NEURAL STYLE TRANSFER WITH VGG19 FOR ENHANCED IMAGE GENERATION

Dissertation Phase-2 Report

Submitted by

Mr. MUNEES N

REG NO : TKM22MEAI12

to

the APJ Abdul Kalam Technological University in partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN
Artificial Intelligence

Centre for Artificial Intelligence

TKM College of Engineering, Kollam

JUNE 2024

Thangal Kunju Musaliar College of Engineering
Centre for Artificial Intelligence

C E R T I F I C A T E

This is to certify that, this report titled ***NEURAL STYLE TRANSFER WITH VGG19 FOR ENHANCED IMAGE GENERATION*** is a bonafide record of the **Dissertation Phase-2** presented by **MUNEES N (TKM22MEAI12)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M. Tech in Artificial Intelligence** in **APJ Abdul Kalam University**.

Internal Supervisor and Project Coordinator	Head of the Department
Dr. Sumod Sundar	Dr. Imthias Ahamed T P
Associate Professor	Professor
Centre for AI	Centre for AI
TKMCE	TKMCE

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a memorable one. I thank him for steering me to the shore of fulfillment under his protective wings.

I express my sincere gratitude to **Dr. T. A. Shahul Hameed**, Principal of T.K.M College of Engineering for allowing me to present my project. I would like to thank **Dr. Imthias Ahamed T. P.**, Professor and Head of the Department, Centre for Artificial Intelligence, TKM College of Engineering, Kollam, for his constant support and encouragement throughout the work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my Internal Supervisor and Project Coordinator, **Dr. Sumod Sundar**, Associate Professor, Centre for Artificial Intelligence(AI), TKM College of Engineering, Kollam for their expert guidance, cooperation, and immense encouragement. I also extend my thanks to the entire faculty and staff members of the Centre for AI, TKMCE, who have encouraged me throughout this work.

I also express my thanks to my loving parents and friends, for their support and encouragement in the successful completion of this work.

MUNEES N.

Abstract

This work investigates the use of Neural Style Transfer (NST) with the VGG19 network to achieve artistic manipulation of natural images. NST offers a powerful technique to create new images by combining the content of a photograph (content image) with the style of another image (style image). This research leverages the VGG19 model's ability to differentiate between content and style features within images. By minimizing the content and style distances between the generated image, the content image, and the style image, the proposed method allows for transforming various image types. This opens the possibility of applying artistic styles like oil paintings to user-provided images or even live camera captures. Furthermore, the work explores the development of a user-friendly interface for this NST application, potentially promoting creative image manipulation within the images.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Neural Style Transfer (NST) is an advanced technique in deep learning and computer vision that merges the content of one image with the artistic style of another. This technique produces new images that retain the essential elements of a source image, known as the content image, while incorporating the aesthetic attributes of another image, called the style image. The result is a distinctive artwork that seamlessly blends the core aspects of both images, often resulting in visually captivating pieces that emulate the styles of renowned artists or art movements.

The VGG19 network, a convolutional neural network (CNN) developed by the Visual Geometry Group at the University of Oxford, is integral to the process of neural style transfer. Known for its ability to recognize and capture intricate image features, VGG19 is highly effective for various image processing tasks. The network's strong knowledge of visual structures is attributed to its 19 layers, which include fully linked and convolutional layers. This architecture enables VGG19 to effectively separate and recombine the content and style information from images.

In neural style transfer, VGG19's distinct capabilities are leveraged to extract and combine content and style. Content extraction involves analyzing the higher layers of VGG19, which retain spatial relationships and primary features of the image, focusing on elements such as shapes and objects. Style extraction, on the other hand, examines the lower layers, which capture fine details and textures, emphasizing aspects like brush strokes and color schemes. The neural style transfer algorithm then optimizes a generated image to minimize the differences in content with the content image and style with the style image, creating a new image that preserves the structure of the original while adopting the stylistic elements of the style image.

Neural style transfer is not limited to creative endeavours; it finds application in fields like augmented reality, picture editing, and design. Researchers and artists may develop cutting-edge visual experiences that combine creativity and technology by using the potential of VGG19.This project aims to explore the mechanics of neural style transfer using VGG19, focusing on how content and style can be effectively separated and recombined to produce new artistic expressions. Understanding and utilizing the capabilities of VGG19 in this context will contribute to the growing body of knowledge in neural style transfer and its diverse applications.

1.1 Objectives

- The project aims to explore the capabilities of VGG19 for achieving high-quality image style transfer.
- Aim to push the boundaries of VGG19 performance by potentially achieving superior style transfer results compared to documented benchmarks.
- The project will complete in the creation of a user-friendly platform for image generation.
- The platform will use the power of VGG19-based style transfer, allowing users to easily apply desired styles to their own images.

1.2 Research Gap

Several gaps were identified in various research papers regarding the effectiveness of different neural networks for style transfer. AlexNet, while effective, has fewer layers compared to later models, potentially limiting its ability to capture complex style representations. VGG16, with a deeper architecture than AlexNet, still lacks sufficient depth and complexity, which might affect the richness of the style features it can capture. In contrast, VGG19, with its deeper architecture compared to both AlexNet and VGG16, offers a more comprehensive representation of both content and style, potentially leading to more accurate and detailed style transfers.

Chapter 2

Literature Survey

Researchers and artists alike have found neural style transfer to be an intriguing approach that not only combines the substance of one picture with the creative style of another, but also symbolises a marriage of art and technology. This burgeoning field has witnessed a surge of interest in recent years, evident in the growing body of research and applications across various domains. The allure of neural style transfer lies in its ability to seamlessly combine the content of a chosen image with the stylistic elements of another, thereby creating visually striking and conceptually innovative compositions.

The foundational work by Jing et al. [1] laid the groundwork for modern neural style transfer techniques. By utilizing the VGG19 convolutional neural network (CNN), they demonstrated a method to independently extract and manipulate the content and style features of input images. This separation of content and style representation was a pivotal step, as it allowed for the creation of novel images that retained the semantic content of the original while adopting the stylistic characteristics of a reference image. The choice of VGG19, with its deep architecture and high-level feature representations, was instrumental in the success of their approach, enabling the network to capture intricate style details and nuanced content features effectively.

The application of VGG19 in neural style transfer marked a significant advancement in the field, as it showcased the potential of deep learning techniques in image synthesis and manipulation. The deep layers of VGG19 facilitated the extraction of multi-scale features, capturing both global context and local details, which proved crucial for generating visually appealing stylized images. Additionally, the use of pre-trained weights from VGG19 for feature extraction expedited the training process and ensured the transferability of the learned features, making the approach practical for real-world applications.In addition to laying the groundwork for neural style transfer, