# Visual Style Transfer Using VGG16 and Masking

Vansh Ahlawat
*Stud. of Computer Science*
*Bennett University*
Greater Noida, Uttar pradesh
e22cseu0403@bennett.edu.in

***Abstract*—This advanced project focuses on Neural Style Transfer and is carried out with the use of a VGG16 convolutional neural network. There will be an image created from mixing the content of one image with the style of another. This is where a third input, namely a mask image, is introduced, whereby it decides how and where the style is going to be applied. High-level content and style features are extracted by such a pretrained deep learning model while carrying out iterative optimization to generate the stylized outputs. A user-defined mask allows certain control of the transfer by enabling selective stylization for specific regions. The technique shows how deep learning is used creatively to manipulate images without training a new model, but rather by leveraging feature extraction and optimization. .**

## I. Introduction

Recently, thanks to deep learning methods, image creation and manipulation witnessed tremendous developments. One such development is Neural Style Transfer, a method that creates newer images by associating the content of one image with the style of another in what is considered an art form. It basically applies to art, design, content creation, and its evolution still continues, now incorporating several new approaches and control mechanisms. In the traditional setup of Neural Style Transfer, one uses a pretrained convolutional neural network, normally VGG16 or VGG19, trained on large datasets like ImageNet. These models are used to extract features for both the content and style images. The NST algorithm then performs gradient descent on a randomly set image (or, a copy of the content image) with respect to a loss combining content and style. This results in a stylized output that closely resembles the structure of one image but appears with the style of another. Despite the improvements attained, classical NST methods often struggle to offer fine-grained control over where to apply the style, limiting their use in real cases where selective stylization is required. To alleviate this, our project proposes a third element: the mask image, through which one may discriminate and selectively control the regions of the content image that will be subject to the style transformation. The mask then acts as a spatial guide that will blend style features in the regions specified while holding to the content features in others. This project achieves stylization using the VGG16 network for feature extraction without retraining the model. Instead, the optimization process focuses on the input image, which is iteratively updated using gradient descent. The system is implemented in Python using PyTorch and executed through a Jupyter Notebook interface for ease of experimentation

## II. Related Work

Neural Style Transfer (NST) represents a remarkable achievement which joins deep learning techniques with artistic image synthesis methods. The paper by Gatys et al. established CNNs especially VGG networks as the base to split content from style in images before restoring the original structure. The authors showed pretrained models could extract both content-level representations and texture-level statistics which their optimization process used to generate stylized images. NST development has mainly concentrated on achieving higher speed and expanded flexibility. The authors Johnson et al. created a real-time image stylizing feed-forward network through individual model training for various styles. The researchers from Ulyanov et al. advanced this framework through instance normalization and achieved excellent visual outcomes with enhanced training stability. Two advancements in style transfer occurred when Adaptive Instance Normalization (AdaIN) alongside GANs enabled non-retraining applications of arbitrary styles to images. The research field investigates several style transfer approaches in addition to methods that maintain image structure throughout processing. Through their work Li and Wand developed a patch-based loss to maintain spatial consistency between images and Huang and Belongie presented innovations for better content and style feature alignment. The optimizationbased method functions as the fundamental basis for style transfer while offering simplicity and interpretability because of its basic design. The proposed work improves upon the Gatys convention by using the VGG-16 network to obtain features from multiple content and style layers. This input optimization method enables high-quality stylization however it results in slower processing speed because it controls and makes the style transfer process easier to interpret

## III. Proposed Methodology

.

### A. Overview

The project focuses on performing a neural style transfer with three input elements of content image, style image, and mask image for specifying the area of style application. The new image is desired to have content from the content image, artistic texture from the style image, and regional constraints from the mask.

### B. Model Selection

The VGG16 model pretrained on the ImageNet dataset has been utilized. We use only the convolutional layers because they can capture relevant hierarchical information from the input images. On the other hand, fully connected layers were discarded, as they are meant primarily for classification purposes and not for feature extraction

### C. Equations

In this section, we describe the mathematical formulations that govern the visual style transfer process with mask-based regional control.

#### 1. Content Loss

The content loss ensures that the generated image retains the structural features of the content image. It is calculated as the mean squared error between the feature representations of the generated and content images at a specific convolutional layer:

$$\mathcal{L}_{\text{content}} = \frac{1}{2}\sum_{i,j}(F_{ij}^{l} - P_{ij}^{l})^2 \tag{1}$$

where $F^l$ and $P^l$ denote the feature maps at layer $l$ from the generated and content images, respectively.

