ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 01, January 2025

Dynamic Neural Style Transfer for Artistic Image Generation using VGG19

[1] Kapil Kashyap, [2] Sean Fargose, [3] Sindhu Nair

[1] [3] [2] Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering

Corresponding Author Email: [1] kapilkashyap3105@gmail.com, [2] fargosesean2808@gmail.com, [3] sindhu.nair@djsce.ac.in

Abstract— **Throughout history, humans have created remark- able works of art, but artificial intelligence has only recently started to make strides in generating visually compelling art. Breakthroughs in the past few years have focused on using convolutional neural networks (CNNs) to separate and manipulate the content and style of images, applying texture synthesis techniques. Nevertheless, a number of current techniques continue to encounter obstacles, including lengthy processing times, restricted choices of style images, and the inability to modify the weight ratio of styles. We proposed a neural style transfer system that can add various artistic styles to a desired image to address these constraints allowing flexible adjustments to style weight ratios and reducing processing time. The system uses the VGG19 model for feature extraction, ensuring high-quality, flexible stylization without compromising content integrity.**

Index Terms— neural style transfer, convolutional neural net- works, VGG19, artistic style blending, customizable style weight ratio.

I. INTRODUCTION

Neural style transfer has transformed digital art by allowing individuals to incorporate distinctive artistic styles into their images. Originally, style transfer models focused on applying a single style to a content image through the use of CNNs to combine style features with the image's structure. While effective, these single-style transfer methods limit the creative flexibility and customization desired in many real-world applications. Multi-style transfer—applying multiple styles to a single image—addresses this need by allowing more nuanced artistic expression. This shift towards multi-style applications introduces challenges in preserving both the content and the integrity of diverse style elements within a single image.

Fig. 1. Example of Single Neural Style Transfer.

An example of single style transfer is shown above, where the content image is of a cat, and the style image is of Hokusai's Great Wave. The generated target image still contains the cat but is stylized with the waves, blue and beige colors, and block print textures of the style image.

II. RELATED WORKS

A. Literature Review

Recent advancements in neural style transfer have led to innovative techniques that enable adaptive, multi-style, and content-aware transformations, with a focus on enhancing aes- thetic quality, computational efficiency, and user interactivity. Chen et al. [1] proposed an incremental learning approach for neural style transfer using a dual-generator architecture. This method allows the system to retain previously learned styles while integrating new ones, achieved through percep- tual and distillation loss functions. It is particularly suitable for real-time systems requiring adaptability and continuous learning.

Shabari and Rajlaxmi [2] introduced a fusion framework that combines GoogleNet and VGG16 to quantify image naturalness by calculating glossiness. Their method optimizes feature weights across the networks, ensuring high aesthetic fidelity and demonstrating the potential of integrating multiple architectures for enhanced perceptual rendering.

Li and Gao [3] presented a nonparametric model for local manipulation of style features via deep feature synthesis. This approach enhances control over texture and content details, enabling faster convergence and reduced computational over- head, making it ideal for high-speed iterative systems.

Yu and Zhou [4] developed the Enhanced-Channel Module (ECM), which modulates style and content feature mapsto generate content-aware weights. Their method excels in preserving local details and stylistic characteristics, proving effective in both image and video applications, especially for high-resolution artistic outputs.

Wang et al. [5] introduced an interactive multi-style transfer system that enables users to segment images and apply mul- tiple styles using pre-trained models like AdaIN and WCT. This system offers significant flexibility for user-defined style control, catering to creative and customized applications.

Gao et al. [6] proposed a video style transfer framework that employs multi-instance normalization and ConvLSTM modules to maintain temporal consistency across frames. This method effectively minimizes flickering in dynamic scenes, making it suitable for real-time video processing with consis-tent stylizations.

[7] proposed a dynamic neural style transfer approach in-troducing a weighting parameter to adjust the style-to-content blending ratio during synthesis. This provides users with greater control over customization, balancing style and content preservation while showing promise for real-time applications with further optimization.

[8] by Chen et al. introduced a dual-generator architecture using perceptual and distillation loss functions for multi-style transfer. This method preserves aesthetic quality while enabling smooth transitions between styles, proving valuable for scalable systems requiring incremental adaptability.

Zhang and Dana [9] introduced the Multi-style Genera -tive Network (MSG-Net), incorporating a CoMatch Layer for matching second-order feature statistics. This network achieves high-quality, real-time performance by preventing checkerboard artifacts and allowing features like brush-size control, enhancing operational flexibility.

Nguyen et al. [10] proposed the MULTAR framework, an extension of AdaIN, introducing noise into the style encoder to enable multimodal style transfer. This approach generates diverse, high-quality outputs efficiently, offering significant improvements to existing unimodal techniques.

Huang and Belongie [11] designed a real-time neural style transfer framework using Adaptive Instance Normalization (AdaIN). This method aligns feature statistics, balancing speed and quality while providing flexibility in content-style trade- offs, making it highly effective for real-time applications.

Gatys et al. [12] enhanced neural style transfer with color-preserving methods such as histogram matching and luminance-only transfer. These methods maintain content in-tegrity while refining the balance between style and content, addressing prior limitations in the algorithm.

Simonyan and Zisserman [13] explored deep convolutional networks with small filters to improve image recognition accuracy and efficiency. Their work, emphasizing model depth, provides a foundational framework for advancements in feature-based style transfer techniques.

Chen et al. [14] introduced a patch-based style transfer approach utilizing a simplified optimization objective that combines style textures and content structure in a single CNN layer. Their method is highly adaptable, allowing arbitrary content and style images while ensuring efficient frame-by-frame video performance. This framework emphasizes prac-tical utility with consistent outputs and intuitive parameter tuning.

Chen et al. [15] developed MXNet, a flexible machine learning library designed for efficient deep learning across heterogeneous systems. Integrating declarative symbolic ex-pressions with imperative tensor computation, the system supports distributed training and optimization.

B. Research Gaps

However, some gaps identified in previous methods include difficulty in handling images with simple textures, which can result in a loss of stylistic detail. Many approaches rely heavily on specific datasets, limiting their generalizability across varied content types. High computational demands also affect the scalability of these systems, especially in real- time applications. Additionally, subjective evaluation methods introduce inconsistency, while video style transfer models, despite improvements, still struggle to maintain stability in flickering scenes. Addressing these issues could enhance scal- ability, stylistic diversity, and overall reliability in future multi- style transfer frameworks.

III. METHODOLOGY

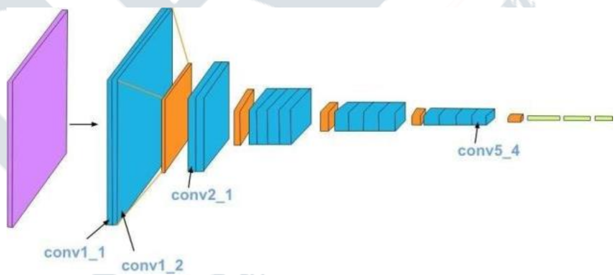


Fig. 2. Illustration of the VGG19 structure used in neural style transfer.

Our work employs the VGG19 model (seen in Fig. 2), a deep convolutional neural network (CNN) widely recognized for its robust architecture and exceptional feature extraction capabilities. The model has 16 convolutional layers (with 5 pooling layers) and 3 fully connected layers (seen in Fig.3). The VGG19 model is particularly adept at capturing hierarchical features from images, enabling it to extract both content and style information effectively. This makes it an ideal foundation for tasks that require the precise manipulationof image characteristics, such as style transfer, where subtle details and broader structural features are equally important.

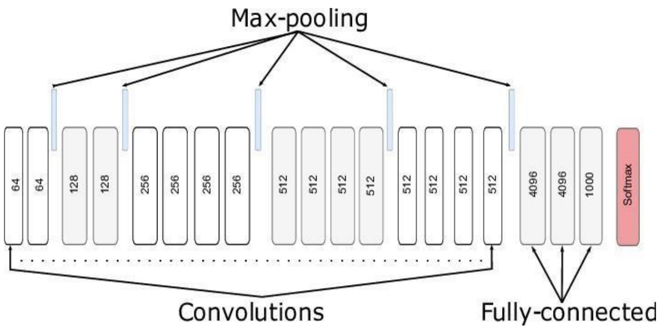


Fig. 3. Detailed view of the VGG19 model.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 01, January 2025

Building upon this foundation, our method integrates a multi-style transfer technique to achieve visually compelling results. The VGG19 model serves as the backbone for extracting content and style representations from input images, leveraging its deep layers to capture nuanced style patterns and maintaining content coherence. The e xtracted features are then processed through an optimization-based approach, designed to seamlessly blend multiple styles into a single cohesive output. By combining the strengths of VGG19's feature e xtract ion with a sophisticated optimization pipeline, our approach ensures the creation of rich ly stylized images that retain the essence of the original content while embodying the desired artistic influences.

