A Deep Learning-Based Non-Photorealistic Rendering (NPR) Generation Method

Wei Zhou
Xi'an University of Technology
Xi'an, China
1944692418@qq.com

Minnan Cang*
Xi'an University of Technology
Xi'an, China
*Corresponding author: cangminnan@xaut.edu.cn

Abstract—With the advancement of deep learning technologies, non-photorealistic rendering (NPR) techniques have made significant breakthroughs in the field of computer graphics. This paper proposes a deep learning-based framework for NPR generation, combining convolutional neural networks (CNNs) and generative adversarial networks (GANs), aimed at automatically generating graphics in various artistic styles, such as sketches, ink paintings, and oil paintings. Through image style transfer and feature learning, the framework optimizes the preservation of details and enhances the artistic effects of the generated images. Experimental results demonstrate that the proposed method significantly outperforms traditional algorithms in terms of style transfer quality, generation efficiency, and adaptability. It shows strong potential for wide applications in animation, game design, and virtual reality, providing robust technical support for digital art creation.

Keywords: Non-Photorealistic Rendering; Deep Learning; Convolutional Neural Networks; Generative Adversarial Networks (GAN)

I. INTRODUCTION

With the rapid development of artificial intelligence, deep learning has demonstrated immense potential across various fields, particularly in non-photorealistic rendering (NPR) generation within computer graphics, which has become a prominent area of research. NPR technology aims to simulate traditional artistic styles and visual effects to generate images with rich artistic expression, widely used in animation production, game design, virtual reality, and other digital art creations. However, traditional NPR methods typically rely on complex graphic algorithms and manually defined artistic rules, which are limited by unnatural style transfer, low generation efficiency, and insufficient detail preservation.

In recent years, deep learning technologies, particularly convolutional neural networks (CNNs) and generative adversarial networks (GANs), have provided new perspectives for NPR techniques. By automatically learning the visual features of artistic works, deep neural networks enable style transfer and artistic generation without manual intervention, thereby enhancing the artistic quality and expressiveness of the images. CNNs efficiently extract low-level image features, such as edges, textures, and colors, while GANs, through adversarial training between the generator and discriminator, further optimize image quality to ensure the generated images possess higher artistic expression.

Although some studies have attempted to apply deep learning to NPR generation, challenges remain in handling complex style transfers, detail preservation, and adaptive generation. Existing methods are often limited to a specific artistic style or lack sufficient attention to detail, resulting in a lack of flexibility and accuracy. Therefore, this paper proposes a deep learning framework combining CNNs and GANs, aiming to optimize the quality of style transfer and graphical detail representation. The innovation of this research lies in the introduction of a multi-scale discrimination approach, which enhances the preservation of details and the consistency of style during the style transfer process. This study designs a framework that integrates CNNs and GANs to improve the style transfer effect, ensuring detail preservation and visual consistency across different artistic styles.

II. RELATED WORK

One of the early style transfer methods based on CNNs, proposed by Gatys et al，has become a significant starting point for NPR research. However, more recent studies (e.g., Zhang et al., 2020)[1] have pointed out limitations in detail preservation and style consistency, especially in complex texture generation. In recent years, advances have focused on overcoming these challenges by integrating more sophisticated neural architectures and improved loss functions. For instance, Huang et al. (2021) [2]improved the CNN-based style transfer method by introducing perceptual similarity metrics, which offer more precise control over image texture and fine details.

CycleGAN, introduced by Zhu et al., remains a cornerstone for unsupervised image translation. However, the method's limitations in detail control have been addressed by recent improvements like those by Chen et al. (2022) [3], who proposed a dual-cycle consistency method to improve the preservation of details in transferred images. Additionally, Zhang and Liu (2022) [4] explored fine-tuning CycleGAN's adversarial loss to enhance the preservation of texture and style consistency during style transfer.

DCGAN and its conditional variant have been used in various applications for generating realistic images. However, recent studies like those by Zhao et al. (2021) [5]have found that these models still struggle with fine-grained style transfer. A promising approach is the use of hybrid GAN models, such as those proposed by Kim et al. (2023)[6], which combine cGAN and attention mechanisms to improve both style and detail preservation.

Recently, style transfer methods based on self-attention mechanisms (Zhu et al., 2020)[7] have gained attention. The self-attention mechanism can generate more detailed and consistent artistic effects by focusing on the relationships

between different regions of an image. However, these methods are typically computationally complex and tend to produce unnatural visual effects when handling style transfer, especially in terms of detail preservation. To address this issue, the multi-scale discriminative approach proposed in this paper processes details at different scales, better adapting to the needs of various artistic styles and enhancing both style consistency and detail preservation.

