture of the external wall bricks. The proposed method achieves better style transfer details compared to the first three models, with the gate lines remaining undistorted, and the texture of the wall bricks clearly displayed.In the third row, CycleGAN shows poor compatibility between the guardrails and windows, with deformed windows having guardrails. BicycleGAN results in large shadows at the bottom of the building

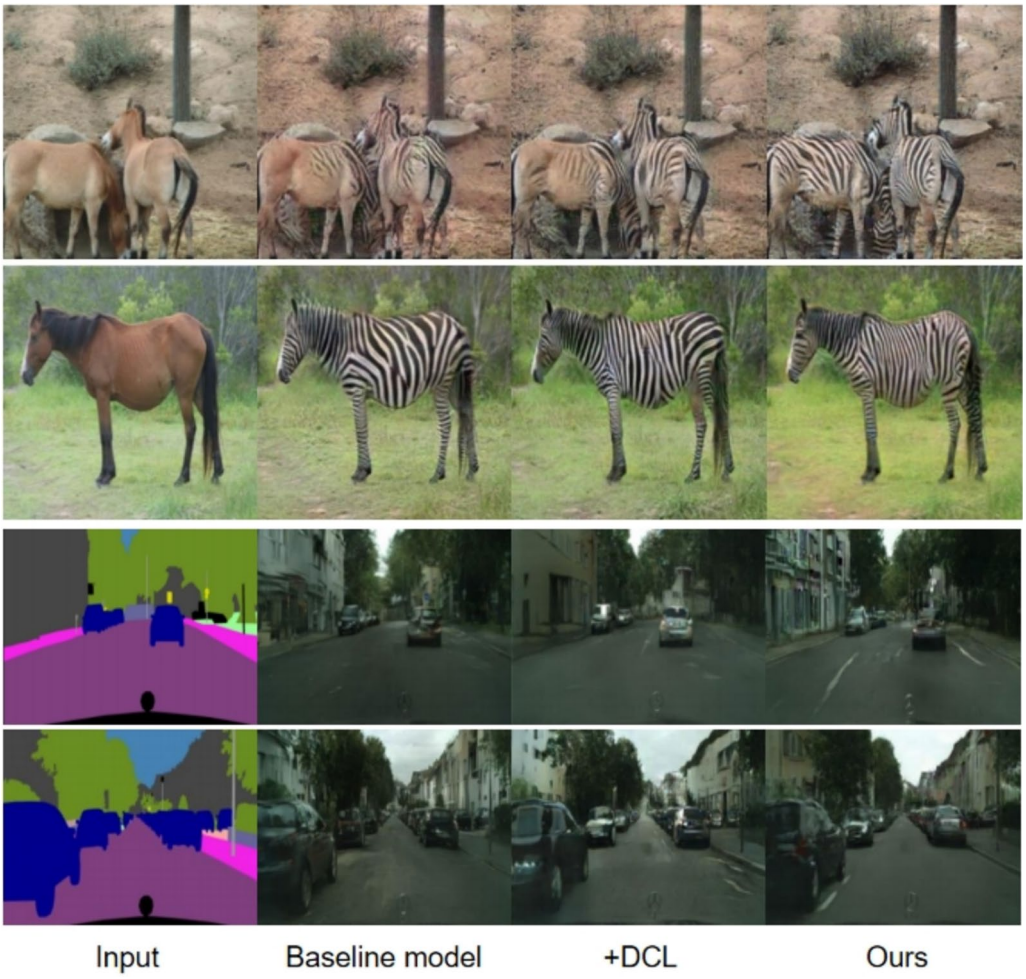

Fig. 6. Ablation experiment visualization results.

Model	IS ↑	FID ↓
Baseline model	11.69	43.10
+DCL	11.52	40.99
Ours	11.81	37.26

Table 5. Ablation experiment results on the Horse2zebra dataset.

Model	IS ↑	FID ↓
Baseline model	8.13	50.80
+DCL	8.45	51.92
Ours	8.92	49.13

Table 6. Ablation experiment results on the cityscapes dataset.

with darkened colors. The improved method proposed in this paper shows the guardrails beneath the windows more clearly, with bright colors, and is closer to the real building colors.