A. Data Preprocessing

The first step involves preprocessing the input images, which consist of both the content and style images. The images are adjusted in size and standardized to guaranteeuniformity and suitability for the pre-trained VGG19 model. Normalization is crucial in preserving the model's stability while e xtract ing features, making sure that pixel values are ata consistent scale.

B. Feature Extraction

For feature e xtract ion, content features are derived from the fourth convolutional layer (conv4 2) of the VGG19 model. This layer captures high-level semantic information from the content image, forming the backbone of the transferred content. Style features, on the other hand, are e xtracted froma set of layers (conv1 1, conv 2 1, conv3 1, conv4 1, andconv5 1). These layers capture textures and patterns at various scales, ranging from fine details to global structural in forma- tion. The combination of features from these layers ensures a comprehensive representation of the style.

C. Style Transfer Process

The process of style transfer starts with the calculation of style features from every style image. The spatial correlations of pixe l values in the feature maps are quantified using Gram matrices to represent these features. Gra m matrices are important for transforming the style of an image by capturing textures and color distributions. These images are utilized as our style images.

Fig. 4. Style Image 1: Monet's painting used for stylization.

The target image is init ially a copy of the content image and is iteratively updated during the optimization process to incorporate elements from the chosen styles. Below is the content image we used.

Fig. 5. Style Image 2: Sandstone texture for stylization.

Fig. 6. Style Image 3: Stone texture for stylization.

D. Loss Calculation

The loss function in style transfer consists of two primary components: content loss and style loss. The content loss measures the difference between the content features of the target image and the content image. The style loss, on the other hand, evaluates the differences between the Gram matrices of the target image and the style images. The total loss is the weighted sum of these two losses, with weights assigned to control the relative influence of each. The optimization process minimizes the combined loss, ensuring that both the content and style are appropriately blended in the target image.

article amsmath

The total loss function is given by:

$$L_{\text{total}} = \alpha \cdot L_{\text{content}} + \beta \cdot L_{\text{style}} \tag{1}$$

where α and β are the weights for the different losses in the loss function. The total loss function, minimized during the optimization process, consists of two primary components: content loss Lcontent and style loss Lstyle.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 01, January 2025

Fig. 7. Content Image.

The content loss quantifies the difference between the content features of the target image and those of the original content image. It is defined as:

to style features obtained from various layers of the VGG19 model: conv1_1 (1.0), conv2_1 (0.75), conv3_1 (0.2), conv4_1 (0.2), and conv5_1 (0.2), demonstrating a struc- tured method for te xture and pattern alignment. A learning rate of 0.003 is utilized to manage the step size of optimization, with 2000 iterations (steps) to guarantee convergence, and intermediate outcomes are visualized every 400 steps.

$$L_{\text{content}} = \frac{1}{2}\sum_i (F_{ic} - F_{it})^2 \tag{2}$$

where F_{ic} and F_{it} denote the content features extracted from layer i of the content and target images, respectively.

The style loss, L_{style}, evaluates the difference between the Gram matrices of the target and style images, capturing the spatial correlations that represent the texture and color distribution of the style. It is defined as:

The optimization procedure reduces the overall loss function, using weights $\alpha = 1$ and $\beta = 10^9$. The VGG19 model utilizes the Re LU (Rectified Linear Unit) activation function, which is defined mathematically as:

$f(x) = \max(0, x)$,

where x is the input. ReLU introduces non-linearity to cap- ture intricate patterns in the data while effectively mitigating vanishing gradient problems.

$$L_{\text{style}} = \sum_l \frac{\sum (G_{lt} - G_{ls})^2}{4N_l^2 M_l^2} \tag{3}$$

where G_{lt} and G_{ls} are the Gram matrices of the target and style images at layer l, and N_l and M_l are the dimensions of the feature map in that layer.

E. Optimization and Multi Style Application

In order to enhance the desired image, we utilize the Adam optimizer, which continuously adjusts the image to reduce the overall loss. Each plan is carried out in sequence, one after the other. By tuning the weights of the content and style losses, we can manage the impact of each styleon the eventual result. This enables the creation of a variedand visually appealing image by blending different styles creatively.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{4}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \tag{5}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \tag{6}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \tag{7}$$

$$\vartheta_{t+1} = \vartheta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}. \tag{8}$$

F. Result Visualization

After the iterative optimization process, the final stylized image is obtained, which incorporates the content of the original image and the artistic ele ments from multiple styles. The output image is displayed to showcase the combined artistic in fluences, highlighting the ability of our method to seamlessly blend different styles while preserving the content of the original image. This method allows for versatile and effective application of various styles to an image, giving more creative influence over the end product. The method not just combines different artistic styles but also lets you change the balance between style and content, making it very versatile for different artistic uses.