In summary, while existing deep learning methods have made significant progress in NPR generation, they often have limitations in detail preservation, style consistency, and adaptive generation. Therefore, this paper proposes a novel framework combining CNNs and GANs, introducing a multi-scale discriminative method to overcome the shortcomings of existing approaches in detail control and style consistency. Through this innovative framework, our research not only improves the quality of style transfer but also ensures the preservation of details and visual consistency in generated images across multiple artistic styles[8][9].

III. DATASET CONSTRUCTION AND PREPROCESSING

This study constructs a diverse art image dataset, covering various styles such as ink painting, oil painting, sketching, and pencil drawings. Each style contains thousands of images, ensuring the diversity and representativeness of the dataset. All images are resized to 256x256 pixels and undergo color normalization to eliminate differences in brightness and contrast, making them suitable for input into deep learning models. The dataset also undergoes data augmentation, including rotation, scaling, and flipping, to enhance the model's generalization ability and robustness. Furthermore, to extract the core features of artistic styles, noise and background interference in the images are removed, ensuring that the model can focus on key elements such as brushstrokes, textures, and color distribution.

IV. DEEP LEARNING MODEL DESIGN

This study proposes a deep learning framework that combines Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) for efficient Non-Photorealistic Rendering (NPR) generation and optimization of style transfer quality[10][11]. In this framework, CNN is responsible for extracting low-level features from input images, while GAN, through an adversarial training mechanism, optimizes the quality of the generated images, ensuring that details are preserved and artistic expressiveness is enhanced during the style transfer process.

A. Application of Convolutional Neural Networks (CNN)

The primary role of Convolutional Neural Networks (CNN) in image processing is to extract low-level features from input images, such as edges, textures, and colors. Through the stacking of multiple convolutional layers, CNNs can progressively capture both local structures and global semantic information in images. These low-level features provide a rich foundation for subsequent style transfer and image generation, enabling the preservation of important detail information during the style transformation process. In this study, residual connections and batch normalization techniques are incorporated into the CNN architecture to effectively mitigate

the vanishing gradient problem and enhance the model's training efficiency and stability. Residual connections, by using skip connections that bypass several layers, help in the effective transmission of information. Batch normalization normalizes the input to each layer in the network, promoting faster convergence during training and improving the model's generalization ability. The convolution operation can be described by the following formula:

$$Y = f(w * X + b) \tag{1}$$

In the equation, X represents the input image, w is the convolutional kernel, b is the bias term, $*$ denotes the convolution operation, and f is the activation function (ReLU).

To further optimize the style transfer effect and preserve image details, Perceptual Loss and Total Variation Loss are introduced. Perceptual loss measures the difference between images in a high-level feature space, ensuring that during style transfer, the image's textures, colors, and other visual features transition smoothly and naturally. Total Variation Loss effectively reduces noise in the image, enhancing its smoothness and detail preservation, thereby improving the overall visual quality. Perceptual loss can be described as:

$$L_{\text{perceptual}} = \sum_i \|\emptyset_i(I) - \emptyset_i(G(I)) -\|_2^2 \tag{2}$$

In the equation, $\emptyset_i(I)$ and $\emptyset_i(G(I))$ represent the activation values of the input image I and the generated image $G(I)$ at a particular layer's feature map. Total Variation Loss aims to reduce noise in the image and enhance its smoothness, which can be described as:

$$L_{TV} = \sum_{i,j}(|I_{i,j} - I_{i+1,j}| + |I_{i,j} - I_{i,j+1}|) \tag{3}$$

B. GAN Part Design

Generative Adversarial Networks (GAN) are the core framework of this study. Through adversarial training between the generator and discriminator, the generator produces new images in each training iteration, while the discriminator evaluates the differences between the generated and real images. By continuously optimizing both the generator and discriminator, GAN can generate high-quality images while enhancing the artistic expression and detail retention of the images.

1) Generator

The task of the generator is to produce images with the target artistic style from the input image. Its network architecture consists of multiple convolutional layers, deconvolutional layers, and activation functions (ReLU). Convolutional layers are used to extract semantic information from the input image, while deconvolutional layers progressively upsample the low-resolution image to a higher resolution, refining the style details.

- Convolutional Layer: Extracts low-level features from the source image and integrates the target artistic style.

- Deconvolutional Layer: Gradually transforms the low-resolution image into a high-resolution image, adding finer details and resolution.

982

- Style Transfer Module: Adjusts the texture, color, and structure of the image to match the target artistic style.