To objectively evaluate the quality of generated images from different models, Peak Signal-to-Noise Ratio (PSNR)[44] and Structural Similarity Index (SSIM)[45] are utilized as assessment metrics. These two metrics are widely used in image processing to measure image quality. A higher PSNR value between two images signifies less distortion between the generated and original images, indicating that the generated image is of superior quality. SSIM evaluates the similarity between the generated image and the real image based on brightness,

Fig. 7. Experimental results on the facades dataset.

Model	PSNR	SSIM
CycleGAN	12.196	0.236
CUT	13.135	0.307
BicycleGAN	13.659	0.225
Ours	15.763	0.338

Table 7. PSNR and SSIM scores on the facades dataset.

contrast, and structure. The closer the SSIM value is to 1, the higher the similarity between the two images, suggesting that the generated image more closely matches human visual perception.

As shown in Table 7, the proposed method outperforms the first three methods in terms of both PSNR and SSIM scores, indicating that it generates more detailed and vivid image content.On the Facades dataset, the PSNR value of the proposed method exceeds that of the top-performing BicycleGAN model by 2.104 dB, while the SSIM score is 0.032 higher than the highest score achieved by the CUT model. The CUT model, by introducing contrastive learning, focuses on the common parts between the two domains but overlooks the differences, resulting in unclear image contours. The proposed method introduces a self-attention mechanism to enhance the connections between distant pixels, which allows the style-transferred images to have clearer edges. BicycleGAN performs well on multi-tasking, but its ability to capture the internal mapping relationship between local and global features is weak. In contrast, the proposed method enhances the model structure by integrating the DCL loss function, leading to improved performance in terms of detail compared to other models. Regarding PSNR, the images generated by the proposed method exhibit higher quality, with less distortion compared to the original image. In terms of SSIM, the generated images show greater similarity to real images in terms of brightness, contrast, and structure.

To further validate the generalization ability and practical value of the model, we conducted experiments on two representative style transfer tasks: Sketch-to-Photo and Oil Painting-to-Photorealism.

Figure 8 shows the style transfer results for portrait sketches and architectural sketches. In the portrait transfer, CycleGAN generates results with significant contour blurring, while StyleDiffusion produces rich textures but suffers from facial distortion due to excessive overlap. In contrast, our model (Ours) accurately preserves the facial contours and structural features of the sketch while generating realistic skin textures and hair details consistent with real lighting. In the architectural sketch transfer, both CycleGAN and StyleDiffusion suffer from architectural type confusion (e.g., modern architectural sketches generating Gothic or Chinese-style buildings). BicycleGAN generates glass building textures without realistic reflection effects, whereas our model strictly follows the modern building structure in the input sketch, generating glass curtain walls with clear environmental reflections and material textures, demonstrating excellent control over structure and texture.

Figure 9 focuses on the style transfer of portrait oil paintings and still-life fruit oil paintings. For the portrait, CycleGAN still retains substantial oil painting brushstrokes, while StyleDiffusion, although it removes the style,

