



(RESEARCH ARTICLE)



## Deep artistry: Blending styles with neural networks

Kavitha Soppari, V. Revathi Chandrika \*, Yogitha Setty and O.Sakshith

*CSE-AI and ML, ACE Engineering College Hyderabad, India.*

International Journal of Science and Research Archive, 2025, 14(01), 630-637

Publication history: Received on 04 December 2024; revised on 10 January 2025; accepted on 13 January 2025

Article DOI: <https://doi.org/10.30574/ijrsra.2025.14.1.0092>

### Abstract

This work investigates the application of Neural Style Transfer (NST) using Convolutional Neural Networks (CNNs), with a specific focus on the VGG16 model. The proposed system combines the structural details of a content image with the artistic characteristics of a style image. By employing a dual-network framework, content and style features are extracted independently, and a stylized image is generated by minimizing a combined loss function through iterative optimization. The research highlights advancements in processing efficiency, enabling potential real-time applications in video processing. The system's adaptability makes it suitable for diverse creative fields such as digital art, graphic design, and multimedia production. Future enhancements aim to incorporate real-time style transfer for dynamic content generation in video and other applications.

**Keywords:** Neural Style Transfer; VGG16; Content Preservation; Style Blending; Image Processing; Deep Learning; Digital Art

### 1. Introduction

Neural style transfer (NST) is a deep learning technique that combines the structural elements of a content image with the artistic features of a style image using convolutional neural networks (CNNs), particularly pre-trained models like VGG16. NST effectively extracts hierarchical features to distinguish between content and style. The aim is to produce a new image that retains the layout and structure of the content while incorporating the artistic characteristics of the style image. This approach has revolutionized digital art and multimedia by automating complex artistic transformations. Recent advancements in NST have improved its efficiency, enabling higher-resolution outputs, reduced computational demands, and even real-time applications on mobile and edge devices. The system utilizes a dual-network approach, where one network extracts content features and the other captures style features. By minimizing a combined loss function (content loss and style loss) through iterative optimization, the model creates a stylized image that balances content retention with artistic transformation. This approach not only enhances the visual appeal of the output images but also optimizes processing speed and resource utilization. Potential applications include digital art, graphic design, and multimedia production.

### 2. Literature survey

#### 2.1. Early Works

[1]. Gatys et al. (2016): Introduced NST using the VGG-19 architecture to extract content and style features. They optimized a loss function combining content loss (to preserve structure) and style loss (to capture textures and patterns) using Gram matrices.

\* Corresponding author: V. Revathi Chandrika

[2].Global Approaches:These methods, like those of Gatys et al. (2016) and Risser et al. (2017), focus on maintaining overall consistency in style transfer across the image. They employ techniques such as Gram matrices and histograms for holistic style representation.

[3].Local Approaches: Researchers like Li and Wand (2016) used patch-based methods and Markov Random Fields (MRFs) for detailed stylization. While effective, these methods sometimes introduce artifacts due to inconsistencies between patches.

[4].Hybrid Approaches: Zhao et al. (2019) proposed models combining global and local techniques to achieve a balance between content preservation and fine-grained stylization. These approaches use fusion loss functions to enhance aesthetic quality while addressing computational challenges.

---

### 3. Objectives

There are 4 objectives. They can be listed as:

- **Implementation of Neural Style Transfer (NST):** The project leverages Convolutional Neural Networks (CNNs) to merge the structural elements of one image with the artistic features of another. A pre trained VGG16 model is utilized to extract multi-level hierarchical features, making it well-suited for this task.
- **Content Preservation and Artistic Application:** The system maintains the essential structure and key features of the content image while incorporating artistic patterns, textures, and colors from the style image. This integration is achieved through iterative optimization processes, ensuring a seamless fusion of content and style.
- **Optimization of Loss Functions:** A combined loss function is employed, which balances content and style retention during the transformation process. This approach enhances the visual appeal the output by ensuring both structural integrity and aesthetic quality.
- **Exploration of Real-World Applications:** The system is designed with potential applications in digital art, creative design, and multimedia production.