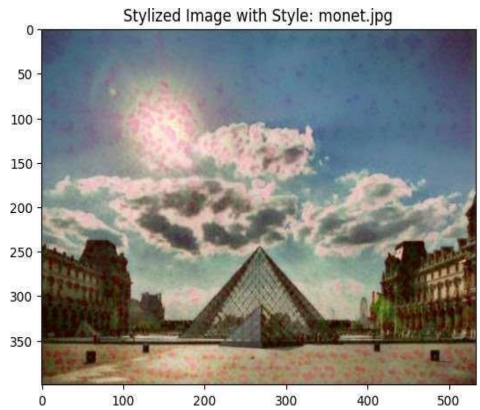


Fig. 8. Stylized Image using Monet

The equations above shows the Adam Optimizer, which updates neural network weights using adaptive moment esti- mation by combining moving averages of gradients (m_t) and their squared values (v_t), corrected for bias, to achieve efficient and stable convergence.

IFERP Explore Your Research Journey...

ISSN (Online) 2394-2320

International Journal of Engineering Researchin Computer Science and Engineering (IJERCSE)

Vol 12, Issue 01, January 2025

G. Hyperparameters

The hyperparameters for the proposed neural style transfer framework are carefully chosen to strike a balance between conserving content and enhancing stylistic effect.

The content weight is established at 1, guaranteeing that the content image's structural integrity is preserved, whereas the style weight is notably greater at 1 10^9, h× ighlighting the stylistic features. Weights specific to each layer are assigned while introducing stylistic features in a controlled manner.

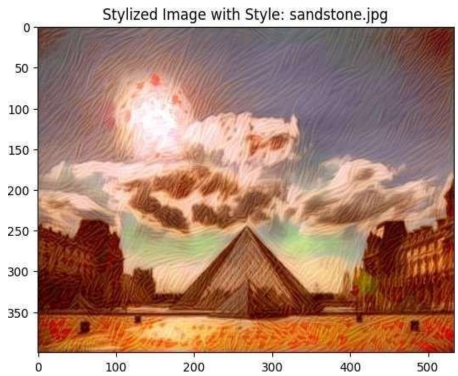


Fig. 9. Stylized Image using Sandstone

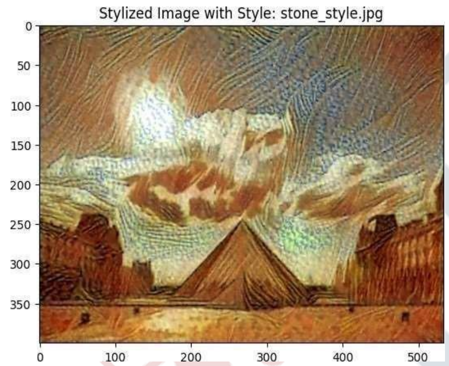


Fig. 10. Stylized Image using Stone Style

Fig. 11. Final Stylized Image

IV. RESULTS

The results of our multi-style neural style transfer approach are presented in this section, showcasing both qualitative and quantitative assessments of the generated images. Our method successfully applies multiple art istic styles to a target image, preserving the content structure while introducing stylistic features in a controlled manner.

Figure 12 showcases the original image and its stylized forms created through the application of various artistic styles. To start, the content image goes through multiple steps in which each style is added individually. The outcome is an ultimate picture that combines all the various styles into one. The visual aspects like te xtu re, brush strokes, and color differences in each image are easily seen while still maintaining the original content structure. The findings demonstrate our method's ability to seamlessly combine various styles while maintaining the content's clarity intact.

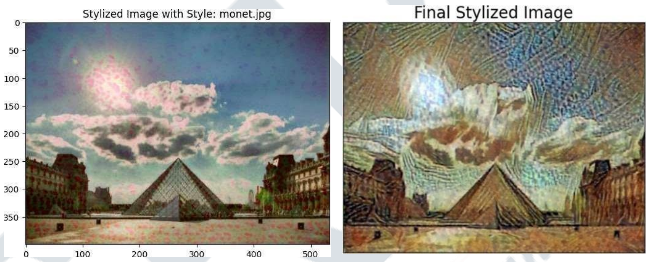


Fig. 12. Final Stylized Image Style 3

Fig. 13. Single Style Image compared to Multiple Stylized Images

Figure 13 compares the outputs of our method with traditional single-style neural transfer techniques. The image on theleft shows the content image with a single style applied, while the image on the right shows the result of applying multiple styles. Our method showcases a visually complex outcome, showing the skill to blend various art istic influences without overshadowing the content.

