

Art And Technology Integration: CNN-Based Image Style Transfer Research

Haoran Tang

Saint Andrew's School, FL, US

haoran.tang@saintandrews.net

Abstract. This research presents a CNN-based image style transfer framework that integrates layered feature extraction with region-aware stylization. Traditional neural style transfer methods often apply styles globally across an image, blurring boundaries between subjects and backgrounds and reducing fine-grained stylistic control. To address this limitation, the proposed method combines multi-layer feature modeling using the VGG19 network with semantic segmentation from DeepLabv3 to distinguish subject and background regions. By separating feature maps using binary masks, the system computes independent style losses for each region while maintaining the overall content structure via a content loss function. Experimental results on a dataset of diverse images—including portraits, architecture, and landscapes—demonstrate that the framework achieves clearer structural preservation and more balanced stylization than conventional global methods. Parameter studies further show that adjusting the weights of content, subject style, and background style enables flexible artistic control. The approach offers potential applications in digital art creation, cultural heritage visualization, and AI-assisted design.

Keywords: Neural Style Transfer; Convolutional Neural Networks; Semantic Segmentation.

1. Introduction

Against the backdrop of rapid advances in artificial intelligence, computer vision has become a key driver of technological innovation. Image style transfer [1], a cutting-edge topic that merges art and technology, has drawn the attention of many researchers and developers. This task aims to retain the original image structure while applying a specific artistic style, generating images that are both semantically consistent and aesthetically appealing. Since Gatys et al. [2] first proposed using Convolutional Neural Networks (CNNs) [3] for image style transfer in 2015, the field has achieved remarkable progress. However, most mainstream methods still rely on fixed-layer or global features, overlooking the varying style needs at the spatial and semantic levels, especially when it comes to fine-detail control and handling complex backgrounds.

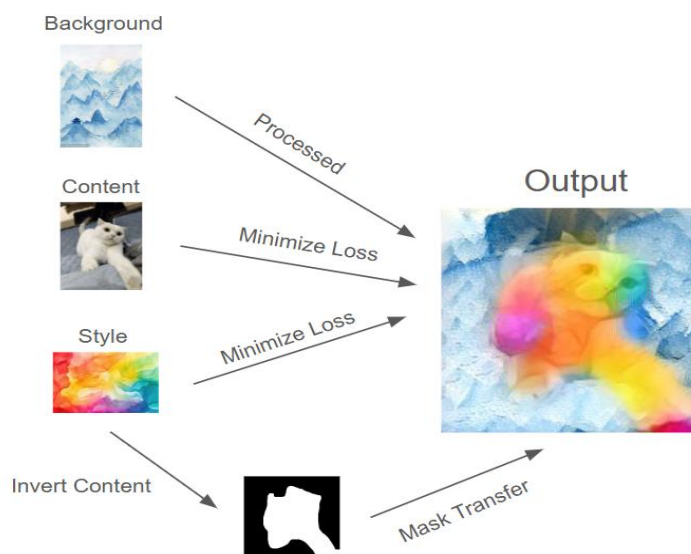


Fig.1 Image Style Transfer Overview

To address these issues, this paper proposes a CNN-based image style transfer approach that combines layered transfer with region recognition. As shown in Fig. 1, it provides a brief overview of how image style transfer works, including the steps and results. This method introduces three innovations: first, based on the VGG19 network [4], it extracts multi-layer features from low to high levels, modeling the image's texture, structure, and semantic hierarchy. During transfer, it calculates style and content losses separately for each layer, ensuring coordination between detail and overall style. Second, it introduces Canny edge detection or the DeepLabv3 model [5] to automatically generate content masks, reinforcing the preservation of primary structures. Especially in images where subjects, such as people or architecture, stand out against complex backgrounds, this significantly improves local style expression. Notably, this project focuses on ink painting styles, using high-resolution Chinese ink paintings as style sources and, in post-processing, combining them with rice paper textures to produce outputs rich in traditional artistic aesthetics, suitable for cultural products, artwork restoration, and creative generation.

Existing studies, such as those by Johnson et al. [6] and Ulyanov et al. [7], primarily focus on real-time performance and image quality, but often rely on shallow layers or single loss functions, thereby lacking fine-grained separation between content and style regions. GAN-based methods, such as CycleGAN [8] and StyleGAN [9], achieve unpaired domain transfer but face instability and limited style controllability.

To bridge the limitations of earlier approaches, this study emphasizes fine-grained, region-aware stylization—a factor often neglected in conventional or real-time models. Previous methods primarily focused on faster processing or improved global aesthetics, but they generally treated the image as a single, uniform domain, resulting in blending between subjects and backgrounds. In contrast, the present work focuses on spatial and semantic differentiation, ensuring that each region follows its own stylistic behavior. By integrating layered feature extraction with regional masking, it establishes a more structured and interpretable framework. Compared to these, our method offers key advantages:

(1) The proposed model uniquely integrates layered transfer with region masking, separating image features across multiple depths to enable targeted stylization and maintain a balanced relationship between structure and artistic detail.

(2) A major innovation lies in the use of DeepLabv3 [5] for precise region identification, allowing the network to distinguish and process subjects and backgrounds independently—a capability rarely emphasized in existing studies.

(3) Its framework demonstrates strong adaptability, functioning effectively across diverse image types such as portraits, landscapes, and architectural scenes, and consistently surpassing conventional single-layer designs in visual coherence.

(4) Beyond technical contributions, the system shows broad potential for application in cultural production, art restoration, digital heritage, and educational contexts, extending its influence well beyond academic exploration.

The introduction explains the project's motivation and research gap; the related work section reviews past CNN-based style transfer studies and their limits. The framework section describes the model design, feature extraction, and loss setup. The experiment section compares this layered method with traditional approaches, demonstrating a more transparent structure and improved style control. The conclusion summarizes the key results and suggests future improvements, such as multi-style fusion and interactive adjustment.

2. Related Work

In recent years, style transfer research has experienced significant growth, particularly with the advent of CNN-based methods. Gatys et al. [2] pioneered the combination of content and style losses via optimization, but this approach proved computationally intensive. Johnson et al. [6] achieved real-time transfer but at the cost of fine detail. Ulyanov et al. [7] improved contrast with instance normalization but lacked spatial modeling. Li and Wand [10] introduced Markov random fields for

style consistency, but texture overflow occurred on complex backgrounds. Chen et al. [11] added semantic segmentation but had weak network coupling. Dumoulin and Shlens [12] explored conditional instance normalization for multi-style learning but lacked region control. Huang and Belongie [13] allowed dynamic adjustments but blurred fine regions. Zhang et al. [14] used attention mechanisms but struggled with multi-style region fusion. Sanakoyeu et al. [15] developed style-aware content loss, but it caused structure distortion. Yao et al. [16] applied dual discriminators but faced instability and overfitting. Li et al. [17] combined CNN feature maps with Markov fields but lacked spatial explicitness. Mechrez et al. [18] improved structure retention but reduced style consistency. Park and Lee [19] integrated style cognition with GANs but relied heavily on the quality of the training data. Kotovenko et al. [20] added adaptive instance normalization with spatial masks but lacked a unified layered framework.

In contrast, our approach uniquely integrates region recognition and layered modeling, enhancing local details, global consistency, cultural fit, and engineering robustness, promising wide applications across cultural, artistic, and cross-modal image tasks.

3. Layered Transfer Algorithm Framework

3.1. Framework Overview

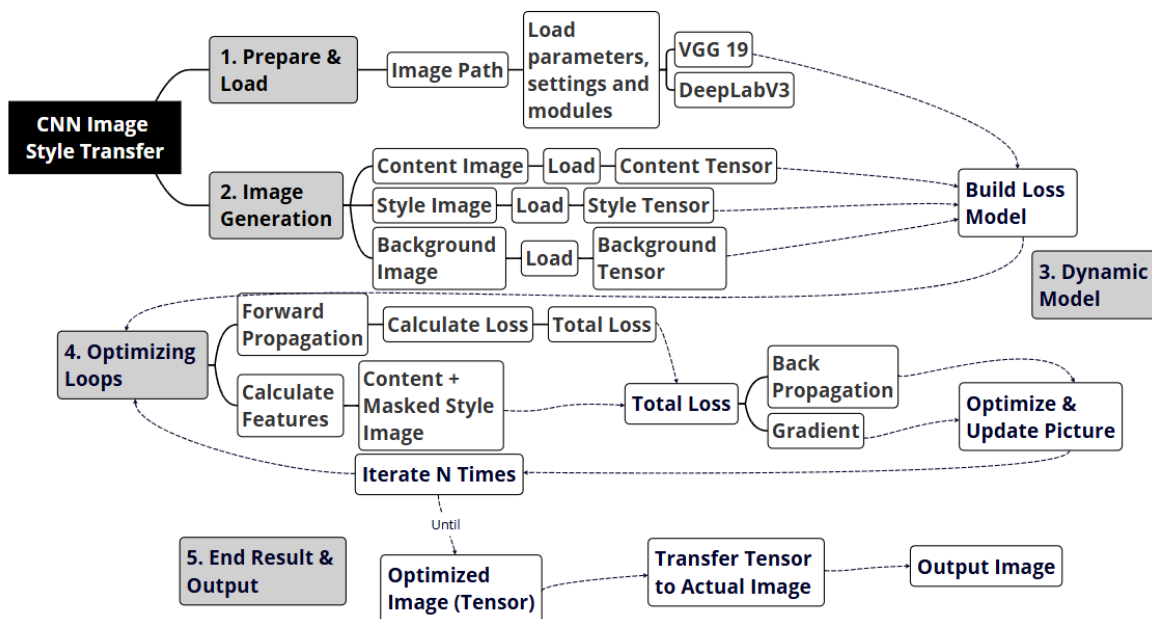


Fig.2 CNN Framework Overview

This section outlines the proposed layered transfer framework, which provides precise control over style transfer by separately handling subject and background regions. As shown in Fig. 2, an overview of the entire image style transfer processing route illustrates how each step works and interacts with the others. Inputs include a content image, a subject-style image (e.g., a painting of a person or object), and a background-style image (e.g., abstract patterns or textures). The grey boxes represent the five main stages of the CNN Image Style Transfer process. In the Prepare & Load stage, the system initializes image paths, imports pretrained networks such as VGG19 and DeepLabV3, and loads necessary parameters, including model weights and optimization settings. In Image Generation, the content, style, and background images are converted into tensors —numerical arrays that the neural network can process efficiently. The Dynamic Model stage constructs the computational framework. It defines the loss functions that guide the evolution of the generated image, combining content, style, and region-based losses for precise control. During Optimizing Loops, the model iteratively computes

losses and backpropagates gradients to update the image, gradually refining its appearance to match the desired style. Finally, in the End Result & Output section, the optimized tensor is converted back into an RGB image to produce the final stylized output. Loading parameters ensures consistent pretrained weights and settings are applied; tensors serve as the numerical representation of images for efficient computation; and the total loss, a weighted sum of content, style, and region-specific losses, directs the optimization process toward producing an image that preserves content structure while applying the chosen artistic style accurately. The goal of CNN Image Style Transfer is to achieve an artistic fusion in which the main subject stands out with refined style while the background complements it harmoniously.