---

### 4. Methodology

#### 4.1. Model Selection

For Neural Style Transfer, we utilized VGG16, a pre-trained Convolutional Neural Network (CNN), due to its effective architecture in extracting both content and style features. With its 16 layers, VGG16 strikes a balance between performance and computational efficiency, making it an ideal option for real-time applications. Although deeper networks like VGG19 provide enhanced accuracy, VGG16 is optimized for tasks requiring efficient processing.

#### 4.2. Pre-processing

The images are adjusted in size and normalized to conform to the required input specifications of VGG16, which typically expects dimensions of 224×224 pixels. This standardization ensures consistent input across various image formats, such as JPEG and PNG, and different resolutions. Normalization adjusts pixel intensity values, facilitating better feature recognition and enhancing the model's ability to process relevant content.

#### 4.3. Style Transfer Process

**Content Feature Extraction:** The deeper layers of VGG16 focus on capturing the content features of the input image, such as object shapes and spatial arrangement.

**Optimization:** To generate the final image, a combined loss function is minimized, which incorporates both content and style losses. This process, carried out iteratively through gradient descent, refines the output, striking a balance between preserving the content and embedding the desired style.

#### 4.4. Output Generation

The final stylized image undergoes post-processing, including resizing and format conversion (e.g., PNG or JPEG). These steps ensure compatibility with various platforms and applications.

## 5. Proposed system

The Deep Artistry system is designed to automate the Neural Style Transfer process, seamlessly integrating the structural elements of one image with the stylistic patterns of another. The system employs a modular architecture, ensuring efficient processing and high-quality outputs. It consists of three main components: the Input Handler, the Neural Network Module, and the Output Generator. Each of these components serves a specific function within the overall workflow.

### 5.1. Input Handler

#### 5.1.1. Purpose

The Input Handler is responsible for preparing the images provided by the user for further processing. It handles the content image, which defines the primary structure or subject, and the style image, which contributes the artistic elements such as textures, patterns, and colors.

#### 5.1.2. Key Features:

- Support for Standard Formats:  
Users can upload images in widely-used formats like JPEG and PNG, ensuring compatibility and ease of integration.
- Pre-Processing Steps:
  - Resizing: Images are resized to meet the input dimension requirements of the model, ensuring optimal processing performance.
  - Normalization: Pixel values are standardized, enhancing the model's capacity to identify and process relevant features, thereby improving consistency and accuracy.

#### 5.1.3. End Result:

At this stage, both the content and style images are fully prepared and optimized for feature extraction in the subsequent module.

### 5.2. Neural Network Module

#### 5.2.1. Purpose:

The primary function of this module is to extract features from both the content and style images and combine them to generate the stylized output image.

#### 5.2.2. Key Functions:

- Feature Extraction Using VGG16:
  - A pre-trained network, VGG16, is leveraged for its efficient capability in detecting and extracting relevant features from images.
  - Content Features: The deeper layers of the VGG16 model focus on capturing structural information, such as the placement of objects and shapes within the content image.
  - Style Features: The more superficial layers of the network are dedicated to identifying and extracting textures, patterns, and color schemes present in the style image.
- Capturing Artistic Style:

The system analyzes the style image's different regions and their interactions. These relationships are mathematically encoded to represent the artistic characteristics of the image.

- Generating the Stylized Image:

The content and style features are merged to produce the stylized image. This process involves iterative refinement, which ensures an optimal balance between the content's structural integrity and the style's artistic patterns.

- Optimization for Refinement:

The generated image undergoes multiple optimization steps to closely match the desired combination of content and style. These adjustments continue until the image achieves an aesthetically pleasing result.

#### 5.2.3. End Result:

The Neural Network Module produces an output image that maintains the content's structural features while successfully incorporating the artistic style derived from the style image.

### 5.3. Output Generator

#### 5.3.1. Purpose:

The Output Generator finalizes the stylized image, ensuring it is polished and ready for use.