The following figures shows the styling of a wall based on3 styles.

Fig. 14. Content Image – Wall

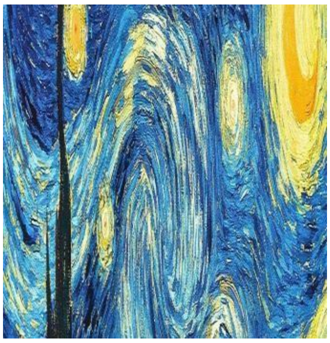
ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 01, January 2025

Style 1

Style 2



Style 3

Fig. 15. Styling Images

Fig. 16. Final Stylized Image

A. Quantitative Results

The following graph consists of two subplots that represent the loss curves for style loss and content loss over iterations. The x-axis in both plots represents the number of iterations, while the y-axis represents the respective loss values. The style loss curve begins at a higher value on the left plot and decreases exponentially with more iterations. At first, decay happens quickly, but it lessens as iterations continue, eventually leveling off. This pattern is commonly seen in style transfer models as the style loss diminishes with time. The content loss curve shown on the right also shows an e xponential decay pattern. Just like the style loss, the content loss begins at a high level and gradually decreases at a slightly slower

V. CONCLUSION

This project effectively investigates the possibilities of neural style transfer for various artistic applications. Us ing the VGG19 network for feature extraction, we created a systemthat can apply various styles to one image, resulting in customizable visual outputs. Optimizing both content and style loss through backpropagation allows for the smooth incorporation of various artistic elements without compromis- ing the fundamental characteristics of the initial image. Our method showcases how deep learning can be adaptable in artistic fields, enabling individuals to e xp lore various styles and modify style intensities. The results show the ability to seamlessly combine various styles, indicating potential for use in fields like interior design and producing artistic NFTs.

VI. FUTURE SCOPE

Future enhancements for the system may include improve- ments in real-time processing for faster and more efficient style transfer. Through the utilization of effective computational techniques, we can greatly diminish processing time, rendering the system more appropriate for real-life scenarios. Moreover, delving into adaptive style blending using image content and testing out more intricate neural network structures could increase the versatility and e xcellence of the results. Potential ways to enhance the reach and practicality of this project include adding support for video style transfer and creating a user-friendly interface for manipulating styles.

REFERENCES

[1] C. Ma, Z. Sun, and C. Ruan, " Style creation: multiple styles transfer with incremental learning and distillation loss," 2024.

[2] S. Nath P and R. Chouhan, "Quantifying image naturalness using transfer learning and fusion model," 2024.

[3] D. Li and W. Gao, " Neural style transfer based on deep feature synthesis," 2023.

[4] X. Yu and G. Zhou, " ECM: arbitrary style transfer via enhanced-channel module," 2023.

[5] X. Wang, Y. Lyu, J. Huang, Z. Wang, and J. Qin, "Interactive artistic multi-style transfer," 2021.

[6] W. Gao, Y. Li, Y. Yin, and M.-H. Yang, "Fast video multi-style transfer," 2020

[7] G. Gao, Z. Wang, X. Li, and M. Li, " Video multi-style transfer with tem- poral consistency using multi-instance normalization and ConvLSTM," 2023.

[8] X. Wang, Z. Lyu, and J. Huang, "Interactive artistic multi-style transfer," 2022.

[9] H. Zhang and K. Dana, "Multi-style Generat ive Network for Real-time Transfer," 2024.
[10] A.-D. Nguyen, S. Choi, W. Kim, and S. Lee, "A Simp le Way of Multimodal and Arbitrary Style Transfer," 2019.
[11] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," 2017.
[12] L. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman, "Preserving color in neural art is tic style transfer," 2016.
[13] K. Simonyan and A. Zisserman, "Ve ry deep convolutional networks for large-scale image recognition," 2015.
[14] T. Q. Chen and M. Schmidt, "Fast patch-based style transfer of arbitrary style," 2016.
[15] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, et al., "MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems," 2015.