3.2. Subject Content Identification (DeepLabv3)

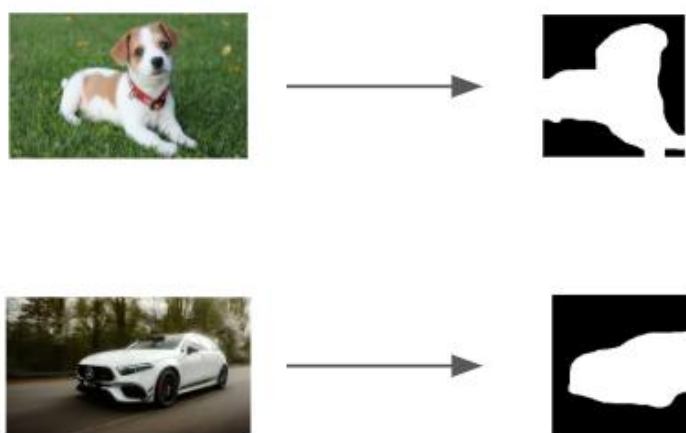


Fig.3 DeepLabV3 Image Mask Generation

The first stage utilizes semantic segmentation with DeepLabv3, a model pretrained on the COCO dataset, to detect various object categories. The segmentation map assigns category IDs to each pixel, with non-background pixels forming a binary subject mask M (where subject = 1 and background = 0). Morphological operations using OpenCV (closing to fill small holes and opening to remove noise) refine the mask, yielding a smooth and clean guide for subsequent feature extraction. As shown in Figure 3, DeepLabV3 produces precise and reliable masks for the subject image. The mask generated provides a reliable path for the image style transfer to continue functioning.

3.3. Style Feature Extraction (VGG19)

VGG19, a deep convolutional neural network, is employed to extract hierarchical image features. Low-level layers capture edges and textures; mid-level layers capture patterns and shapes; deep layers capture semantic information. Each layer outputs feature maps F^1 with dimensions $(C \times H \times W)$, where C is the number of channels, H is the height, and W is the width. The feature map's role is to provide a rich representation for computing both content and style losses.

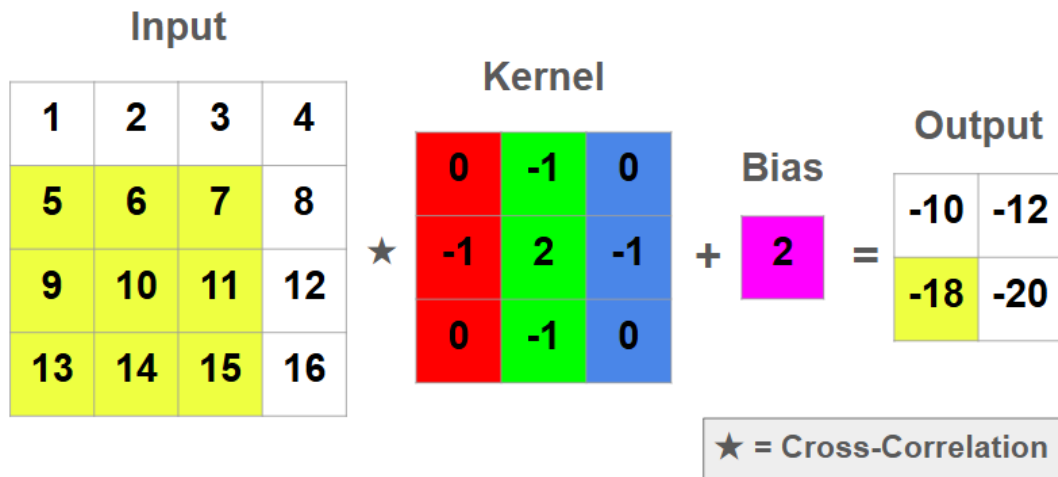


Fig.4 Convolutional Layer Overview

As shown in Fig. 4, the core operation behind convolutional neural networks is convolution, or more precisely, cross-correlation. A small 3x3 filter, known as a kernel, slides across a larger input matrix, such as a grayscale image, performing element-wise multiplication between the kernel and each overlapping 3x3 region of the input. The results are summed, and a bias value is added to produce a single number in the output matrix, referred to as the feature map. In this example, the kernel enhances the center pixel while reducing the influence of surrounding pixels, effectively highlighting essential patterns. This process continues across the entire input, generating a smaller output that captures local visual features. The resulting 2x2 output matrix represents transformed information that helps identify textures, edges, and structures, forming the foundation for deeper layers in networks like VGG19.

3.4. Loss Functions

3.4.1 Content Loss Function

Content loss ensures the generated image x retains the high-level structure of the original content image p . It is calculated as the normalized mean square error (MSE) between feature maps:

$$L_{\text{content}}(\vec{p}, \vec{x}, l) = \left(\frac{1}{N_l D_l} \right) \times \sum_{ij} \left(F_{ij}^l(\vec{x}) - P_{ij}^l(\vec{p}) \right)^2$$

The content loss function measures how well the generated image \vec{x} preserves the high-level content structure of the original content image \vec{p} . It does this by comparing their feature representations extracted from a specific convolutional layer l in a pretrained neural network, such as VGG19. Specifically, the function looks at the activations (feature maps) at that layer: $F_{ij}^l(\vec{x})$ represents the activation at position (i,j) in layer l for the generated image, while $P_{ij}^l(\vec{p})$ represents the activation at the same position for the content image. Here, i indexes over the number of feature channels (filters), and j indexes over the spatial positions (height and width combined). The variables N_l and D_l refer to the number of channels and the spatial size (height times width) of the feature maps at layer l , respectively.

