Research on Style Transfer of Unpaired Images Based on Improved CycleGAN

Mengzhuo Zhao
School of Computer Science and Engineering
Xi’an Technological University
Xi’an, China
E-mail: 3513353899@qq.com

Xiaoyi Lan
School of Computer Science and Engineering
Xi’an Technological University
Xi’an, China
E-mail: lanxiaoyi2004@163.com

Abstract—**To address the issues of uneven texture distribution, background mis-migration, and training imbalance in traditional CycleGAN for style transfer of non-paired horse and zebra images, this study proposes an improved model integrating dynamic attention mechanisms, semantic segmentation constraints, and adaptive training strategies. By embedding lightweight space-channel hybrid attention modules in generator residual blocks, the model enhances feature extraction in target regions. A lightweight semantic segmentation network is introduced to enforce local style transfer constraints, preventing redundant background migration. A two-stage adaptive loss weight adjustment strategy is designed to improve training stability. Experiments on the horse2zebra dataset using the PyTorch platform demonstrate that the improved model generates zebra stripes that conform to the subject structure, with non-target region pixel changes below 15.8%. Compared to the original CycleGAN, the model achieves an 8.3% SSIM improvement and 1.5dB PSNR enhancement. This model effectively resolves core limitations of traditional methods, providing a superior solution for style transfer of non-paired animal images.**

Keywords: Cyclegan; Image Style Transfer; Attention Mechanism; Semantic Segmentation; Adaptive Training Strategy

I. Introduction

With the rapid advancement of deep learning in computer vision, image style transfer technology has emerged as a research hotspot due to its extensive applications in artistic creation, film and television special effects, and image editing. The core of image style transfer lies in integrating the content of source domain images with the style of target domain images to generate new images that maintain both content integrity and stylistic consistency.

The CycleGAN model introduced in 2017 revolutionized the field by eliminating the need for paired training data in traditional methods. Through its cyclic consistency loss mechanism, it enabled cross-domain transfer between unpaired images, dramatically expanding the application scope of style transfer [1]. However, practical implementations still face notable limitations: the model often produces images with uneven texture distribution and blurred details, while its style transfer lacks specificity, frequently causing unintended background transfer. Moreover, during later training phases, the imbalance between discriminator and generator often leads to pattern collapse.

Style transfer between horses and zebras, a classic task in animal image transformation, faces the core challenge of accurately generating zebra-like stripes while preserving the horse's structural integrity, without introducing stylistic contamination in the background. Current approaches predominantly focus on single-module optimization, lacking coordinated design for feature extraction, transfer constraints, and training stability. To address these issues, this study implements multi-dimensional enhancements to traditional CycleGAN, aiming to improve the precision, stability, and visual quality of style transfer between equine and zebra images.

To address the aforementioned limitations, this study proposes a multi-module collaborative improvement CycleGAN model. Through systematic experiments on the horse2zebra dataset, we validate the model's effectiveness through both qualitative visual assessment and quantitative evaluation metrics. The proposed approach not only enhances the quality of style transfer between horse and zebra images but also provides a reference solution for style transfer tasks involving other animal images that are not paired.

The main research contents include:

First, a lightweight dynamic attention module is introduced into the generator to achieve differentiated feature extraction between target and background regions, thereby enhancing the precision of stripe texture generation.

Secondly, the lightweight semantic segmentation network is integrated to establish a local style transfer constraint mechanism, which effectively suppresses style diffusion in the background region.

Thirdly, a two-stage adaptive loss weight adjustment strategy is designed to balance adversarial loss and cyclic consistency loss, preventing pattern collapse during the later stages of training.

The following sections will provide a detailed elaboration on the aforementioned research topics. Regarding related work, we will review the current research status in areas such as non-paired image style transfer and attention mechanisms to clarify the research focus of this paper. Within the algorithm framework, we will elaborate on the overall architecture of the improved model, the design of its core modules, and algorithmic details. Subsequently, experimental validation will be conducted to demonstrate the model's effectiveness, with results analyzed from multiple dimensions including training stability, quantitative metrics, and visual effects. Finally, the research findings, limitations, and future prospects will be summarized.

II. RELATED WORK

This chapter focuses on the collaborative improvement of traditional CycleGAN through three major modules. It reviews the current research status in the relevant field, identifies the shortcomings of existing methods and the research gaps in this paper, and provides theoretical support for subsequent model design and experimental validation.