#### 5.3.2. Key Functions:

- Post-Processing:

The system adjusts the final image for visual quality. It also allows resizing based on user preferences, making the output versatile for various purposes, such as digital prints or wallpapers.

- Format Conversion:

The stylized image is saved in user-specified formats, such as JPEG or PNG, to ensure compatibility with different platforms and applications.

#### 5.3.3. End Result:

The Output Generator delivers a refined and stylized image, ready for download or use in creative projects.

### 5.4. Workflow Summary

The complete system workflow consists of the following steps:

#### 5.4.1. User Upload:

Users upload their content and style images in supported formats.

#### 5.4.2. Pre-Processing:

The uploaded images are resized and normalized to meet the neural network's input requirements.

#### 5.4.3. Style Transfer:

The Neural Network Module extracts features from the images, blends them, and produces a stylized output.

#### 5.4.4. Post-Processing and Delivery:

The output image is refined, resized, and saved in the user's preferred format.

### 5.5. Applications

The Deep Artistry system has numerous applications in creative fields, including:

- Digital Art: Transform photographs or original artwork into distinctive stylized pieces, enabling artists to explore creative variations.
- Design: Generate custom patterns for use in textiles, branding materials, posters, and other graphic design projects, enhancing creative flexibility.
- Multimedia Production: Improve the visual impact of video and animation projects by incorporating unique artistic styles, expanding possibilities for content creators in industries such as advertising, film, and gaming.

## 6. Algorithms

### 6.1. Input Handling

The process starts with preparing the content and style images for processing.

- Loading Images:
  - The content image ( $I_c$ ) and style image ( $I_s$ ) are loaded into the system.
- Resizing Images:
  - Both images are resized to the same dimensions ( $H \times W$ ) to ensure the generated image matches the required size.
- Image Normalization:
  - The pixel values of the images are normalized to align with the input requirements of the pre-trained model (e.g., VGG16).

The normalization formula is: where:

$$I_{norm} = \frac{I - \mu}{\sigma}$$

- $I$ : The image matrix.
- $\mu$ : Mean pixel value of the training dataset.
- $\sigma$ : Standard deviation of pixel values from the training dataset.

### 6.2. Feature Extraction

The pre-trained VGG16 model is used to extract meaningful features from the input images.

#### 6.2.1. Extracting Content Features:

Content features are derived from a deeper layer  $l$  of VGG16 to capture the structural aspects of the content image:

$$F_c^l = \text{VGG16}_l(I_c)$$

Where  $F_l$  represents the feature map of the content image at layer  $l$ .

#### 6.2.2. Extracting Style Features:

Style features are represented through Gram matrices, capturing the relationships between feature maps in the style image. The Gram matrix for layer  $l$  is computed as:

$$G_s^l = F_s^l (F_s^l)^T$$

where:

- $F_l$ : Feature map of the style image at layer  $l$ .
- $G$ : Gram matrix summarizing the texture and patterns of the style image.

### 6.3. Optimization

The stylized image ( $I_g$ ) is created by minimizing a loss function that combines content and style losses.

#### 6.3.1. Content Loss:

The content loss measures the similarity between the feature map of the generated image ( $F_l$ ) and the content image ( $F_c^l$ ):

$$L_{content} = \frac{1}{2} \sum_{i,j} (F_g^l(i,j) - F_c^l(i,j))^2$$

where:

- $F_l$ : Feature map of the generated image.
- $F_c$ : Feature map of the content image.

### 6.3.2. Style Loss:

The style loss quantifies the difference between the Gram matrices of the generated image ( $G$ ) and the style image ( $G_s$ ):

$$L_{style} = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_g^l(i,j) - G_s^l(i,j))^2$$

where:

$N_l$ : Number of feature maps in layer  $l$ .

$M_l$ : Size of each feature map (height  $\times$  width).

### 6.3.3. Total Loss Function:

The total loss combines content and style losses, with weights  $\alpha$  and  $\beta$  controlling their relative importance:

$$L_{total} = \alpha L_{content} + \beta L_{style}$$

### 6.3.4. Optimizing the Loss:

The generated image is refined iteratively to minimize the total loss using optimization algorithms (e.g., Adam or L-BFGS):

$$I_g = \operatorname{argmin}_I L_{total}(I)$$

## 6.4. Output Handling

After the optimization process, the resulting image undergoes additional steps for final user delivery.