3.4.2 Style Loss Function

$$G_{ij}^l(\vec{x}) = \sum_k F_{ik}^l(\vec{x}) F_{jk}^l(\vec{x})$$

The style loss in neural style transfer relies on the Gram matrix, which captures an image's style or texture by measuring inter-channel correlations within its feature maps. Specifically, for a given layer l , the Gram matrix $G_{ij}^l(\vec{x})$ for the generated image \vec{x} is computed by taking the sum over k of the product $F_{ik}^l(\vec{x})F_{jk}^l(\vec{x})$, where $F_{ik}^l(\vec{x})$ represents the activation of the i -th feature channel at the k -th spatial position in layer l . Here, i and j index over the feature channels (or filters), while k indexes over the spatial positions, which combine height and width.

In addition, the Gram matrix calculates how strongly each pair of feature channels i and j are correlated across all spatial locations k . By comparing the Gram matrices of the generated image and the style image, the style loss evaluates how well the generated image replicates the texture and style patterns — such as colors, brush strokes, or repeating motifs — that are independent of the specific spatial arrangement.

3.4.3 Masked Style Loss Functions

$$L_{ss(\vec{x}, \vec{s}_s, M)} = \sum_l w_l \cdot \text{MSE}(\text{Gram}(F_{\text{subject}}^l(\vec{x})), \text{Gram}(F^l(\vec{s}_s)) \cdot r_1)$$

$$L_{bs(\vec{x}, \vec{s}_b, M)} = \sum_l w_l \cdot \text{MSE}(\text{Gram}(F_{\text{background}}^l(\vec{x})), \text{Gram}(F^l(\vec{s}_b)) \cdot (1 - r_1))$$

The subject style loss L_{ss} and the background style loss L_{bs} are designed to separately measure how well the generated image \vec{x} matches the style of the subject and background regions, respectively. These losses are computed as weighted sums over selected layers l in the neural network, where each layer can contribute differently via a weight w_l .

For the subject style loss $L_{ss(\vec{x}, \vec{s}_s, M)}$, the feature maps of the subject region in the generated image, denoted $F_{\text{subject}}^l(\vec{x})$, are compared to the feature maps of the subject style image \vec{s}_s . This comparison is based on their Gram matrices, which capture inter-channel correlations, and the difference between these matrices is measured using the mean squared error (MSE). The result is then weighted by a region mask r_1 , which ensures that only the subject region contributes to the loss.

Similarly, the background style loss $L_{bs(\vec{x}, \vec{s}_b, M)}$ compares the background regions. $F_{\text{background}}^l(\vec{x})$ represents the background feature maps of the generated image, and $F^l(\vec{s}_b)$ comes from the background style image \vec{s}_b .

3.4.4 Feature Separation Functions

We downsample the mask to match the dimensions of each VGG layer's feature map using interpolation, then split the features into subject:

$$F_{\text{subject}}^l(\vec{x}) = F^l(\vec{x}) \odot M_l$$

And background components:

$$F_{\text{background}}^l(\vec{x}) = F^l(\vec{x}) \odot (1 - M_l)$$

The two equations define how the feature maps of the generated image \vec{x} are separated into subject and background regions at a given layer l . Specifically, $F_{\text{subject}}^l(\vec{x})$ is calculated by taking the element-wise (Hadamard) product \odot between the full feature map $F^l(\vec{x})$ and a binary mask M_l , where M_l has values of 1 for the subject region and 0 elsewhere. This effectively keeps only the activations corresponding to the subject, setting the background parts to zero.

Similarly, $F_{\text{background}}^l(\vec{x})$ is calculated by multiplying $F^l(\vec{x})$ by $(1 - M_l)$, which flips the mask, selecting the background region instead. $(1 - M_l)$ equals 1 in the background areas and 0 in the

subject areas, so this operation preserves the background activations while zeroing out the subject part.

3.4.5 Total Loss Function

A normalization technique compensates for the reduced feature space caused by masking, using the ratio rl of subject pixels, ensuring fair loss weighting. The total loss function is then defined as:

$$L_{total} = \alpha \cdot L_{content} + \beta_s \cdot L_{ss} + \beta_b \cdot L_{bs}$$

The total loss L_{total} combines all the key components needed to guide the style transfer process. Specifically, it is calculated as a weighted sum of the content loss $L_{content}$, the subject style loss L_{ss} , and the background style loss L_{bs} . The parameter α controls the relative importance of the content loss, ensuring that the generated image retains the high-level structure and semantic content of the original content image. The parameter β_s weights the subject style loss, which governs how well the style of the subject in the generated image matches the subject style image. Similarly, β_b weights the background style loss, focusing on how well the background style is transferred. By tuning these weights (α , β_s , and β_b), one can adjust the balance between preserving content, applying subject style, and applying background style, thus achieving flexible and controllable artistic results in the final image.

4. Experimental Validation and Analysis

4.1. Experiment Design

The experimental setup includes a dataset of approximately 50 diverse content images, including portraits, street views, and natural landscapes. Subject style sources are drawn from figurative art such as oil portraits and detailed sketches, while background style sources are drawn from abstract textures, color gradients, and minimalist patterns. To explore the behavior of the system, we systematically vary the hyperparameters α , β_s , and β_b , observing how they influence the balance between preserving content structure and imposing stylistic features.