The core breakthrough in unsupervised image style transfer in this study originates from the introduction of CycleGAN. However, the original CycleGAN lacked targeted feature extraction, resulting in blurry details and unreasonable texture distribution in generated images. The aggregation-residual transform architecture proposed by Xie Sen et al. provided a solution for optimizing feature extraction [3]. These studies demonstrate that module structure optimization is an effective approach to enhance transfer performance, yet the challenges of capturing local features and balancing training for specific animal images remain unresolved. Therefore, this paper proposes an improved model integrating dynamic attention mechanisms, semantic segmentation constraints, and adaptive training strategies. The following section reviews the current developments of these three modules.

A. Application of Attention Mechanism in Style Transfer

The attention mechanism dynamically adjusts feature weights to focus the model on key regions, providing an effective approach to address local style transfer. Sanghyun Woo et al. 's CBAM attention module significantly enhances the feature recognition capability of convolutional neural networks through the synergistic interaction of channel attention and spatial attention [4], offering valuable insights for designing lightweight attention modules in this study. Wang's edge feature self-attention-based CycleGAN further validates the effectiveness of attention mechanisms in improving detail precision during style transfer [5].

Research on integrating attention mechanisms into CycleGAN has made notable progress. Li et al. combined edge detection with self-attention mechanisms [6], significantly improving style transfer performance on unpaired datasets. However, existing approaches often employ complex attention architectures, resulting in substantial computational overhead. Moreover, they lack customized designs tailored to the physiological characteristics of horses and zebras, making it challenging to accurately generate stripe textures.

B. Integration of Semantic Segmentation and Style Transfer

Semantic segmentation technology enables pixel-level category classification in images, enabling precise control over style transfer scope. Fully convolutional networks overcome the input size limitations of traditional convolutional neural networks, achieving efficient semantic segmentation.

Liu Zhe Liang et al. combined FCN with CycleGAN to locate target regions through semantic segmentation [10], effectively mitigating background mis-migration issues. However, their cascaded architecture resulted in complex training procedures and failed to address dynamic equilibrium during training. Moreover, existing fusion methods are primarily designed for general scenarios, lacking targeted optimization for the specific structures of animal images, leaving room for further improvement in transfer performance.

C. Analysis of Training Strategy Optimization Methods

Training imbalance is a prevalent challenge for GAN models. Current solutions primarily involve static strategies like adjusting learning rates or optimizing loss function weights, yet these approaches struggle to adapt to varying training requirements. The context loss proposed by Mechrez et al. [7] offers a novel direction for texture consistency optimization, while Ramachandran et al.'s research on activation functions [8] provides valuable insights for enhancing network training stability.

In summary, while existing research on each module has made progress, several limitations remain. Firstly, feature extraction lacks specificity, making it difficult to accurately capture the subject area characteristics of animals. Secondly, the scope of style transfer is not precisely controlled, with prominent background mistransfer issues. Thirdly, the training strategy lacks dynamic adaptability, leading to potential imbalance in later stages. To address these challenges, this paper proposes a multi-module collaborative improvement CycleGAN model, achieving high-quality style transfer for non-paired images of horses and zebras.

III. ALGORITHM FRAMEWORK

This chapter addresses the core limitations of traditional CycleGAN by designing a multi-module collaborative improvement framework. The approach focuses on three key dimensions: generator optimization, transfer range constraints, and training strategies. It provides a detailed explanation of each core module's architecture, operational principles, and loss function construction, offering a comprehensive methodological foundation for subsequent experimental validation.

A. Overall Model Architecture

The proposed improved CycleGAN model maintains the original "dual generator + dual discriminator" ring structure, with three key enhancements: a dynamic attention generator, a semantic segmentation constraint module, and an adaptive training strategy. The model architecture is illustrated in Figure 1.

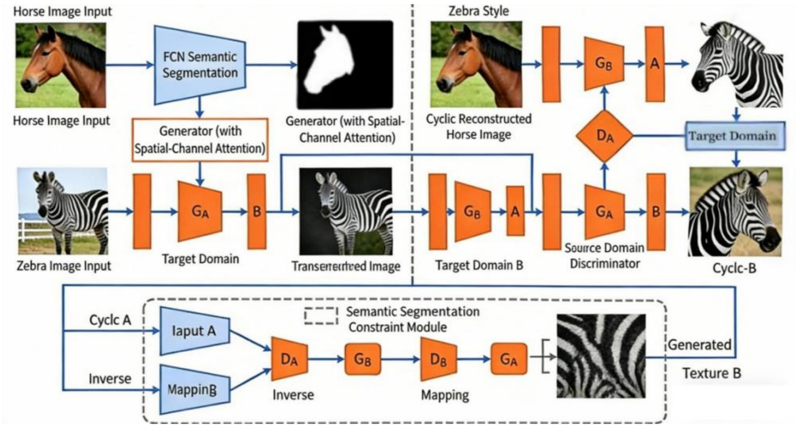

Figure 1. Schematic diagram of the improved CycleGAN model architecture

The model's core objective is to learn the mapping function G$_{A\to B}$ from the source domain (Horses) to the target domain (Zebra) and its inverse mapping G$_{A\to B}$. Through the coordinated operation of cyclic consistency loss, adversarial loss, semantic constraint loss, and texture consistency loss, it achieves precise and stable style transfer.