#### 2. Style Loss

The style loss measures the difference in correlations between feature maps of the generated and style images using Gram matrices. It captures the texture and style information:

$$\mathcal{L}_{\text{style}} = \sum_{l=1}^{L} w_l \cdot \frac{1}{4N_l^2 M_l^2}\sum_{i,j}(G_{ij}^{l} - A_{ij}^{l})^2 \tag{2}$$

Here, $G^l$ and $A^l$ represent the Gram matrices of the generated and style images at layer $l$, $N_l$ and $M_l$ are the number of feature maps and size of each feature map, and $w_l$ is the weight assigned to that layer.

#### 3. Masked Style Loss

The mask guides where the style should be applied in the generated image. The masked style loss ensures that the style is applied only to specified regions:

$$\mathcal{L}_{\text{masked\_style}} = \sum_{x,y} M(x,y)\cdot(I_{\text{gen}}(x,y) - I_{\text{style}}(x,y))^2 \tag{3}$$

where $M(x,y)$ is the value of the mask at pixel location $(x,y)$, and $I_{\text{gen}}$, $I_{\text{style}}$ are pixel values from the generated and style images, respectively.

#### 4. Total Loss

The total loss is a weighted combination of the content, style, and masked style losses. It drives the optimization of the generated image:

$$\mathcal{L}_{\text{total}} = \alpha\cdot\mathcal{L}_{\text{content}} + \beta\cdot\mathcal{L}_{\text{style}} + \gamma\cdot\mathcal{L}_{\text{masked\_style}} \tag{4}$$

where $\alpha$, $\beta$, and $\gamma$ are scalar hyperparameters that control the influence of each component.

### D. Feature Extraction

For the content image, features are extracted from an intermediate layer of VGG16 that constitutes an adequate level representation of the structural elements of the image. For the style images, features are extracted from multiple layers so that the style features can be represented by a wide range of textures and artistic patterns. These features extracted from style image and the content image form our style and content representations.

### E. Role of the Mask

The mask image defines which areas of the content image are affected by the style. Essentially, it is converted into a greyscale image and resized to the size of input images. At the loss calculation stage, the mask underlines style transfer into specific regions and keeps the rest unchanged

### F. Optimization Process

The final image is initialized as a clone of the content image. Using gradient-based optimization, updates are applied iteratively to the image by minimizing combinations of content loss, style loss, and maybe an additional mask-guided constraint. The content loss intends that the generated image reflects the structure of the original, while the style loss ensures that it matches the artistic texture of the style image. Masks are important in restricting wrinkling to regions of interest.

### G. Output Formation

After several iterations of optimization, the final stylized image is born. It adopts the structure of the content from the first image but carries the artistic style from the second-image style, being applied only in areas as specified by the thirdimage mask. The output looks very well, but the mask, along with loss weights, can still be tuned.

### H. Figures and Tables

Table 1 demonstrates the impact of various masks on stylization process in terms of visual style transfer. The experiments check different masks with regard to content and style preservation. In Experiment 1 (None), a global style transfer occurs without any mask, with a content loss of 23.4 and a style loss of 115.3, and used as a baseline to compare against. Experiment 2 (Center) uses a mask which targets the center of the image without affecting the edges. This approach minimally increases the content loss to 24.1 while the style loss decreases to 110.9, while the periphery of the content is

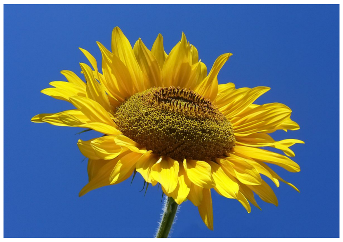

Fig. 1. Content Image

preserved while the center is stylized. Experiment 3 (Face) focuses on the face region, and the content loss is 18.9 while style loss is 105.7 indicating that this mask preserves the face better while applying the style more strongly in other regions. Experiment 4 (Top-half) places a mask on to stylize the top half of the picture down to a content loss of 20.8 and style loss of 120.1. This mask accentuates styling elements such as sky whilst maintaining the lower part of the picture. Experiment 5 finally follows a full-mask approach, which means that, everything in the image gets a stylized form and the average content loss here is 19.8 and average style loss here is 113.6. This experiment evens out the style across the whole image. These experiments show how various mask applications can regulate the trade off between content preservation and style transfer in diverse areas of the image.

