International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Neural Style Transfer for Artistic Image Generation Using CNN and Deep Learning

Mr. R. Rajesh[1], A. Murari[2], J. Rajesh Reddy[3], M. Sai Kumar [4]

[1] Assistant Professor of Department of CSE(AI & ML),

[*2,3,4]Students of Department of CSE(AI & ML), ACE Engineering College, An Autonomous Institution, Ghatkesar, Hyderabad, Telangana, India

1. ABSTRACT

Neural Style Transfer (NST) is a deep learning-based technique that generates artistic images by blending the content of one image with the style of another. This project implements NST using a pre-trained VGG-19 convolutional neural network, where the network extracts hierarchical feature representations of images. A combined loss function, consisting of content loss and style loss, is optimized through iterative updates to produce a stylized image that maintains the semantic structure of the content while reflecting the artistic characteristics of the chosen style. The proposed system demonstrates applications in digital art, photo editing, film production, and personalized creativity tools. Furthermore, the project highlights the potential of extending this technique to real-time video style transfer, mobile applications, and multistyle generation, making neural style transfer a powerful tool in the intersection of art and artificial intelligence.

This project not only bridges the gap between technology and art but also empowers non-artists to create visually appealing designs with ease. The experimental results confirm that NST can successfully generalize across diverse image domains and artistic styles.

2. INTRODUCTION

Neural Style Transfer (NST) is one of the most fascinating advancements in deep learning, enabling machines to merge the artistic style of one image with the structural content of another. With the evolution of deep learning and convolutional neural networks, computers can now understand textures, colours, and patterns from famous artworks and apply them creatively to ordinary photographs. This transformation is possible due to the hierarchical feature-learning capability of CNNs, which extract deep visual representations from images. NST has become an influential technique in digital art, media, entertainment, and visual design, offering an automated way to produce aesthetically pleasing artwork without human expertise. The ability to generate unique, stylized images has created new opportunities in both academic research and the creative industry, showcasing how AI can be used for artistic expression.

The motivation for developing NST-based systems stems from the need for automated, flexible, and high-quality artistic generation tools that do not require professional artistic skills. Today's digital environment demands fast and visually appealing content for media production, marketing, and entertainment. NST provides an efficient way to generate such content by intelligently combining artistic attributes through deep learning. It not only democratizes art creation for everyday users but also enhances creative workflows for designers, filmmakers, and visual storytellers. The growing popularity of AI-driven applications—including mobile filters, AR effects, and AI-powered editing tools—further strengthens the relevance of Neural Style Transfer in real-world usage.

Despite its strengths, the process of style transfer presents challenges such as computational cost, optimization time, and the need to balance content preservation with stylistic transformation. These challenges encourage ongoing research in improving speed, output quality, and flexibility. With advances in CNN architectures, GPU acceleration, and feature-learning techniques, the performance and applicability of NST continue to evolve, enabling more sophisticated and creative outputs.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

3. LITERATURE REVIEW:

ABOUT THE PROJECT

This project focuses on implementing Neural Style Transfer (NST) for generating artistic images by combining the structural content of one image with the artistic patterns of another using deep learning techniques. It utilizes convolutional neural networks—particularly a pre-trained VGG-19 model—to extract multi-level feature representations that separately encode content information and stylistic textures. By computing a content loss and a style loss from these feature maps and optimizing a generated image to minimize both, the system produces visually compelling outputs that reflect the artistic style while preserving the original content. The project explores essential components such as feature extraction, Gram matrix computation, loss optimization, and image reconstruction. It also highlights key improvements introduced in modern NST research, such as instance normalization and perceptual losses, to enhance quality and efficiency.

[1] Title: Neural Algorithm of Artistic Style (2015)

Authors: Leon A. Gatys, Alexander S. Ecker, Matthias Bethge

The Neural Algorithm of Artistic Style by Gatys, Ecker, and Bethge (2015) introduced the foundational concept of Neural Style Transfer (NST), demonstrating how convolutional neural networks can separately represent image content and artistic style. The method uses a pre-trained VGG-19 model to extract hierarchical features from both content and style images. Content information is taken from deeper layers that capture structure, while style information is extracted from Gram matrices computed across shallower convolutional layers that encode textures and patterns. By defining a combined loss function—content loss plus style loss—and optimizing the pixels of a generated image, the method produces a new image that blends structure from one image with the artistic characteristics of another. The work revolutionized computational creativity by showing that CNNs inherently learn semantic and texture-based representations that can be recombined. Although the approach yields high-quality artistic results, it relies on iterative optimization, making it extremely slow and computationally expensive. This work laid the foundation for future improvements toward faster, more flexible, and more controllable style transfer techniques.