2) Multi-Scale Discrimination Method

To further enhance the retention of image details and artistic expressiveness, a multi-scale discrimination method is introduced. Traditional GAN discriminators typically focus only on global image features, which may lead to loss of local details or blurry images. To overcome this, a multi-scale discriminative mechanism is employed, evaluating the image at multiple scales to precisely capture both the local details and global structure of the image.

- Local Feature Level: The low-resolution discriminator focuses on the image's local details, such as texture, edges, and small-scale features, ensuring that no detail is lost or blurred during style transfer.

- Detail Fine-tuning Level: The discriminator further assesses the texture, brushstrokes, and lighting details to optimize subtle differences in the artistic style, enhancing the image's artistic expressiveness.

By combining these three levels, the multi-scale discrimination method effectively addresses the limitations of traditional NPR generation methods, such as loss of details and inconsistency in style. It enhances the generator's ability to recover details and ensures a natural style transfer.

3) Ablation Study

An ablation study was conducted to evaluate the contribution of each module by removing or modifying key components. The tested configurations include:

- Without Perceptual Loss: Removal led to degradation in texture and color fidelity, emphasizing its importance in preserving high-level features during style transfer.

- Without Total Variation Loss: This caused increased noise and less smooth transitions, highlighting its role in preserving fine details.

- Without Multi-Scale Discrimination: Replacing it with a single-scale discriminator resulted in blurred details and reduced image quality, especially in textures and stylistic consistency.

The results confirm that each component—Perceptual Loss, Total Variation Loss, Multi-Scale Discrimination, and Residual Connections—significantly improves the quality of generated images, demonstrating their necessity for superior NPR style transfer performance.

V. Model Training and Evaluation

A. Model Training

The training of this model uses the Adam optimizer with a learning rate of 0.0002, a batch size of 16, and 50 epochs. Due to Adam's adaptive learning rate, it effectively addresses sparse gradients, helping the model converge faster in complex deep learning tasks.

During training, perceptual loss and total variation loss were chosen as key loss functions. A validation set is used to evaluate the quality of the generated images in real-time, allowing dynamic adjustments of hyperparameters and network structure based on validation results. The validation set contains images from multiple artistic styles, ensuring the model can adapt to style conversion across different genres. By comparing the generated images with real artwork, the hyperparameters (such as learning rate and batch size) and network architecture are fine-tuned to optimize the model's output. The learning rate of 0.0002 was selected after experimental tuning to promote stable convergence, preventing rapid weight updates that could cause instability. The batch size of 16 was chosen to balance training speed and memory usage, ensuring efficient training under reasonable hardware configurations. Both the learning rate and batch size significantly influence the convergence speed, with larger batch sizes speeding up convergence but potentially destabilizing the training, requiring careful adjustment with other hyperparameters.

B. Model Evaluation

After training, the model underwent multiple rounds of testing to assess its performance in terms of style conversion, detail retention, and generation efficiency.

1) Style Conversion Performance

Style conversion performance is a key indicator of whether the model can effectively transform source images into target artistic styles. The evaluation methods include:

- Visual Similarity Evaluation: Expert evaluations of the generated images are compared with the target style images, assessing their similarity in terms of color distribution, brushstroke effects, and texture reproduction. The goal is to ensure the results are natural and smoothly transition to the intended style.

- Computational Metrics: The Structural Similarity Index (SSIM) is used to quantify the style conversion performance, assessing how well the generated image matches the target style, considering brightness, contrast, and structure.

Results show that the model excels in style conversion across multiple artistic styles. For example, the ink painting style conversion retained the main structure and color layout of the source image, presenting typical brushstroke and ink shading effects. The oil painting style exhibited strong oil painting textures, including thick paint layers and lighting effects, while the sketch style successfully recreated smooth lines and light-dark contrasts (see figure 1).
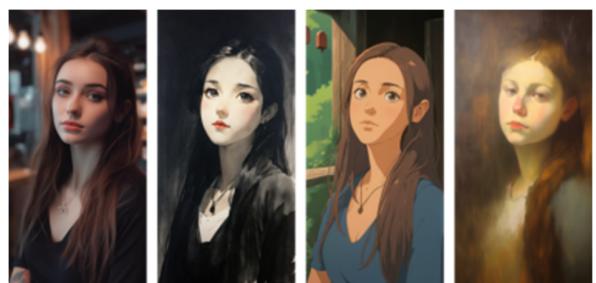

Figure 1. Style Transfer Results

2) Detail Retention

Detail retention is a crucial standard for evaluating image generation. To ensure the generated images preserve source image details during style conversion, this study uses two metrics: SSIM and Peak Signal-to-Noise Ratio (PSNR).

- SSIM: Measures the structural similarity between the generated and real images, focusing on details such as edges, texture, and contrast. It reflects whether important details are preserved during style conversion.