B. Design of Dynamic Attention Generator

The generator employs an "encoder-converter-decoder" architecture, where lightweight space-channel hybrid attention modules are embedded in the residual blocks of the converter to replace the traditional feature extraction method with fixed weights.

Generator Infrastructure

The encoder consists of three convolutional blocks, each containing a 7×7 convolution layer, an instance normalization layer, and a ReLU activation function with a stride of 2. This architecture performs image downsampling and feature extraction, ultimately outputting 25664×64 feature vectors.

The converter is designed for 256×256 input resolution, and it uses six improved residual blocks to realize the inter-domain feature transformation, which can retain the main structure features and integrate the target style features.

The decoder consists of two up-sampling blocks and one output convolutional layer. The up-sampling block includes a transposed convolutional layer, an instance normalization layer, and a ReLU activation function, while the output convolutional layer employs a tanh activation function to generate a 256×256×3 target image.

Lightweight Space-Channel Hybrid Attention Module

The attention module consists of a channel attention branch and a spatial attention branch, as shown in Figure 2.

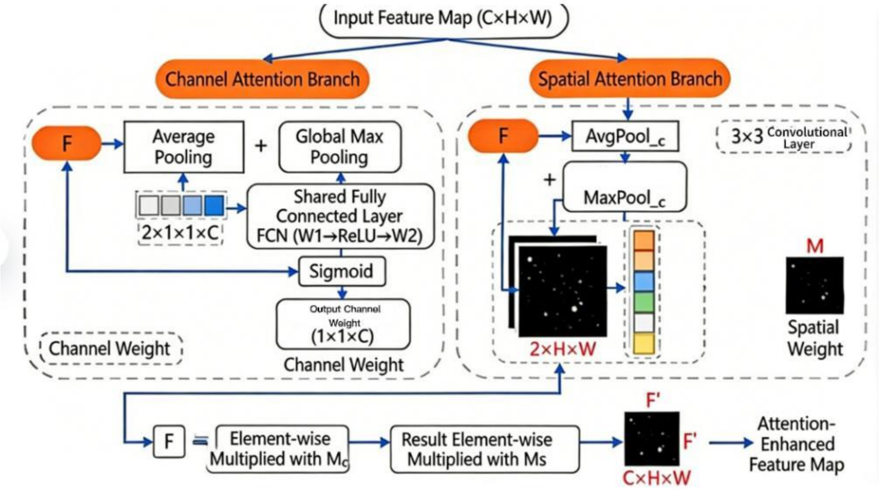

Figure 2. Structural diagram of lightweight space-channel hybrid attention module

The algorithm enhances key channels (e.g. texture and contrast) while suppressing background redundancy by learning weights of zebra stripe style correlation feature channels, as shown in Equation (1).

$$M_C\left(F\right)=W_2\delta\left(W_1\delta\left(W_1\begin{bmatrix}AvgPool\left(F\right);\\ MaxPool(F)\end{bmatrix}\right)\right) \quad (1)$$

Here, AvgPool and MaxPool represent global average pooling and global maximum pooling respectively, W_1 and W_2 denote the weights of the fully connected layer, δ is the ReLU activation function, and σ is the Sigmoid activation function.

Spatial attention computation involves learning the subject's contour mask to focus on core regions such as the torso and limbs, with the calculation process illustrated in Equation (2):

$$M_C\left(F\right)=\sigma\left(Conv_{3\times3}\left(\begin{bmatrix}AvgPool\left(F\right);\\ MaxPool(F)\end{bmatrix}\right)\right) \quad (2)$$

$AvgPool_C$ and $MaxPool_C$ are channel-wise average pooling and max pooling respectively, while $\mathrm{Conv}_{3\times3}$ denotes a 3×3 convolutional layer.

Feature fusion: The channel weights are multiplied with the spatial weights element-wise and applied to the original feature map to enhance key region features, as shown in Equation (3):

$$F^{'}=F\otimes M_C\left(F\right)\otimes M_s\left(F\right) \quad (3)$$

C. Semantic Segmentation Constraint Module

The lightweight FCN semantic segmentation network [11] is introduced as a pre-constraint module to achieve precise control of transfer scope, as shown in Figure 3.