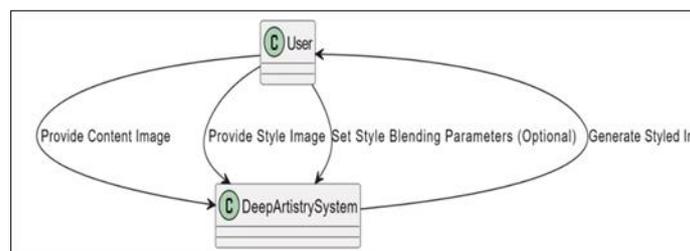
### 6.4.1. Post-Processing:

The pixel values of the generated image are adjusted to their original range ([0, 255], [0, 255], [0, 255]) to ensure accurate color reproduction.

### 6.4.2. Saving the Stylized Image:

The final image is saved in a format chosen by the user, such as PNG or JPEG.

This algorithm processes the input images by extracting their features and blending them through optimization. By minimizing the combined loss function, it produces a new image that maintains the structure of the content image while incorporating the artistic patterns of the style image.

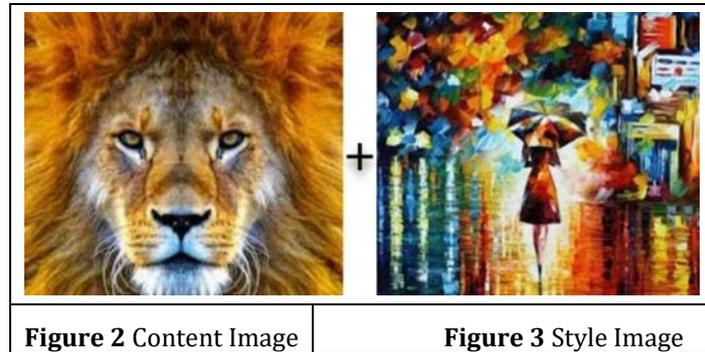


**Figure 1** Artistic style transfer

Fig. 1 illustrates the workflow of the Deep Artistry System, enabling users to create stylized images. The process begins with the user providing a content image, which serves as the main subject or visual element, and a style image, which dictates the artistic features or style to be applied. Optionally, the user can adjust style blending parameters to control

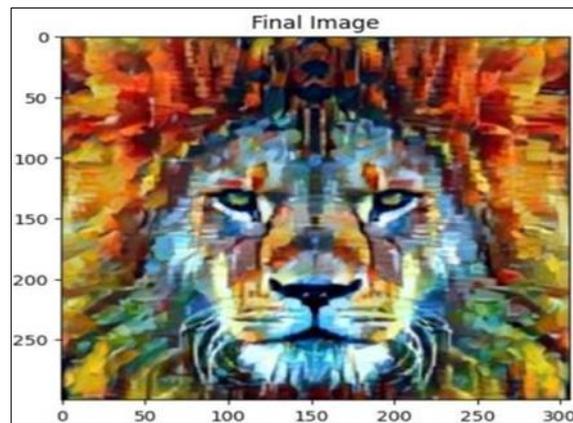
the extent of style application. The system then processes these inputs and generates the final image, merging the content's structure with the stylistic elements to produce a unique artistic version of the original content.

## 7. Result



The results showcase the effectiveness of neural style transfer in blending the structural characteristics of a content image with the artistic elements of a style image. Figure 4 depicts the final stylized output, created by combining the structural details of the content image (Figure 2) with the vivid textures and patterns from the style image (Figure 3). The output successfully retains the key structural elements of the content image while adopting the vibrant colors and artistic textures of the style image, achieving a harmonious integration of both.

This outcome emphasizes the capability of neural style transfer algorithms to maintain critical aspects of the content image, such as its structure and composition, while effectively applying stylistic attributes like color schemes and textures from the style image.