- PSNR: Assesses the quality of the image, where a higher PSNR indicates better image quality and less detail loss.

Experimental results demonstrate excellent detail retention by the model. The average SSIM value was 0.85, with a peak value of 0.92. The average PSNR value was 33.5 dB, indicating low noise levels and high image quality, with excellent detail retention.

3) Generation Efficiency

Generation efficiency refers to the model's speed and computational cost. This study evaluates the model's potential for real-world applications by measuring training time and inference speed.

- Training Time: The time taken to complete training with different batch sizes, learning rate settings, and hardware conditions is recorded to ensure the training efficiency meets practical needs.

- Inference Speed: The time required to generate an image is calculated, evaluating the model's real-time response capability on different hardware platforms.

On an NVIDIA RTX 3080 GPU platform, the model completed 50 epochs of training in 8 hours, with each image generated in 1.2 seconds. This demonstrates that the model can generate high-quality images within a reasonable timeframe, showing strong potential for practical applications.

4) Visual Quality Scoring

In addition to quantitative evaluation, this study also introduced subjective visual quality scoring, combining expert ratings and audience feedback to comprehensively assess the generated images.

- Ink Painting Style Conversion: The average score was 4.6. Experts praised the natural ink layers and diffusion effects, noting that the style conversion was smooth and realistic.

- Oil Painting Style Conversion: The average score was 4.3. Experts highly rated the brushstrokes and paint layer textures, but some images still showed room for improvement in detail presentation.

- Sketch Style Conversion: The average score was 4.5. Experts appreciated the fine detail in the generated images, with smooth lines and natural light-dark contrasts.

The expert scores highlighted the naturalness, artistic perception quality, and style retention of the generated images.

VI. CONCLUSION

Combining both quantitative and qualitative evaluations, the CNN-GAN-based deep learning model proposed in this study performed exceptionally well in style transfer, detail preservation, and generation efficiency. Compared to traditional Non-Photorealistic Rendering methods, the model not only retains high-quality details effectively but also accurately performs style transfer across various artistic styles.

Both style transfer and detail retention reached high levels, and the generation efficiency met real-time demands, especially on mid-to-high-end GPUs, where inference speed and training time were ideal. These results suggest that the model has significant application potential in digital art creation, animation production, game design, and other fields. Despite the model's excellent performance in most tests, there is room for further optimization. For example, incorporating more complex network architectures could further enhance style transfer quality. Additionally, optimization for inference speed and generation efficiency on lower-performance hardware could extend its applicability to a wider range of use cases.

REFERENCES

[1] Zhang, L., Zhang, Y., & Zhang, D. (2020). "Adaptive neural texture synthesis for style transfer." IEEE Transactions on Image Processing, 29, 4512-4524.

[2] Huang, J., Li, L., & He, J. (2021). "Improved CNN-based style transfer using perceptual similarity metrics." Pattern Recognition, 108, 107539.

[3] Chen, W., Liu, Y., & Zhang, X. (2022). "Dual-cycle consistency for detail-preserving style transfer in CycleGAN." IEEE Transactions on Neural Networks and Learning Systems, 33(9), 4236-4249.

[4] Zhang, H., & Liu, Y. (2022). "Fine-tuning CycleGAN for style consistency and texture preservation." Computer Vision and Image Understanding, 214, 103347.

[5] Zhao, J., Liu, X., & Huang, T. (2021). "Improved detail preservation in GAN-based style transfer." Journal of Visual Communication and Image Representation, 75, 103070.

[6] Kim, H., Choi, K., & Lee, J. (2023). "Hybrid GAN models with attention mechanisms for style transfer and detail preservation." IEEE Transactions on Image Processing, 32(1), 184-195.

[7] Zhu, J., Xu, H., & Wang, Y. (2020). "Self-attention-based methods for style transfer." International Journal of Computer Vision, 128(5), 1232-1243.

[8] González, A., & Pérez, P. (2023). "Adversarial Training for Artistic Image Synthesis: A GAN Approach." Journal of Visual Communication and Image Representation, 88, 103365.

[9] Bennett, C., & Robinson, C. (2023). "Non-Photorealistic Rendering for Interactive Graphics using Deep Neural Networks." Computer Graphics Forum, 42(6), 1610-1622.

[10] Boubekeur, T., & Kato, H. (2023). "Artistic style transfer with high-level feature loss for non-photorealistic rendering." Computer Graphics Forum, 42(1), 59-72.

[11] Schulz, M., & Kähler, C. (2023). "Style-preserving GANs for artistic rendering in 3D modeling." Computer Graphics Forum, 42(2), 75-88.