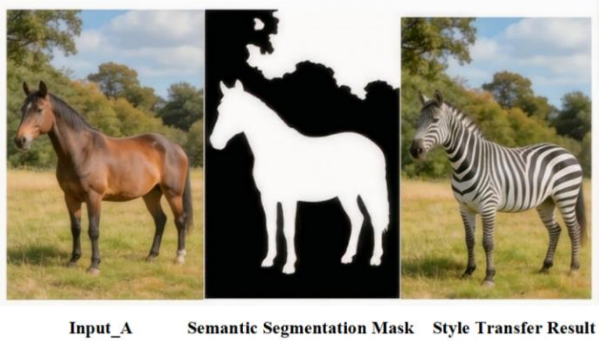

Figure 3. Semantic Segmentation Flowchart

The semantic segmentation network training employs MobileNet as the backbone network [9], replacing the fully connected layers of traditional FCN to construct a lightweight semantic segmentation model. The training dataset consists of annotated data extended from the horse2zebra dataset, containing 200 labeled images. The annotation data was labeled using the LabelMe tool, focusing on the main contour of horses, with an annotation accuracy exceeding 85%. The network output is a binary mask Y of the same size as the input image, where $\mathrm{Y}(i,j)=1$ indicates the target migration region at pixel (i,j), and $\mathrm{Y}(i,j)=0$ indicates the non-migration region.

The semantic constraint loss function is designed to enforce the generator to perform style transfer exclusively within the target region. It introduces the semantic constraint loss L_{seg}, which calculates the pixel-wise matching loss between the generated image and the mask image, as shown in Equation (4):

$$L_{seg}=\frac{1}{H\times W}\sum\nolimits_{i=1}^{H}\sum\nolimits_{j=1}^{W}\ |Y\left(i\cdot j\right)\cdot \left(G_{A\to B}\left(X_A\right)\left(i\cdot j\right)-X_A\left(i\cdot j\right)\right) \quad (4)$$

Here, X_A denotes the source domain horse image, $\mathrm{G}_{A\to B}(\mathrm{X}_A)$ represents the generated zebra image, Y is the mask image, and $H\times W$ indicates the image dimensions. The loss function penalizes pixel variations outside the target region to suppress background mis-migration.

D. Adaptive Training Strategies

A two-stage adaptive training strategy is designed, combining dynamic loss weight adjustment and texture consistency loss to address the imbalance issue in the later training phase.

1) Two-Phase Training Division

In the first phase, the focus is on feature alignment and structural preservation. The cyclic consistency loss weight is set to $\lambda=15$,and the adversarial loss weight to $\alpha=0.5$,prioritizing the retention of the horse's main structure in the generated images to prevent early pattern collapse.

In the second phase, the focus shifts to style refinement and training balance. The adaptive weight factor β is introduced to dynamically adjust the adversarial loss weight based on the loss difference between the discriminator and generator, as shown in Equation (5):

$$\alpha=\begin{cases}1.2 & \text{if } |L_D-L_G|>0.5 \text{ and } L_D<0.1\\ 0.8 & \text{if } |L_D-L_G|>0.5 \text{ and } L_G>1.0\\ 0.5 & \text{otherwise}\end{cases} \quad (5)$$

Here, L_D denotes the discriminator's average loss and L_G the generator's average loss. When the discriminator becomes overconfident ($L_D < 0.15$), increasing α enhances the generator's adversarial capability; conversely, when the generator underperforms ($L_G > 0.9$), decreasing α prevents suppression.

2) *Texture Consistency Loss*

To ensure the rationality of zebra stripe generation, the texture consistency loss $\mathcal{L}_{\text{texture}}$ [12] is introduced to calculate the statistical differences in texture between the generated image and the real zebra image, as shown in Equation (6):

$$\mathcal{L}_{\text{texture}} = \frac{1}{C \times C} \sum_{i=1}^{C} \sum_{j=1}^{C} \left| \begin{matrix} \text{Gram}\left(\text{G}_{A \to B}(\text{X}_A)\right) \\ (i, j) - \text{Gram}(\text{X}_B)(i, j) \end{matrix} \right| \quad (6)$$

Here, Gram denotes the Gram matrix calculation, X_B represents the real zebra image, and C is the number of feature channels. This loss constraint ensures that the generated stripes maintain the same direction and density as the real zebra.