**Figure 4** Output image

## 8. Conclusion

The Deep Artistry system enhances Neural Style Transfer (NST) by utilizing the VGG16 model, which effectively balances content preservation with artistic style integration. It addresses key challenges, including format compatibility, processing efficiency, and ensuring high-quality outputs. Its flexibility and scalability make it ideal for creative industries such as advertising, gaming, and virtual reality. By automating artistic transformations, the system democratizes access to advanced tools, fostering innovation. This project investigates the application of NST with Convolutional Neural Networks (CNNs), focusing on VGG16 to merge content and style from different images. The goal is to generate visually striking images that retain the structure of the content while incorporating the textures, patterns, and colors of the style. The system uses a dual-network approach: one captures content features, while the other extracts style features. Through iterative optimization and minimizing the combined content and style losses, the model produces a balanced stylized output. This approach not only enhances the aesthetic quality but also improves efficiency, addressing speed and resource management. It holds potential in fields like digital art, graphic design, and multimedia

production, with possible real-time applications in video processing for dynamic content generation in industries like gaming and advertising.

### 8.1. Future enhancement

- 3D Model Style Transfer: Extending style transfer to 3D models would enable the application of artistic effects to 3D objects. This could benefit industries like gaming and animation by creating more dynamic and immersive visual experiences.
- Live Stream Style Transfer: Adding real-time style transfer for live video streams would allow users to apply artistic effects while broadcasting. This feature would enhance live events, social media streams, and gaming broadcasts with creative visual effects.
- Cloud-Based Real-Time Collaboration: Implementing real-time collaboration in cloud platforms would let users work together on projects simultaneously. This would benefit teams in creative fields, enabling seamless collaboration and feedback on designs from different locations.

---

### Compliance with ethical standards

#### *Disclosure of conflict of interest*

No conflict of interest to be disclosed.

---

### References

- [1] Gatys, L.A., Ecker, A.S., & Bethge, M., "A Neural Algorithm of Artistic Style Using VGG-19 Architecture for Image Feature Extraction", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. Available at: <https://ieeexplore.ieee.org/document/7780634>.
- [2] Risser, E., Wilmot, P., & Barnes, C., "Global Methods for Artistic Style Transfer Using Gram Matrices and Histogram- Based Techniques," 2017. Available at: <https://arxiv.org/abs/1701.08893>.
- [3] Li, Y., & Wand, M., "Local Approaches for High-Quality Stylization Using Patch-Based Methods and Markov Random Fields," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. Available at: <https://ieeexplore.ieee.org/document/7780641>
- [4] Zhao, S., Zhang, L., & Wu, Y., "Hybrid Approaches for Neural Style Transfer Using Fusion Loss Function to Balance Content Preservation and Detailed Stylization," 2019. Available at: [https://www.researchgate.net/publication/350184156\\_Evaluate\\_and\\_improve\\_the\\_quality\\_of\\_neural\\_style\\_tr](https://www.researchgate.net/publication/350184156_Evaluate_and_improve_the_quality_of_neural_style_tr).
- [5] Zhao, H.-H., "Image Neural Style Transfer With Global and Local Optimization Fusion," 2019. Available at: [https://www.researchgate.net/publication/350184156\\_Evaluate\\_and\\_improve\\_the\\_quality\\_of\\_neural\\_style\\_tr](https://www.researchgate.net/publication/350184156_Evaluate_and_improve_the_quality_of_neural_style_tr)
- [6] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M., "Artistic Style Transfer: Preserving Color Fidelity Using VGG- 19 and Luminance Transfer," IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 11, pp. 3365-3385, Nov. 2020. Available at: <https://ieeexplore.ieee.org/document/8732370>.
- [7] Prathamesh Jadi , P., Desai, V., Bidargaddi, A.P., Ransubhe, S., & Mummigatti, S., "Optimizing Color Preservation in Artistic Style Transfer with VGG-19 and Luminance Transfer Approach," 2024 IEEE. Available at: <https://ieeexplore.ieee.org/document/10593426>.