[2] Title: Perceptual Losses for Real-Time Style Transfer (2016)

Authors: Justin Johnson, Alexandre Alahi, Li Fei-Fei

Johnson, Alahi, and Fei-Fei (2016) addressed the major limitation of Gatys' optimization-based NST by introducing a feed-forward style transfer network trained using perceptual losses. Instead of optimizing output pixels for every input image, they trained a transformation network offline to generate stylized images in a single forward pass. The key innovation was the use of perceptual loss functions derived from feature maps of a pre-trained VGG network, which ensured that the output preserved high-level content while accurately capturing stylistic patterns. This approach enabled near real-time performance even on consumer GPUs, making NST practical for videos and interactive applications.While the quality was comparable to the original optimization method, the main drawback was the need to train a separate model for each style, limiting flexibility. Nevertheless, this work significantly advanced computational efficiency in NST and became a major stepping stone for subsequent fast and arbitrary style transfer methods.

[3] Title: Instance Normalization and Texture Networks (2016)

Authors: Dmitry Ulyanov, Andrea Vedaldi, Victor Lempitsky

Ulyanov, Vedaldi, and Lempitsky(2016) proposed key improvements to fast style transfer by introducing Instance Normalization (IN), which proved far more effective for stylization than Batch Normalization. They observed that style transfer benefits from normalizing feature statistics on a per instance basis, as this enhances consistency in texture patterns across images. Their work on texture networks demonstrated that feed-forward generative networks can synthesize high-quality textures and stylized outputs efficiently by training with appropriate perceptual and texture losses. Instance Normalization became a crucial component in almost all modern style transfer architectures, as it helps decouple content and style representations while maintaining stable training. The study highlighted how subtle architectural changes, such as replacing batch statistics with per-image normalization, can significantly improve style consistency and reduce artifacts. Overall, this research provided a strong theoretical and empirical foundation for improving visual quality in real-time style transfer systems.

[4] Title: Universal Style Transfer via Whitening and Colouring Transform (2017)

Authors: Yongcheng Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, Ming-Hsuan Yang

Li et al. (2017) introduced a major breakthrough in NST by proposing a universal, training-free method capable of transferring arbitrary artistic styles without needing separate models. Their technique applies Whitening and Colouring Transforms (WCT) to align the statistical distributions of deep feature maps from the content image to those of the style image. By using an encoder– decoder architecture based on VGG, the system modifies feature covariance structures to embed style characteristics while preserving semantic content. This was one of the first approaches to make arbitrary style transfer fast and generalizable across unlimited styles. The decoder reconstructs the stylized output without iterative optimization, making it efficient and flexible. Although WCT may produce slightly less detailed styles compared to optimization-based approaches, its ability to support any style image without retraining greatly expanded the applicability of NST. The method laid the groundwork for meta-style learning and adaptive normalization techniques in later research.

[5] Title: Attention-Aware Multi-Stroke Style Transfer (2019)

Authors: Kai Zhang, Yihao Liu, Shuhang Gu, Wangmeng Zuo, Lei Zhang

Zhang and colleagues (2019) extended NST by adding semantic awareness and multi-stroke capabilities through self-attention mechanisms. Their approach identifies salient regions of the content image and applies style transformations more intelligently by focusing on important structural areas. The multi-stroke feature allows the system to generate stylization results with varying stroke sizes, adding greater artistic diversity and finer control over output appearance. By integrating self-attention modules, the network becomes better at preserving important content regions such as faces and objects, while still applying detailed texture patterns from the style image. This work significantly improved local style control, enabling users to achieve different artistic effects within a single stylized image. The method supports region-adaptive stylization, smoother texture transitions, and enhanced detail reconstruction. It represents a key advancement in making NST more controllable, expressive, and semantically aligned with human artistic intent.