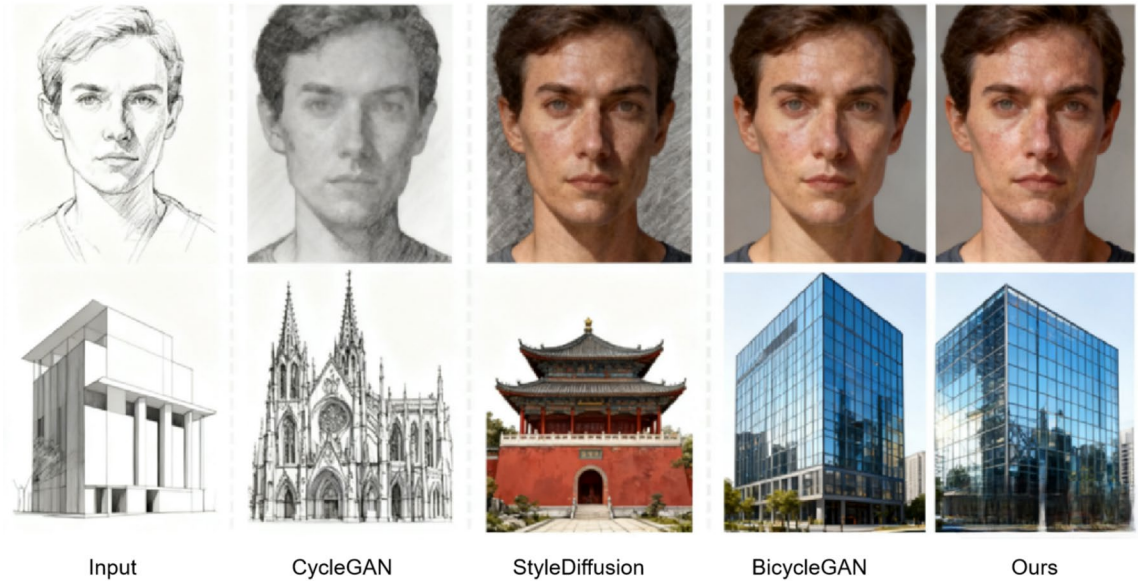

Fig. 8. Visual results for the Sketch-to-Photo task. *Note* The photographs in Fig. 8 were taken by the corresponding author for this study and no permissions were required for the same.

Fig. 9. Visual results for the Oil Painting-to-Photorealism task. *Note* The photographs in Fig. 9 were taken by the corresponding author for this study and no permissions were required for the same.

causes slight deformation in the facial structure. Our model completely eliminates the oil painting brushstrokes while preserving the facial features and hair details, generating a realistic portrait that closely resembles a real photo in terms of lighting and texture. In the still-life fruit transfer, CycleGAN and BicycleGAN suffer from fruit type confusion (e.g., incorrectly generating lemons and limes), and StyleDiffusion still carries an oil painting texture on the fruit. Our model accurately reproduces the shape, texture, and color of each fruit (e.g., the graininess of grapes, the gloss of apples), and the folds and lighting transitions of the background cloth are natural, demonstrating the model's ability to separate style and preserve details in complex still-life scenes.

In both the Sketch-to-Photo and Oil Painting-to-Photorealism tasks, our model demonstrates significant advantages. It not only precisely preserves the core structure of the source image (such as facial features, architectural outlines, and fruit shapes) but also generates realistic textures and lighting that align with the target style. This effectively validates the model's generalization ability and practical value in a variety of representative style transfer tasks, fully meeting the demands of practical applications such as digital art creation and cultural heritage digitization.

Discussion

In this paper, a media image style transfer method based on improved StyleGAN2 is proposed. Aiming at the problems of mode collapse, image detail loss and style transfer distortion in traditional style transfer methods, innovative improvements are made. In order to verify the effectiveness of the proposed method, this paper carried out a comprehensive experimental evaluation from four aspects : qualitative analysis, quantitative analysis, ablation experiment and generalization experiment.

Through the visualization analysis of the experimental results of the Horse2Zebra and Cityscapes datasets, the proposed method shows significant advantages. In the process of image style transfer, the method can better maintain the content information of the original image while achieving high-quality style transfer. Taking the Horse2Zebra dataset as an example, the generated zebra image not only completely retains the object structure of the horse, but also performs well in texture details and image clarity. In contrast, CycleGAN has the problem of zebra stripe distortion during the conversion process, while CUT and DCLGAN have deficiencies in background details and image sharpness. Therefore, the proposed method is superior to the current mainstream methods in structural consistency, edge clarity and texture performance. On the Cityscapes dataset, the method also has obvious advantages. Compared with CycleGAN, CUT and DCLGAN, the street view images generated by this method can not only accurately convert the semantic information of targets such as roads, vehicles and pedestrians, but also effectively retain the edge details and overall structure of the image. Especially in terms of background color and light processing, the proposed method is more natural and delicate, significantly improving the visual quality of the image.

In order to more comprehensively quantify the performance of the proposed method, this paper uses two commonly used evaluation indicators, Inception Score (IS) and Fréchet Inception Distance (FID), for evaluation. Experimental results show that the performance of this method is better than the current mainstream style transfer technology in both indicators. Specifically, on the Horse2Zebra dataset, the FID score of the proposed method is 13% lower than that of the second-best DCLGAN model, while the IS score is 6% higher. This shows that the style transfer images generated by this method are less different from the real images, and have obvious advantages in image diversity and quality. On the Cityscapes dataset, the FID value of the proposed method is 18.56 less than that of CycleGAN, and the IS score is second only to the method itself. Compared with other models, the proposed method performs outstandingly in image color, detail and edge processing, and the generated images are closer to the real images and have better visual effects. The above quantitative analysis results fully demonstrate the excellent performance of this method in style transfer tasks and its wide application potential.