TABLE I
EFFECT OF DIFFERENT MASKS ON STYLIZATION

| Exp. | Mask | Content ↓ | Style ↓ | Notes |
|---|---|---|---|---|
| 1 | None | 23.4 | 115.3 | Global style |
| 2 | Center | 24.1 | 110.9 | Edges preserved |
| 3 | Face | 18.9 | 105.7 | Focused style |
| 4 | Top-half | 20.8 | 120.1 | Sky only styled |
| 5 | White | 19.8 | 113.6 | Full-style again |

## RESULTS AND DISCUSSION

Our project end result illustrates how neural style transfer reaches success by utilizing selective masking features. The model transforms a source picture by keeping its basic structure intact then applies artistic design elements from the style reference to create a new visual depiction. An additional mask image establishes precise control of stylization in specific areas thus protecting all content image elements outside of these targeted regions from change. The processed photos in visual form showcase how the model successfully maintains important content elements and structural borders and layout and adds stylistic attributes within masked image areas. During an all-white mask condition the style transfer process applies the style to every part of the content image leading to whole-image stylization. The style transformation effect of the model occurs only within the regions highlighted by the mask image which allows users to guide its artistic modifications. The system achieves alternative levels of user flexibility because it operates beyond traditional standard style transfer systems.

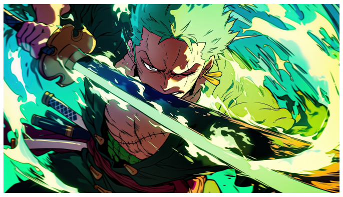

Fig. 2. Style image

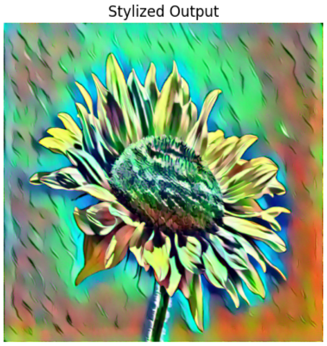

Fig. 3. Generated Image

Technical benefits emerge from implementing VGG16 as a pretrained model because it performs efficient feature extraction across multiple layers. Basic textures along with edges appear in lower layers of the system and deeper layers contain more abstract information about content. The model acquires a balanced style-content fusion through its integration of multiple feature levels. The optimization-based approach uses resources intensively but achieves top-quality output results through multiple refining steps. The results obtain aesthetic appeal and user-intent consistency by using proper style-to-content loss ratio settings and choosing appropriate layers. .

## CONCLUSION

This project succeeded in performing Neural Style Transfer through VGG-16 model implementation that fused the style of one image into the content of another. Different convolutional network layers allowed our system to extract both feature representations which we used to separate content from style then recombined them effectively. The design which used one content layer along with several style layers produced images that maintained content structure but displayed reference image stylistic patterns effectively. Deep learning successfully completes intricate artistic transformations through a program-

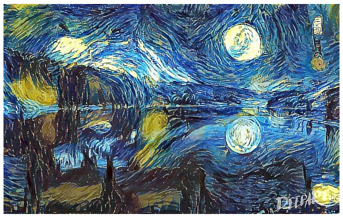

Fig. 4. Another Output

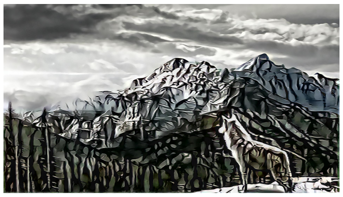

Fig. 5. Masking on wolf and mountains

ming system that looks beyond conventional rule-based methods. The developed project displays neural network processing capabilities in image generation tasks and creates potential possibilities including real-time style transfer systems and personalizable filters and AI-powered art-making tools. The next stage of research should combine optimization methods for enhancing speed while adopting transformer-based models for boosted stylization output.

## References

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv preprint* arXiv:1508.06576, 2015. [Online]. Available: https://arxiv.org/abs/1508.06576

[2] F. Johnson and P. Jain, "Image style transfer using convolutional neural networks," *J. Image Proc. Comput. Vis.*, vol. 9, no. 3, pp. 134–140, Mar. 2017.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint* arXiv:1409.1556, 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[4] R. Zhang, "An improved approach to neural style transfer using region masks," unpublished.

[5] A. Vedaldi and K. Lenc, "MatConvNet – convolutional neural networks for MATLAB," in *Proc. ACM Int. Conf. Multimedia*, 2015, pp. 689–692.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.

[7] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.

[8] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, "Recurrent neural network-based language model," in *Interspeech*, 2010, pp. 1045–1048.

[9] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, 2016, pp. 694–711.

[10] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint* arXiv:1607.08022, 2016. [Online]. Available: https://arxiv.org/abs/1607.08022