[6] Title: Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization (2017)

Authors: Xun Huang, Serge Belongie

Huang and Belongie (2017) proposed Adaptive Instance Normalization (AdaIN), a breakthrough method that enables arbitrary style transfer in real time using a single trained network. Unlike earlier fast style transfer approaches that required training a separate model for each artistic style, AdaIN dynamically adjusts the statistics of the content image features to match those of the style image. Specifically, it aligns the mean and variance of the content feature maps with those of the style features extracted from a pre-trained VGG encoder, effectively embedding stylistic information without retraining. The architecture consists of an encoder (fixed VGG network), the AdaIN transformation layer, and a decoder trained to reconstruct the stylized image. This design allows the system to generalize to unseen styles during inference, making it highly flexible and efficient. The method achieves a strong balance between speed, quality, and generalization, producing visually appealing results comparable to optimization-based NST while operating in real time. Additionally, AdaIN allows user control over style intensity through interpolation between content and style features. This work significantly influenced later research by demonstrating that feature-statistics alignment is sufficient for high-quality stylization, paving the way for adaptive normalization techniques widely used in modern generative models.

S. No.	Author(s)	Title	Proposed Methodology	Findings from the Reference Paper
1	Gatys et al., 2015	Neural Algorithm of Artistic Style	Used VGG-19 features and Gram matrices to optimize an image combining content and style.	Demonstrated that CNNs can recombine semantic content with artistic textures, producing high-quality results but with high computational cost.
2	Johnson et al., 2016	Perceptual Losses for Real-Time Style Transfer	Trained a feed-forward network using perceptual loss for fast stylization.	Achieved real-time style transfer with quality comparable to optimization-based methods.
3	Ulyanov et al., 2016	Instance Normalization and Texture Networks	Applied Instance Normalization instead of Batch Normalization in stylization networks.	Applied Instance Normalization instead of Batch Normalization in stylization networks.
4	Li et al., 2017	Universal Style Transfer via Whitening and Colouring Transform	Used Whitening and Coloring Transform to match content and style feature statistics.	Enabled universal style transfer without retraining for new styles.
5	Zhang et al., 2019	Attention-Aware Multi-Stroke Style Transfer	Integrated self-attention and multi-stroke modules for region-aware stylization.	Enhanced content preservation and allowed controllable artistic effects.
6	Huang & Belongie, 2017	Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization	Used Adaptive Instance Normalization to align content and style features dynamically.	Allowed a single model to generalize to arbitrary styles efficiently in real time.

4. RESEARCH GAPS IN EXISTING SYSTEMS

4.1 Optimization Limitations

Many early Neural Style Transfer methods rely on iterative optimization to generate stylized images, where the output image is repeatedly updated to minimize content and style loss. While this produces high-quality artistic results, the process is computationally expensive and slow, making it unsuitable for real-time or large-scale applications. Each new image requires a fresh optimization cycle, leading to high processing time and resource consumption. This limitation restricts practical deployment in mobile devices, video stylization, and interactive systems.

4.2 Lack of Real-Time Flexibility

Feed-forward approaches introduced to overcome optimization delays can perform stylization quickly, but most of them require training a separate neural network for each artistic style. This creates scalability issues, as adding a new style demands retraining and storing additional models. Such dependency increases storage requirements and reduces adaptability in dynamic environments where users may want to apply multiple or unseen styles instantly.

4.3 Limited Style Generalization

Although later techniques like Adaptive Instance Normalization and Whitening–Coloring Transform support arbitrary styles, they sometimes fail to generalize well to complex artistic patterns or highly detailed textures. These models may oversimplify style characteristics, resulting in loss of fine artistic details or inconsistent texture reproduction. Achieving both strong generalization and high-fidelity stylization remains an open challenge.

4.4 Inadequate Content Preservation and Semantic Awareness

Several NST methods treat all regions of an image equally, without understanding the semantic importance of objects such as faces, buildings, or backgrounds. This can lead to distortion of important structures when style patterns are applied uniformly. Although attention-based models attempt to address this, precise region-aware stylization is still limited, indicating the need for better integration of semantic understanding into style transfer systems.

4.5 Trade-off Between Speed and Visual Quality

A persistent challenge across existing approaches is balancing computational efficiency with artistic quality. Optimization-based models deliver rich, detailed outputs but are slow, whereas fast feed-forward models sacrifice stylistic depth for speed. This trade-off prevents current systems from achieving both real-time performance and high-quality stylization simultaneously, highlighting the need for more efficient architectures that maintain both accuracy and speed.

5. Background and Fundamentals

5.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks form the core foundation of Neural Style Transfer, as they are capable of extracting hierarchical image features such as edges, textures, shapes, and semantic structures. In NST, pre-trained CNNs like VGG-19 are not used for classification but as feature extractors that encode different levels of image representation. Lower layers capture fine textures while deeper layers represent high-level content, enabling separation of style and structure. This property makes CNNs highly suitable for artistic image synthesis tasks.