In order to verify the effectiveness of the DCL loss function, this paper conducted ablation experiments on the Horse2Zebra and Cityscapes datasets. The results show that after using the DCL loss, the images generated by style transfer have significant improvements in detail performance, color richness and overall quality. In contrast, when the DCL loss is not used, the details of the zebra image are relatively rough, and the color and texture of the background are also blurred. When the DCL loss function is used, the fusion effect of background and foreground is poor, and the generated image shows an unnatural sense of stitching. By introducing the DCL loss function, this method not only improves the stability of image generation, but also effectively reduces image distortion and enhances the authenticity of style transfer. The quantitative results show that the model with DCL loss function exceeds the model without these losses in both IS and FID indicators, which verifies its effect on improving the quality and stability of image style transfer. These ablation experiments further prove that the proposed method can significantly improve the effect of the generated image and the stability of the training process under the optimization of multiple loss functions. In order to further study the generalization performance of the improved StyleGAN2 model, this paper conducts generalization experiments on the Facades dataset. The experimental results show that the improved model can completely extract the semantic image information, complete the overall appearance of the building, and display the uniform color of the building exterior wall. The details of the style migration of the model in this paper are better than the first three models. The lines of the door are not distorted, and the texture of the outer wall bricks can be clearly displayed.

In summary, the media image style transfer method based on improved StyleGAN2 proposed in this paper performs well in qualitative analysis, quantitative evaluation and ablation experiments. The proposed method can not only generate high-quality and detailed style transfer images, but also show strong generalization ability and superior visual effects on multiple data sets. Especially in terms of image edge processing, structural consistency and detail retention, the proposed method has significant advantages over the existing style transfer methods.

Conclusion

This paper focuses on the improved version of StyleGAN2 and proposes a new image style transfer method to address the problems of mode collapse, detail blur and style distortion in traditional style transfer methods. By integrating the generator of the ResNet structure, the PatchGAN discriminator and the DCL loss function into the StyleGAN2 framework, this study constructs a more stable style transfer scheme that generates more realistic images.

First, from a qualitative analysis perspective, the style transfer experiments on the Horse2Zebra and Cityscapes datasets show significant visual advantages. Compared with traditional methods such as CycleGAN, CUT and DCLGAN, the proposed method performs better in maintaining the structural information of the original image, and the generated images are more prominent in details and clarity. For example, in the Horse2Zebra dataset, the method can accurately restore the zebra texture and avoid texture distortion and blur. On the Cityscapes dataset, the street view images after style transfer not only maintain coherence in content, but also show higher clarity and structural stability.

Secondly, quantitative evaluation using two commonly used indicators, Inception Score (IS) and Fréchet Inception Distance (FID), further confirmed the advantages of this method. On the Horse2Zebra dataset, the FID score of the proposed method is 13% lower than the current best DCLGAN, while the IS score is 6% higher. On the Cityscapes dataset, this method outperforms CycleGAN, CUT, and DCLGAN in both FID and IS indicators, especially in image edge detail processing and color contrast. These quantitative results show that the proposed method can generate style transfer effects that are closer to real images.

Finally, the effectiveness of DCL loss and FSeSim loss is verified by ablation experiments. The experimental results show that the generated image has obvious defects in detail and structure without using DCL loss and FSeSim loss. After introducing these loss functions, the details, color, light and structure of the image are significantly improved. Through quantitative analysis, experiments show that DCL loss and FSeSim loss can effectively improve the quality of style transfer images, making the overall image more natural and realistic. Through generalization experiments, it is verified that the generalization ability of the improved StyleGAN2 model on different data sets has been strengthened, which proves that the model is not only suitable for specific tasks, but also performs well in various image style transfer tasks.

In summary, the media image style transfer method based on improved StyleGAN2 proposed in this paper has made significant progress in many aspects. By introducing innovative generator structure and loss function, this method successfully solves some core problems in traditional media style transfer methods, and shows better generated image quality. Future research can further explore the application of this method in more complex media scenes and other fields, such as video style transfer, 3D image generation, etc., to expand its application scope.