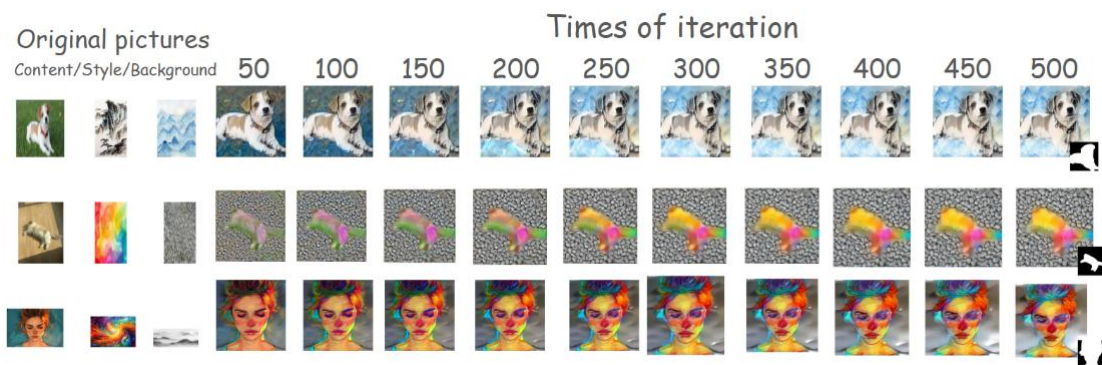


Fig.5 Different subject / Different style experiment group #1



Fig.6 Different subject / Different style experiment group #2

Figures 5 and 6 (see Different Subject / Different Style Experiment Groups #1 and #2) illustrate the results under multiple combinations of content, style, and background images. Each row represents a distinct experimental configuration, with iterative outputs from 50 to 500 steps. The progression shows how the layered transfer model gradually enriches the image with stylistic elements while maintaining content integrity. At early iterations (50–150), the generated images retain most structural details, with only subtle stylistic traces emerging. By iteration 200, distinctive colors and textures appear, while the content remains recognizable. Between iterations 300–400, style integration intensifies, achieving harmony between subject and background, and by iteration 500, a clear artistic coherence is reached—the subject is well-defined, and the background contributes without distraction (see Fig.5–6).

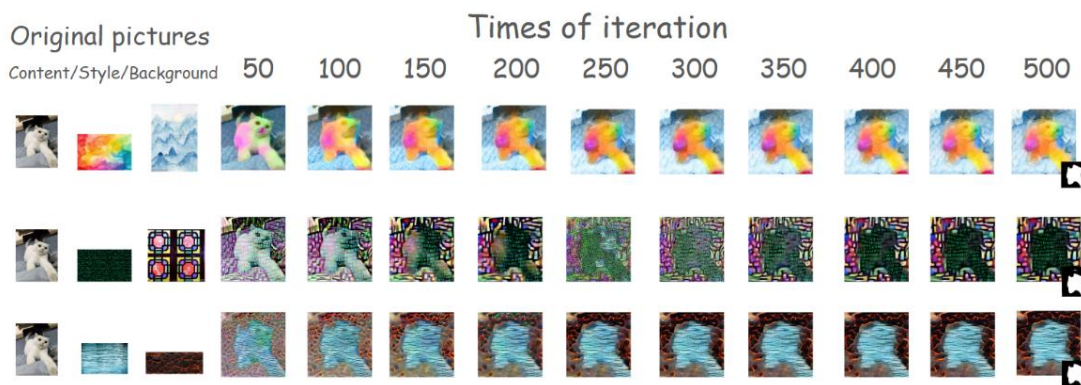


Fig.7 Same subject / Different style experiment group #1

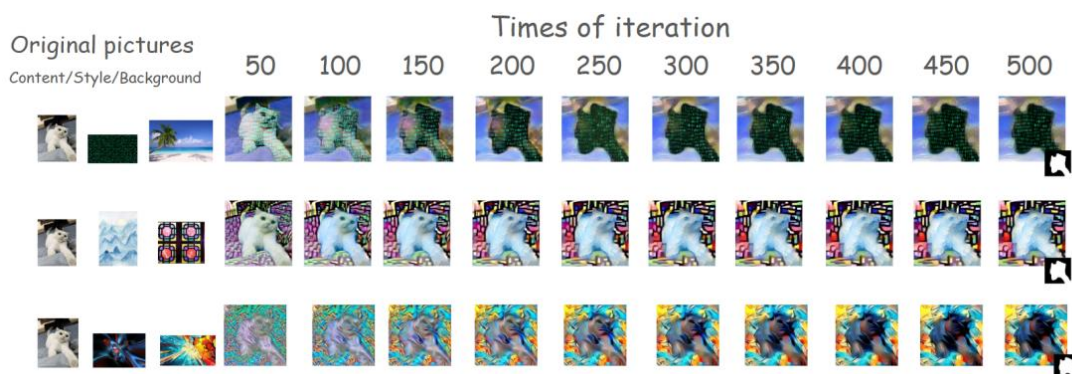


Fig.8 Same subject / Different style experiment group #2

Figures 7 and 8 (see Same Subject / Different Style Experiment Groups #1 and #2) further validate the adaptability of the proposed framework by fixing the same content image while varying subject and background styles. Across iterations from 50 to 500, the subject remains semantically consistent, while distinct styles are applied successfully. For instance, colorful abstract, geometric, and wave-like textures produce entirely different aesthetics, yet the subject structure and boundaries remain stable. These results demonstrate that the layered transfer method achieves controllable stylization and effectively separates subject and background features, outperforming uniform global methods in both precision and expressiveness (see Fig.7–8).

4.2. Comparison of the results

Experiment 1: Effect of Subject and Background Style Weights on Stylization Balance

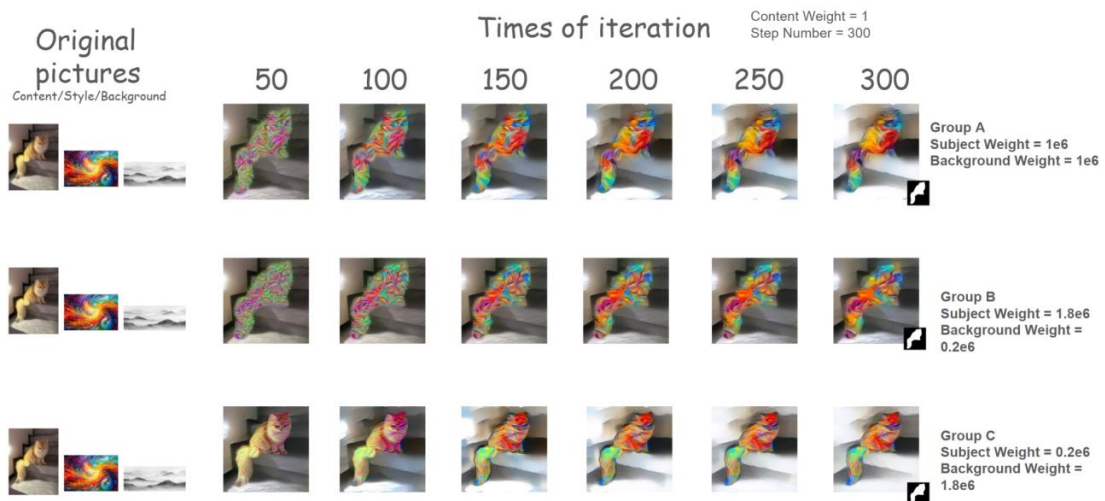


Fig.9 Different Subject / Background Weight experiment group



Fig.10 Iteration Step vs Subject Loss graph

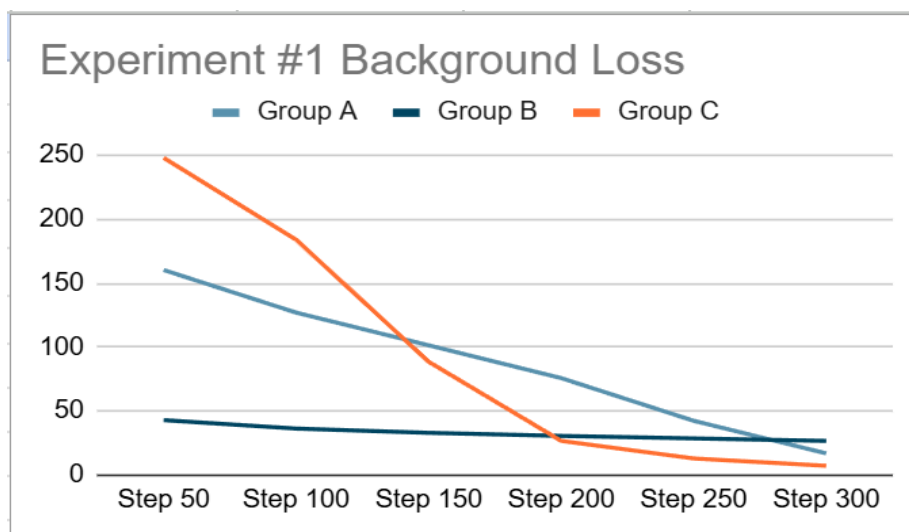


Fig.11 Iteration Step vs Background Loss graph

This experiment investigates how the balance between the subject style weight (β_s) and background style weight (β_b) affects the separation and harmony of styles within the final image. When one weight is significantly more substantial than the other, its region becomes dominant. It may cause style blending across boundaries, whereas equal weights help maintain clear regional distinctions and visual balance.

To ensure consistent testing, the content weight (α) was fixed at 1, with 300 training iterations. Three groups were compared: balanced weights ($\beta_s = \beta_b = 1.0e6$), enhanced subject ($\beta_s = 1.8e6, \beta_b = 0.2e6$), and enhanced background ($\beta_s = 0.2e6, \beta_b = 1.8e6$). As shown in Figure 9, balanced weights achieved even stylization across both regions. A higher subject weight made the subject stand out with stronger color and texture, but softened the background. In contrast, a higher background weight emphasized environmental details, occasionally blending into the subject and reducing clarity.

Figures 10 and 11 illustrate that the loss curves for each group converge differently, depending on the weight ratios. The subject loss dropped faster when β_s was larger, showing stronger stylization in the subject, whereas the background loss decreased most when β_b was higher, confirming stronger background adaptation. These results highlight that balancing β_s and β_b is essential for maintaining structural separation and visual harmony.

Additional ablation studies reinforce these findings. When semantic segmentation using DeepLabv3 was applied, the system maintained sharper subject edges and prevented background textures from bleeding into the subject area, significantly improving overall coherence. Similarly, comparing single-style and dual-style inputs revealed that dual-style configurations yielded richer artistic layering and improved spatial depth. Finally, parameter sensitivity analysis confirmed that proper tuning of α , β_s , and β_b yields the most stable and aesthetically pleasing images. Together, these experiments demonstrate that both segmentation and balanced parameter design are key to achieving clear structure, realistic focus, and natural artistic composition in CNN-based style transfer.

Experiment 2: Trade-off Between Content Weight and Stylization Strength

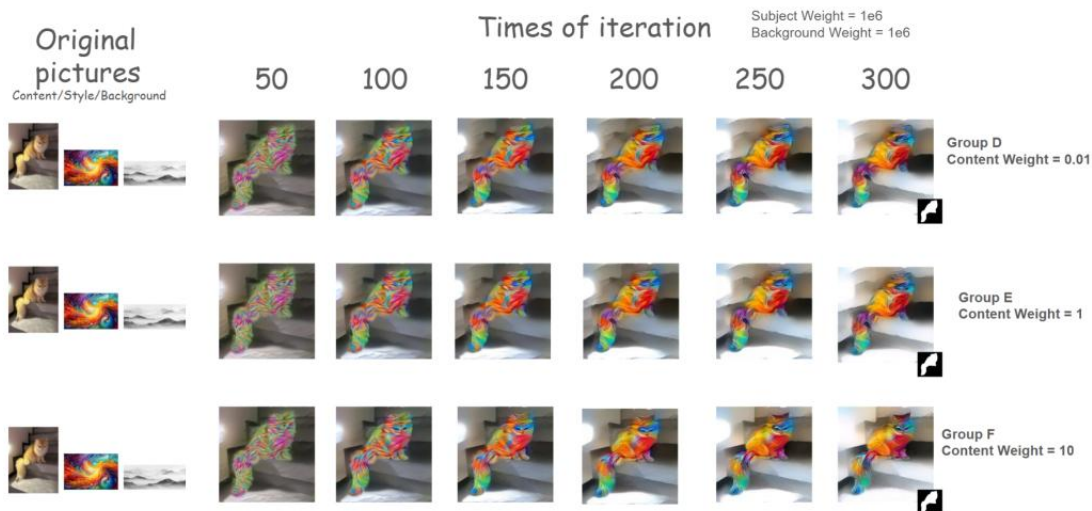


Fig.12 Different Content Weight experiment group

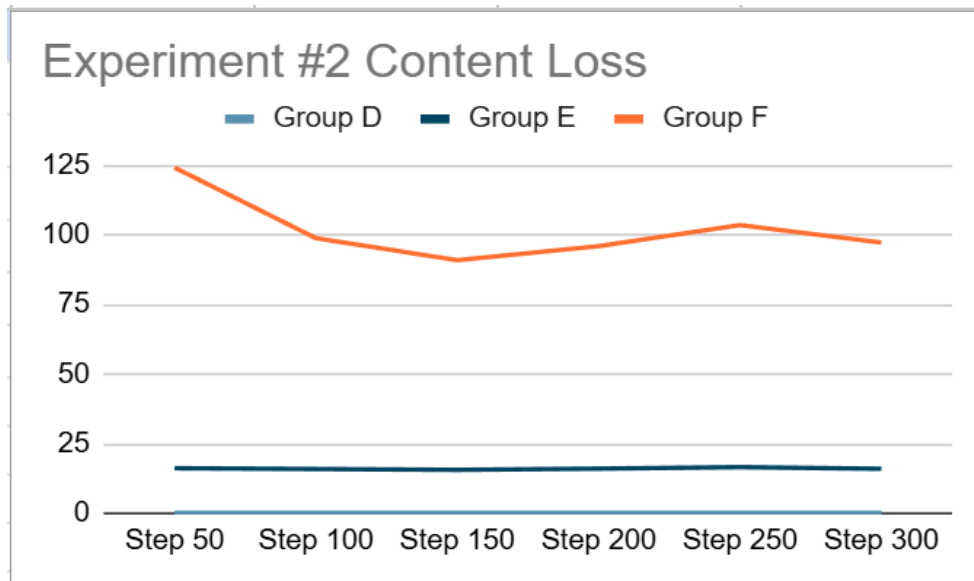


Fig.13 Iteration Step vs Content Loss graph

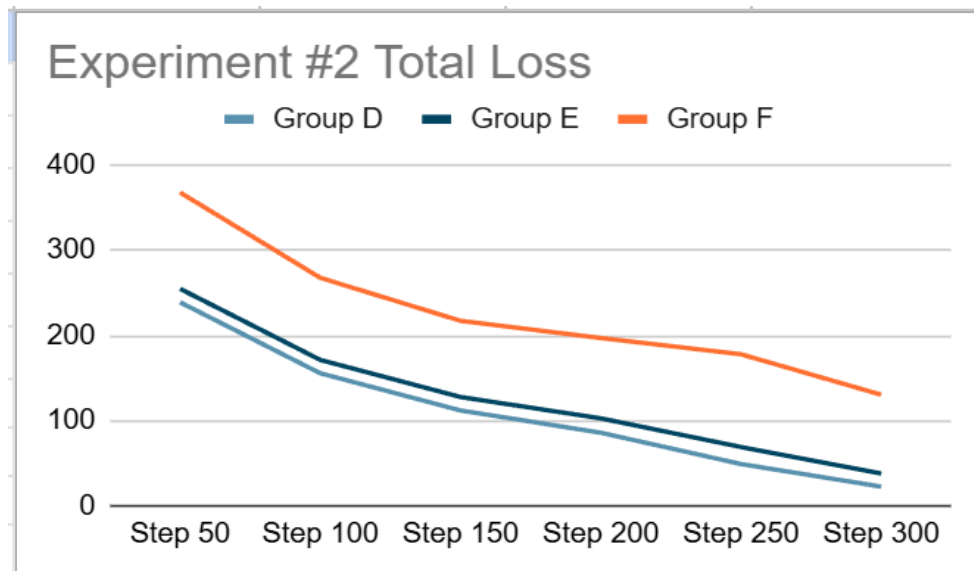


Fig.14 Iteration Step vs Total Loss graph

This experiment focuses on how the content weight (α) influences the relationship between structural preservation and artistic stylization. The general idea is that increasing α makes the generated image retain more of the original content, but reduces the strength of the style. Conversely, decreasing α produces more creative and expressive results, but may distort shapes and details. To maintain consistency, the subject and background style weights were kept constant at $\beta_s = \beta_b = 1.0e6$, and three settings were tested: weak content weight ($\alpha = 0.1$), balanced content weight ($\alpha = 1.0$), and substantial content weight ($\alpha = 10.0$).

As shown in Figure 12, Group D ($\alpha = 0.1$) yielded highly expressive and freeform stylization, characterized by vivid textures and artistic color transitions, although the overall structure was partially lost. Group F ($\alpha = 10.0$) produced nearly realistic images, maintaining most of the original structure but with minimal artistic transformation. Group E ($\alpha = 1.0$) achieved the most visually balanced output, where both content and style were clearly visible, forming a smooth blend of realism and creativity.

Figures 13 and 14 provide further evidence. The content loss curve demonstrates that Group F converged to the lowest values, confirming stronger content retention, while Group D remained the highest due to weaker structural constraints. Meanwhile, the total loss curve exhibits an inverse trend: Group D achieved lower style losses more quickly, reflecting a higher stylization intensity, whereas

Group F's larger α prevented complete stylistic adaptation. These results highlight a clear trade-off between detail preservation and stylistic freedom, emphasizing that optimal visual quality lies in the middle range of α .

The ablation studies reinforce these observations. When semantic segmentation (DeepLabv3) is enabled, the model achieves more precise boundaries and more accurate style separation between subject and background, preventing texture overlap and improving overall composition. Comparing single-style and dual-style setups reveals that dual-style inputs offer greater visual depth and contrast, with subjects exhibiting fine artistic detail and backgrounds providing softer, complementary textures. Finally, adjusting α , β_s , and β_b demonstrates that balanced parameters—especially when β_s is approximately equal to β_b —produce the most stable, harmonious images with realistic structure and consistent style blending.

5. Conclusions and directions for further study

5.1. Core Contributions

This study proposes a layered style transfer framework designed to achieve region-aware, controllable artistic stylization across diverse visual domains. By integrating semantic segmentation with multi-layer feature extraction, the model effectively addresses one of the key limitations of traditional CNN-based methods—the lack of coordination between subject and background styles. Through independent style loss computation and region-specific masking, the system enables balanced stylization where subjects retain clarity and visual prominence, while backgrounds complement them with coherent texture and tone. Furthermore, the framework supports customizable dual-style fusion, enabling users to combine artistic influences from different sources selectively. This not only enhances creative flexibility but also broadens the applicability of neural style transfer in cultural art restoration, digital heritage visualization, and AI-assisted creative design.

5.2. Future prospects

Future research will explore several directions to enhance both artistic expressiveness and structural intelligence. One key area involves advancing style representation beyond traditional Gram-matrix correlations to include higher-order texture and brushstroke modeling, enabling the system to capture more nuanced visual patterns reflective of specific artistic schools. Another promising direction lies in composition-aware optimization, incorporating principles from art theory—such as the rule of thirds, symmetry, and visual flow—into the loss design to guide aesthetic balance automatically. Additionally, future models may integrate attention mechanisms or user-interactive modules that allow real-time control over style intensity and spatial influence, bridging human artistic intuition with computational creativity. These developments aim to transform the proposed framework into a more intelligent, interpretable, and artistically autonomous system for the next generation of AI-driven image creation.

References

- [1] Image Style Transfer, the earliest formalization inspired by Gatys et al. (2015).
- [2] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A Neural Algorithm of Artistic Style. arXiv:1508.06576.
- [3] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition (LeNet). Proceedings of the IEEE.
- [4] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG). arXiv:1409.1556.
- [5] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLabv3: Semantic Image Segmentation. arXiv:1706.05587.

- [6] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. ECCV.
- [7] Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2016). Instance Normalization: The Missing Ingredient for Fast Stylization. arXiv:1607.08022.
- [8] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ICCV.
- [9] Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for GANs. CVPR.
- [10] Li, C., & Wand, M. (2016). Combining Markov Random Fields and CNNs for Image Synthesis. CVPR.
- [11] Chen, Y., et al. (2017). Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks. arXiv.
- [12] Dumoulin, V., Shlens, J., & Kudlur, M. (2017). Learned Representations for Artistic Style. ICLR.
- [13] Huang, X., & Belongie, S. (2017). Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. ICCV.
- [14] Zhang, H., Goodfellow, I., et al. (2019). Self-Attention Generative Adversarial Networks. ICML.
- [15] Sanakoyeu, A., Kotovenko, D., et al. (2018). A Style-Aware Content Loss for Real-Time HD Style Transfer. ECCV.
- [16] Yao, A., et al. (2018). Dual Discriminator GANs. NeurIPS.
- [17] Li, C., & Wand, M. (2016). Precomputed Real-Time Texture Synthesis with Markovian GANs. ECCV.
- [18] Mechrez, R., Talmi, I., & Zelnik-Manor, L. (2018). The Contextual Loss for Image Transformation with Non-Aligned Data. ECCV.
- [19] Park, J., & Lee, K. M. (2019). Style-Aware GANs for Realistic Image Synthesis. CVPR.
- [20] Kotovenko, D., Sanakoyeu, A., et al. (2019). Content and Style Disentanglement for Artistic Style Transfer. ICCV.