5.2 Feature Representation in Deep Networks

NST relies on the concept that deep neural networks learn multi-level feature representations of images. Content representation is obtained from deeper feature maps that preserve spatial arrangement, while style representation is captured through correlations between features across layers. These representations allow the system to manipulate visual appearance without altering the underlying structure of the image. Understanding feature extraction is essential to performing meaningful style transfer.

5.3 Gram Matrix for Style Extraction

The Gram matrix is a mathematical tool used to measure correlations between feature maps in a CNN layer, effectively capturing texture, color distribution, and artistic patterns. By computing Gram matrices for the style image and matching them with the generated image, NST encodes stylistic characteristics independent of spatial information. This approach enables replication of painting styles, brush strokes, and textures in the synthesized output.

5.4 Loss Functions in Neural Style Transfer

NST is driven by optimization of multiple loss functions, primarily content loss and style loss. Content loss ensures that the generated image maintains the structural similarity to the original image, while style loss enforces resemblance to the artistic reference. A weighted combination of these losses guides the network in producing visually balanced outputs. Some modern methods also include total variation loss to reduce noise and improve smoothness.

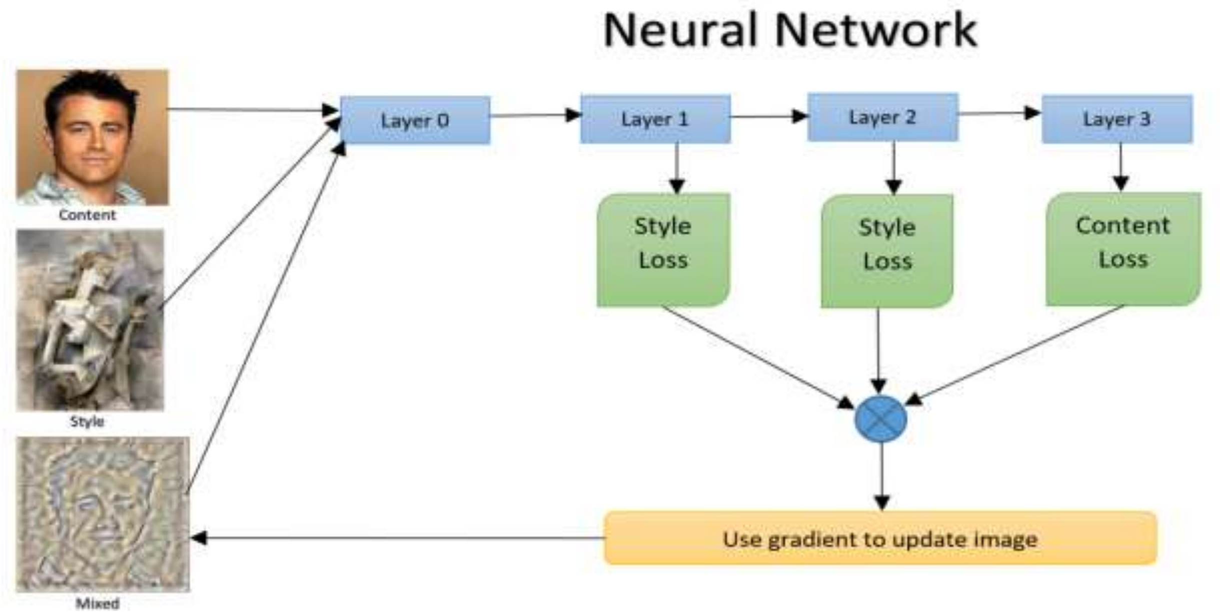
5.5 Image Optimization and Reconstruction

Once feature representations and losses are defined, NST reconstructs the stylized image by iteratively updating pixel values or by passing the image through a trained feed-forward network. Optimization-based reconstruction focuses on accuracy but is slow, whereas feed-forward reconstruction provides faster results suitable for real-time use. This balance between reconstruction quality and efficiency is a central concept in the development of modern style transfer systems.