Data availability

The datasets used in this study are publicly available and can be accessed through the following links: - Horse-2zebra dataset:https://opendatalab.org.cn/OpenDataLab/Horse2zebra. - Cityscapes dataset:https://opendatalab.org.cn/OpenDataLab/Cityscapes. - Facades datase: https://opendatalab.com/OpenDataLab/facade_cyclegan.

Received: 7 August 2025; Accepted: 21 November 2025
Published online: 04 December 2025

References

1. Huang, Q., Zheng, Z., Hu, X., Sun, L. & Li, Q. Bridging the gap between label-and reference-based synthesis in multi-attribute image-to-image translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14628–14637 (2021).
2. Briot, J., Hadjeres, G. & Pachet, F. Deep learning techniques for music generation—A survey. arXiv preprint arXiv:1709.01620 (2017).
3. He, X. & Deng, L. Deep learning for image-to-text generation: A technical overview. *IEEE Signal Process. Mag.* **34**, 109–116 (2017).
4. Liu, C., Gu, J., Yao, L. & Zhang, Y. Research on embroidery style migration model based on texture cycle GAN. *Int. J. Cloth. Sci. Technol.* **37**, 138–153 (2025).
5. Guo, X. & Ma, J. Heritage applications of landscape design in environmental art based on image style migration. *Results Eng.* **20**, 101485 (2023).
6. Borji, A. Pros and cons of GAN evaluation measures. *Comput. Vis. Image Underst.* **179**, 41–65 (2019).
7. Gao, Y. *et al.* High-fidelity and arbitrary face editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16115–16124 (2021).
8. Chen, Y. & Li, N. A study of style migration generation of traditional Chinese portraits based on Dualstylegan. *J. Eng. Des.* **1**, 1–14 (2024).
9. Zhoua, J., Wangb, Y., Gongc, J., Dongd, G. & Mae, W. Research on image style convolution neural network migration based on deep hybrid generation model. *Acad. J. Comput. Inf. Sci.* **4**, 83–89 (2021).
10. Creswell, A. et al. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **35**, 53–65 (2018).
11. Chen, Y., Shibata, H., Chen, L. & Takama, Y. Synthesis of comic-style portraits using combination of Cyclegan and pix2pix. *J. Adv. Comput. Intel. Intel. Inf.* **28**, 1085–1094 (2024).
12. Huang, J. *et al.* Design and verification of a wearable micro-capacitance test system for Poc biosensing. *IEEE Transactions on Instrumentation and Measurement* (2025).
13. Zhu, J., Park, T., Isola, P. & Efros, A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232 (2017).
14. Choi, Y. *et al.* Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8789–8797 (2018).
15. 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*, pp. 8188–8197 (2020).
16. Durall, R., Jam, J., Strassel, D., Yap, M. & Keuper, J. Facialgan: Style transfer and attribute manipulation on synthetic faces. arXiv preprint arXiv:2110.09425 (2021).
17. Liu, W., Chen, W., Yang, Z. & Shen, L. Translate the facial regions you like using self-adaptive region translation. *Proc. AAAI Conf. Artif. Intel.* **35**, 2180–2188 (2021).
18. Miao, J., Ning, X., Hong, S., Wang, L. & Liu, B. Secure and efficient authentication protocol for supply chain systems in artificial intelligence-based internet of things. *IEEE Internet of Things Journal* (2025).
19. Bai, R. & Bai, B. The impact of labor productivity and social production scale on profit-induced demand: Function and analysis from the perspective of marx's economic theory. *J. Xi'an Univ. Financ. Econ.* **37**, 3–17 (2024).
20. Wei, Y. *et al.* Maggan: High-resolution face attribute editing with mask-guided generative adversarial network. In *Proceedings of the Asian Conference on Computer Vision* (2020).
21. Qi, H. et al. Capacitive aptasensor coupled with microfluidic enrichment for real-time detection of trace sars-cov-2 nucleocapsid protein. *Anal. Chem.* **94**, 2812–2819 (2022).
22. Xing, X., Wang, B., Ning, X., Wang, G. & Tiwari, P. Short-term OD flow prediction for urban rail transit control: A multi-graph spatiotemporal fusion approach. *Inf. Fus.* **118**, 102950. https://doi.org/10.1016/j.inffus.2025.102950 (2025).
23. Shynar, Y. et al. Comprehensive analysis of blockchain technology in the healthcare sector and its security implications. *Int. J. E-Health Med. Commun.* **15**, 1–45 (2024).
24. Li, S., Hu, J., Zhang, B., Ning, X. & Wu, L. Dynamic personalized federated learning for cross-spectral palmprint recognition. *IEEE Transactions on Image Processing* (2025).

25. Kodipalli, A., Fernandes, S. L., Dasar, S. K. & Ismail, T. Computational framework of inverted fuzzy c-means and quantum convolutional neural network towards accurate detection of ovarian tumors. *Int. J. E-Health Med. Commun.* **14**, 1–16 (2023).
26. Hao, M. et al. A prompt regularization approach to enhance few-shot class-incremental learning with two-stage classifier. *Neural Netw.* **188**, 107453 (2025).
27. Bermano, A. et al. State-of-the-art in the architecture, methods and applications of stylegan. *Comput. Graphics Forum* **41**, 591–611 (2022).
28. DeMatteo, C. et al. The headaches of developing a concussion app for youth: Balancing clinical goals and technology. *Int. J. E-Health. Med. Commun.* **15**, 1–20 (2024).
29. Richardson, E. *et al.* Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2287–2296 (2021).
30. Huang, J., Liao, J. & Kwong, S. Unsupervised image-to-image translation via pre-trained stylegan2 network. *IEEE Trans. Multimed.* **24**, 1435–1448 (2021).
31. Back, J. Fine-tuning stylegan2 for cartoon face generation. arXiv preprint arXiv:2106.12445 (2021).
32. Wu, W., Huo, L., Yang, G., Liu, X. & Li, H. Research into the application of Resnet in soil: A review. *Agriculture* **15**, 661 (2025).
33. Almayyan, W. I. & AlGhannam, B. A. Detection of kidney diseases: Importance of feature selection and classifiers. *Int. J. E-Health Med. Commun.* **15**, 1–21 (2024).
34. Liao, J. *et al.* A machine learning-based feature extraction method for image classification using resnet architecture. *Digital Signal Processing* 105036 (2025).
35. Ma, W., Karakus, O. & Rosin, P. Patch-gan transfer learning with reconstructive models for cloud removal. arXiv preprint arXiv:2501.05265 (2025).
36. Seibel, M., Kepp, T., Uzunova, H., Ehrhardt, J. & Handels, H. Assessing spatial bias in medical imaging: An empirical study of patchgan discriminator effectiveness. In *BVM Workshop*, pp. 172–177 (Springer, 2025).
37. Xuan, S. & Zhang, S. Decoupled contrastive learning for long-tailed recognition. *Proc. AAAI Conf. Artif. Intel.* **38**, 6396–6403 (2024).
38. Su, D., Fan, B., Zhang, Z., Fu, H. & Qin, Z. Dcl: Diversified graph recommendation with contrastive learning. *IEEE Trans. Comput. Soc. Syst.* **11**, 4114–4126 (2024).
39. Barath, D. & Matas, J. Graph-cut ransac. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6733–6741 (2018).
40. Han, J., Shoeiby, M., Petersson, L. & Armin, M. Dual contrastive learning for unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 746–755 (2021).
41. Salimans, T. *et al.* Improved techniques for training gans. *Advances in neural information processing systems* **29** (2016).
42. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826 (2016).
43. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **30** (2017).
44. Dietz, T. *et al.* A study in dataset distillation for image super-resolution. arXiv preprint arXiv:2502.03656 (2025).
45. Martini, M. Measuring objective image and video quality: On the relationship between ssim and psnr for dct-based compressed images. *IEEE Transactions on Instrumentation and Measurement* (2025).

Author contributions
Yixuan Qin: Research design, Core model, Data collection, Experimentation, Results analysis, Writing original draft, Manuscript revisions.

Funding
No.

Declarations

Competing interests
The authors declare no competing interests.

Consent for publication
All authors of this manuscript have provided their consent for the publication of this research.

Additional information
Correspondence and requests for materials should be addressed to Y.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.