3) *Total Loss Function*

The total loss function of the model is the weighted sum of all loss terms, as shown in Equation (7):

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{gan}} + \lambda \mathcal{L}_{\text{cycle}} + \gamma \mathcal{L}_{\text{seg}} + \mu \mathcal{L}_{\text{texture}} \quad (7)$$

In this framework, $\lambda = 0.8$ represents the semantic constraint loss weight, $\mu = 0.5$ denotes the texture consistency loss weight, λ is the cyclic consistency loss weight, and α is the adversarial loss weight.

E. Flowchart of the Overall Model Implementation

To clearly present the entire workflow logic and inter-module collaboration, we created a comprehensive flowchart as shown in Figure 4, which specifies the inputs, processing logic, and outputs for each step, enabling visualized and traceable workflow management.

Figure 4 presents the implementation flowchart of the enhanced CycleGAN model, which systematically integrates four core improvement modules. The semantic segmentation module serves as a pre-training constraint, while the dynamic attention module and adaptive strategy are incorporated into the two-stage training process. Notably, the flowchart also highlights the feedback loop between the validation phase and the training phase, where the performance metrics obtained from the validation dataset are fed back to adjust the hyperparameters of the adaptive strategy in real time, ensuring the model converges to the optimal state efficiently.

This diagram clearly illustrates the complete implementation chain of "pre-training-training-validation-iteration", highlighting the collaborative relationships among modules and providing clear guidance for model replication.

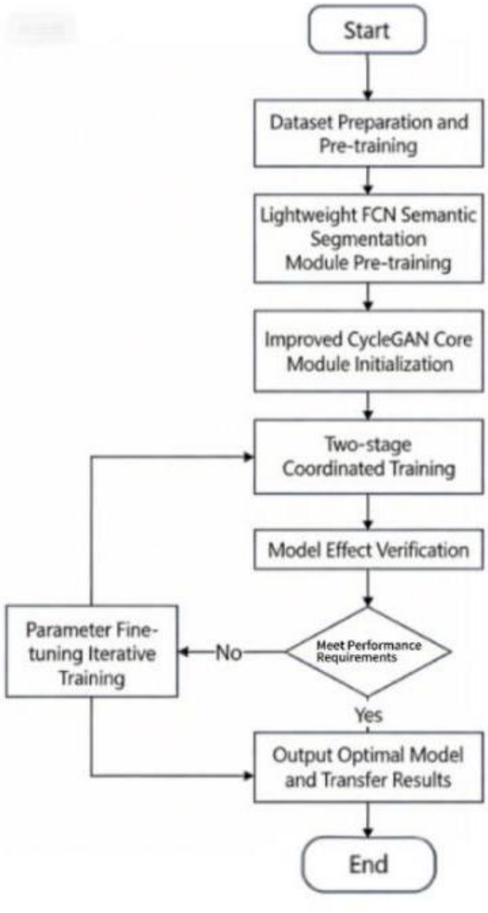
Start
Dataset Preparation and Pre-training
Lightweight FCN Semantic Segmentation Module Pre-training
Improved CycleGAN Core Module Initialization
Two-stage Coordinated Training
Model Effect Verification
Meet Performance Requirements
No
Parameter Fine-tuning Iterative Training
Yes
Output Optimal Model and Transfer Results
End

Figure 4. Schematic flowchart of the improved CycleGAN model implementation

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section specifies the hardware and software environment, dataset configuration, and preprocessing protocols, establishing a stable and standardized foundation for subsequent model training and performance evaluation.

A. Experimental Environment and Dataset

Building upon the improved CycleGAN architecture and implementation process outlined earlier, this chapter establishes a standardized experimental environment to validate the efficacy of core modules and overall model performance. By specifying dataset configurations and hyperparameter settings, systematic experiments will be conducted to analyze training stability, quantitative metrics, and visual effects across multiple dimensions.

1) Experimental Environment

The experiment was conducted using the PyTorch deep learning framework on an AMD Ryzen 7 5800U processor running Windows 11, with Python 3.9 as the programming language. The required libraries included Torch 1.12.0, OpenCV 4.5.5, NumPy 1.21.6, and Matplotlib 3.5.3. As shown in Table 1:

TABLE I. EXPERIMENTAL ENVIRONMENT PARAMETER CONFIGURATION TABLE

Environmental category	Configuration
hardware environment	CPU: AMD Ryzen 7 5800U
operating system	Windows 11
programming language	Python 3.9
core dependency library	Torch 1.12.0、OpenCV 4.5.5、NumPy 1.21.6、Matplotlib 3.5.3
Deep learning framework	PyTorch

By integrating model refinement with a two-stage training strategy and conducting multiple rounds of debugging, we identified the core hyperparameters. The key parameters are listed in the table below, with values optimized for experimental conditions and dataset characteristics to ensure stable training and transfer performance. As shown in Table 2:

TABLE II. CORE HYPERPARAMETER CONFIGURATION TABLE OF THE MODEL

Hyperparameter name	short-cut process
Image resolution	256×256×3
batch size （BatchSize）	16
initial learning rate	0.0002
Total training rounds	200 (in two phases)
reciprocal consistency loss weight (λ)	15 (Phase 1) 10 (Phase 2)
Semantic/textural loss weight (γ / μ)	0.8/0.5
optimizer	Adam

2) data set

The experiment utilized the standard Horse2Zebra dataset, which contains unpaired images of horses and zebras. The dataset was divided into three sets: training set, validation set, and test set. The training set comprised 1,067 horse images (trainA) and 1,334 zebra images (trainB), the validation set included 100 horse images and 100 zebra images, and the test set consisted of 120 horse images (testA) and 140 zebra images (testB).

All images were uniformly resized to a resolution of 256×256 . Data augmentation techniques such as random flipping, translation, and brightness adjustment were employed to expand the training set, thereby enhancing the model's generalization ability.

B. Experimental Results and Analysis

The loss variation during training serves as a core metric to evaluate model stability and convergence. To comprehensively analyze the regulatory effects of each improvement module on training dynamics, the model's validity is validated through multi-dimensional curve comparison.

1) Analysis of Training Process

The loss evolution during training is illustrated in Figure 5. The x-axis represents training iterations, while the y-axis shows the discriminator's average loss value. This curve directly reflects the training intensity of the discriminator and the model's training equilibrium. The improved model incorporating dynamic attention mechanisms, semantic segmentation constraints, and adaptive training strategies demonstrates significantly enhanced stability in its discriminator loss. Throughout the training process, the loss value remains consistently stable without extreme fluctuations or excessive convergence to zero.

This positive trend stems from the synergistic interaction of multiple modules. The adaptive training strategy dynamically adjusts loss weights in two phases, effectively balancing the adversarial relationship between the discriminator and generator to prevent overemphasis on either component. The dynamic attention module enhances feature extraction in target regions, reducing the discriminator's reliance on irrelevant features and minimizing abrupt loss fluctuations. Semantic segmentation constraints further define the transfer scope, allowing the discriminator's evaluation to focus specifically on stripe generation quality within the horse's main body area, thereby improving training stability.

The comparison demonstrates that the improved model, through multi-module collaborative optimization, successfully resolved the imbalance issue in the later stages of original CycleGAN training. It achieved stable convergence of discriminator loss, establishing a solid training foundation for generating high-quality zebra images. This also validates the effectiveness of the added modules in regulating training dynamics and preventing pattern collapse.

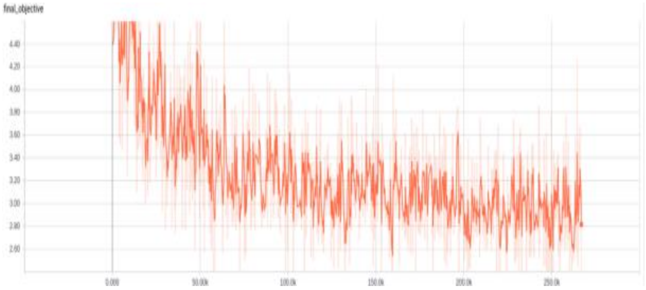

Figure 5. Overall objective function loss curve

The analysis of the training process shows that the whole model proposed in this paper achieves the dynamic balance of the training process through the multi-module cooperation, and solves the problem of the imbalance in the later stage of the traditional model.

2) *Quantitative Indicator Comparison*

Quantitative metrics validate the improved model's advantages in image quality, transfer accuracy, and training stability through objective data analysis, while the visual effects and detail representation of generated images require further qualitative validation. By comparing style transfer results between the improved model and the original CycleGAN using typical test set samples, we focus on analyzing stripe texture distribution, background contamination control, and subject structure preservation to visually demonstrate the practical effectiveness of the improved module, as shown in Figure 3.

In conclusion, compared with the original CycleGAN, the improved model demonstrates significant optimization across four core metrics, with comprehensive enhancements in image quality, transfer accuracy, and training stability, which directly validates the effectiveness of multi-module collaborative improvement.

TABLE III. COMPARISON OF QUANTITATIVE INDICATORS

Evaluating indicator	Original CycleGAN	improved model	Enhance
peak signal to noise ratio	28.3 dB	29.8 dB	+1.5 dB
structural similarity	0.72	0.78	+8.3%
non target area change rate	23.6%	15.8%	-15.8%
loss stability	0.189	0.082	-55.8%

3) *Qualitative Visual Effect Comparison*

The visual effects of style transfer for each model are shown in Figure 6.

The input image depicts a horse standing on a grassy background with sharp contours and a clean, texture-free background. The original CycleGAN output exhibits significant flaws: zebra stripes are unevenly distributed, with broken and distorted patterns on the torso and limbs. Moreover, the grass background is heavily contaminated by zebra stripes, resulting in extensive mis-migration that compromises the image's overall authenticity.

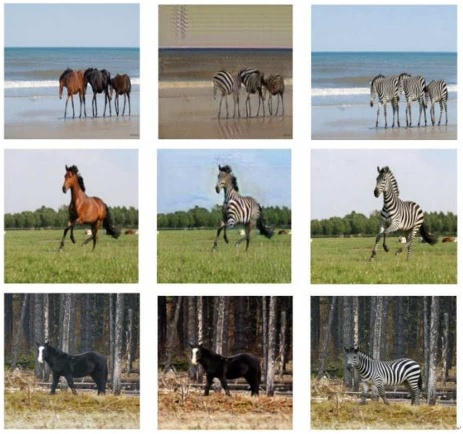

Figure 6. Comparison of visual effects for style transfer

The enhanced method demonstrates superior image quality, thanks to the synergistic effects of the dynamic attention module, semantic segmentation constraints, and adaptive training strategy. The stripes appear natural without breakage or overlap issues, while the semantic segmentation effectively isolates the background. The zebra texture pollution in grassy areas is significantly reduced, with only a few edge pixels affected, fully preserving the original background features. The generated images maintain the integrity of the horse's main form while retaining the authentic texture of zebra stripes, achieving a visual effect that far surpasses the original CycleGAN.

The qualitative analysis shows that the improved model can accurately generate zebra stripes conforming to physiological characteristics in the main region of the horse, effectively avoiding background mis-migration, and the generated images have remarkable realism and detail integrity.

4) *Ablation Experiment Analysis*

To validate the independent contributions of each improvement module, ablation experiments were conducted, with results presented in Table 4.

The experimental results show that the three improved modules have positive contributions to the model performance, and there are obvious synergistic benefits.

The implementation of dynamic attention mechanisms in the original CycleGAN model achieved significant performance improvements: PSNR increased from 28.3dB to 29.5dB, SSIM improved from 0.72 to 0.77, and non-target region change rate decreased from 23.6% to 18.4%. These results demonstrate that dynamic attention effectively directs the model's focus to the image's main structure while reducing background interference, thereby enhancing the structural consistency of generated images.

TABLE IV. ABLATION EXPERIMENT RESULTS

Model configuration	PSNR（dB）	SSIM	Change rate of non-target area I (%)	loss stability σ
Original CycleGAN	28.3	0.72	23.6	0.189
Original CycleGAN+ Lightweight Spatial-Channel Mixed Attention Module	29.5	0.77	18.4	0.153
Original CycleGAN+ Lightweight FCN Segmentation Module	29.1	0.74	16.2	0.108
Original CycleGAN+ Adaptive Training Strategy	28.7	0.73	22.1	0.091
Full strategy	29.8	0.78	15.8	0.082

When semantic segmentation constraints were applied separately, the PSNR improved to 29.1dB and SSIM reached 0.74, with the non-target region change rate dropping significantly to 16.2%. These results demonstrate the critical role of semantic segmentation constraints in precisely controlling background migration extent and suppressing background contamination.

The introduction of adaptive training strategy significantly reduced the loss stability from 0.189 to 0.091, indicating that this strategy effectively mitigated loss fluctuations during training and enhanced the model's convergence stability.

When the three modules operate in synergy, the model achieves optimal performance with a PSNR of 29.8dB, SSIM of 0.78, and a reduction in non-target region change rate to 15.8%, while loss stability further improves to 0.082. These results conclusively demonstrate the effectiveness of multi-module collaborative optimization. The independent contributions of each module and their synergistic effects mutually corroborate, indicating that the proposed improvement strategy in this study possesses clear interpretability and reproducibility.

To visually demonstrate the independent functions and synergistic effects of each improvement module, Figure 7 presents a comparative visualization of ablation experiment data. All images are extracted from typical test set samples, uniformly using a horse against grass background as input to ensure fair comparison.

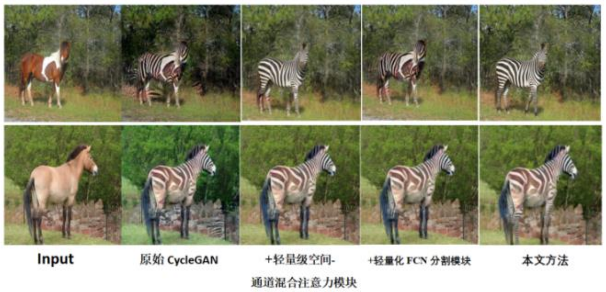

Figure 7. Comparison of Ablation Experiments

The image input is a horse with clear outline standing on grass background, without redundant texture interference, which is convenient for intuitive observation of the migration effect.

The original CycleGAN results exhibited chaotic zebra stripes with distorted trunk patterns, fragmented limb stripes, and noticeable grass background contamination. When incorporating the dynamic attention module, stripe distribution became more aligned with the horse's body and limbs, significantly reducing distortion and fragmentation. However, slight texture contamination in the grass background persisted, demonstrating the attention module's enhanced feature extraction. The semantic segmentation module alone effectively suppressed background noise, leaving grass areas free from stripe interference, though localized misalignment remained, resulting in limited quality improvement. The integrated strategy achieved optimal performance: stripes became naturally slender, perfectly matching the horse's body curvature, joint contours, and limb lines without fragmentation or overlap. The background remained virtually uncontaminated, while image clarity and structural consistency reached peak levels.

The data in Table 4 and the visualization in Figure 7 demonstrate that while individual modules can only address specific issues, their combined operation creates complementary advantages. This synergy resolves the core limitations of traditional models, thereby validating the rationality and effectiveness of the multi-module collaborative improvement strategy proposed in this study.

V. SUMMARY AND CONCLUSIONS

This paper proposes a multi-module collaborative improvement model to address the issues of uneven texture distribution, background mis-migration, and imbalance in the later training phase of traditional CycleGAN for style transfer between non-paired images of horses and zebras.

The design incorporates a lightweight space-channel hybrid attention mechanism to achieve precise extraction of target region features, enhancing the rational distribution and detail integrity of zebra stripes. By integrating a lightweight semantic segmentation network and constructing a semantic constraint loss function, the system maintains non-target region pixel change rates below 15.8%, effectively addressing background mis-migration issues. Finally, a two-stage adaptive training strategy is proposed, combining dynamic loss weight adjustment with texture consistency loss to maintain dynamic balance during training, thereby preventing post-training imbalance and pattern collapse.

Experiments on the horse2zebra dataset demonstrate that the improved model achieves a PSNR of 29.8dB and SSIM of 0.83, outperforming both the traditional CycleGAN and the single-improved model. The generated images exhibit significantly enhanced visual fidelity and realism.

Building on the findings of this study, we propose expanding the research scope by investigating end-to-end joint training between semantic segmentation and CycleGAN to optimize module compatibility. Furthermore, the model is transferred to other animal texture transfer tasks such as cats, tigers, and wolf dogs to validate its cross-scenario generalization capability, thereby enhancing its practical application value.

Future research will focus on deepening studies in core directions. A semantic segmentation and CycleGAN end-to-end joint training framework will be constructed, with customized attention mechanisms designed for zebra stripes. The framework will be extended to multi-animal image transfer tasks while optimizing performance in low-resolution and complex background scenarios, thereby enhancing the model's generalization ability and practical value.

REFERENCES

[1] Zhu J Y, Park T, Isola P, Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2242-2251.

[2] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2414-2423.

[3] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1492-1500.

[4] Woo S, Park J, Lee J Y, Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. 2018: 3-19.

[5] Wang L, Wang L, Chen S. ESA-CycleGAN: Edge feature self-attention based cycle-consistent generative adversarial network for style transfer[J]. IET Image Processing, 2022, 16(1): 176-190.

[6] Li C, Wand M. Combining Markov random fields and convolutional neural networks for image synthesis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2479-2486.

[7] Mechrez R, Talmi I, Zelnik-Manor L. The contextual loss for image transformation with non-aligned data[C]//Proceedings of the European Conference on Computer Vision. 2018: 768-783.

[8] Ramachandran P, Zoph B, Le Q V. Searching for activation functions[C]//Proceedings of the 35th International Conference on Machine Learning. 2018: 4095-4104.

[9] Liu J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[10] Wang H, Lu X, Deng F. Improving CycleGAN for Image-to-Image Style Transfer by DenseNet[C]//2022 7th International Conference on Computer and Communication Systems. 2022: 326-330.

[11] Castillo C, De S, Han X T, Castillo C, De S, Han X T, et al. Son of Zorn's lemma: targeted style transfer using instance-aware semantic segmentation[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. 2017: 1348-1352.

[12] Azadi S, Fisher M, Kim V, Azadi S, Fisher M, Kim V, et al. Multi-content GAN for few-shot font style transfer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 7333-7341.