6. Methodology

The methodology takes a content image and a style image as inputs and preprocesses them for analysis using a pre-trained VGG-19 network.
The network extracts deep features to represent content and uses Gram matrices to capture the artistic style. A combined loss function measuring content and style differences is computed to guide the transformation. The generated image is iteratively optimized until it blends the structure of the content image with the appearance of the style image.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930



6.1 Input Acquisition

The system begins by selecting two images: a **content image**, which provides the structural layout, and a **style image**, which provides artistic patterns such as texture, color, and brush strokes. These images serve as the primary inputs for the Neural Style Transfer process.

6.2 Image Pre-processing

Both images are resized to a uniform resolution and normalized according to the requirements of the pre-trained CNN model. Pre-processing ensures compatibility with the VGG-19 network and improves feature extraction accuracy.

6.3 Feature Extraction using VGG-19

A pre-trained **VGG-19 Convolutional Neural Network** is used as a feature extractor. The network is not trained again; instead, it identifies hierarchical image features:

- **Deeper layers** extract content-related structures such as object shapes and spatial arrangement.
- **Shallow and intermediate layers** extract style-related features like textures and patterns.

6.4 Style Representation using Gram Matrix

To capture artistic style, Gram matrices are computed from selected feature maps of the style image. These matrices measure correlations between features, enabling the system to encode texture, color distribution, and artistic strokes independent of image structure.

6.5 Loss Function Formulation

The generated image is evaluated using a combination of loss functions:

- **Content Loss:** Ensures similarity between generated and content image structures.
- **Style Loss:** Ensures the generated image reflects artistic patterns of the style image.
- **Total Variation Loss (optional):** Reduces noise and smoothens the output.

The total loss is calculated as a weighted sum of these components.

6.6 Optimization Process

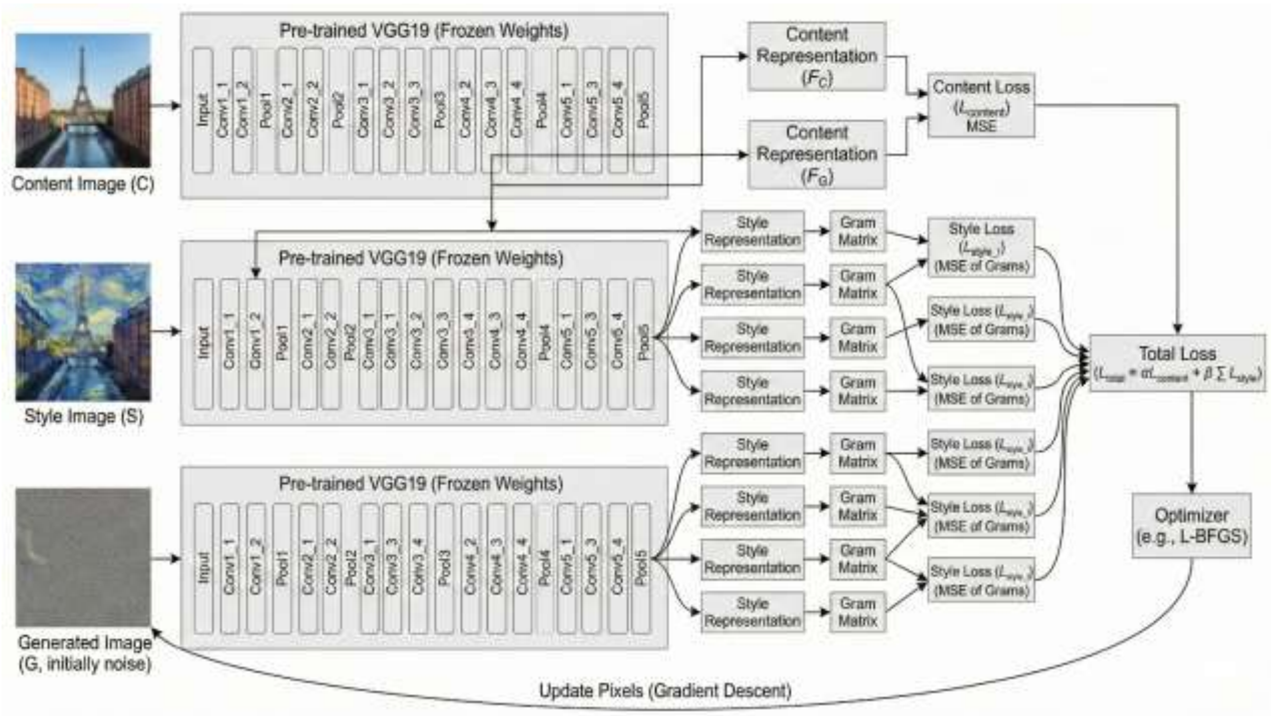
An initial image (random noise or a copy of the content image) is iteratively updated using backpropagation and gradient descent to minimize total loss. During this process, the image gradually acquires stylistic textures while maintaining structural integrity.

6.7 Image Reconstruction and Output Generation

After sufficient optimization, the final stylized image is reconstructed. The output successfully combines the semantic content of the original image with the artistic appearance of the reference style image.

6.8 System Architecture

The system architecture is designed to combine deep feature extraction with optimization-based image synthesis for effective Neural Style Transfer. The process begins with two inputs: the content image and the style image, which are pre-processed to match the required input size and normalization of the CNN model. A pre-trained VGG-19 network acts as a fixed feature extractor, identifying hierarchical representations where deeper layers capture content structure and earlier layers capture stylistic patterns. Gram matrices are computed from selected feature maps to model the texture and artistic characteristics of the style image. These representations are then used to calculate content and style losses, which are combined into a total loss function. An optimization module iteratively updates the generated image through backpropagation to minimize this loss. The final output is a stylized image that preserves the semantic layout of the content while reflecting the artistic features of the style reference.



7. Challenges and Limitations

7.1 High Computational Cost

Optimization-based Neural Style Transfer methods require multiple iterations of gradient descent to generate a single stylized image. This makes the process computationally expensive and time-consuming, especially for high-resolution images, limiting real-time usability.

7.2 Trade-off Between Speed and Quality

Fast feed-forward models can generate stylized images quickly, but they often compromise on fine artistic details compared to optimization-based approaches. Achieving both high-quality results and real-time performance remains a major challenge.

7.3 Limited Control Over Stylization

Many NST methods apply style uniformly across the image without understanding object boundaries or semantic importance. This can distort key regions such as faces or important structures, reducing realism and user control.

7.4 Generalization to Arbitrary Styles

Although modern techniques support multiple styles, some models struggle to accurately reproduce highly complex or unseen artistic patterns. Maintaining consistency across diverse styles is still difficult.

7.5 Resource and Memory Requirements

The use of deep CNNs like VGG-19 demands significant GPU memory and processing power. This restricts deployment on low-resource environments such as mobile devices or embedded systems.

8. Conclusion and Future Scope

In conclusion, this project demonstrates the effectiveness of Neural Style Transfer in generating artistic images by combining the structural content of one image with the stylistic features of another using deep convolutional neural networks. By leveraging feature extraction through a pre-trained VGG-19 model, Gram matrix-based style representation, and loss-driven optimization, the system successfully produces visually appealing stylized outputs. Despite achieving promising results, challenges such as computational cost, limited real-time flexibility, and reduced control over region-specific stylization remain. Future work can focus on developing more efficient architectures, integrating attention mechanisms for better semantic awareness, supporting high-resolution and video style transfer, and deploying lightweight models suitable for mobile and real-time creative applications, thereby expanding the usability of NST in digital art, design, and multimedia industries.

9. References

[1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A Neural Algorithm of Artistic Style. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[2] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. Proceedings of the European Conference on Computer Vision (ECCV).

[3] Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2016). Instance Normalization: The Missing Ingredient for Fast Stylization. arXiv preprint arXiv:1607.08022.

[4] Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., & Yang, M.-H. (2017). Universal Style Transfer via Feature Transforms. Advances in Neural Information Processing Systems (NeurIPS).

[5] Zhang, K., Liu, Y., Gu, S., Zuo, W., & Zhang, L. (2019). Attention-Aware Multi-Stroke Style Transfer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[6] Huang, X., & Belongie, S. (2017). Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. Proceedings of the IEEE International Conference on Computer Vision (ICCV).

[7] Chen, T. Q., & Schmidt, M. (2016). Fast Patch-Based Style Transfer of Arbitrary Style. arXiv preprint arXiv:1612.04337.

[8] Dumoulin, V., Shlens, J., & Kudlur, M. (2017). A Learned Representation for Artistic Style. Proceedings of the International Conference on Learning Representations (ICLR).

[9] Sheng, L., Lin, Z., Shao, J., & Wang, X. (2018). Avatar-Net: Multi-Scale Zero-Shot Style Transfer by Feature Decoration. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[10] Park, D. Y., & Lee, K. H. (2019). Arbitrary Style Transfer with Style-Attentional Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).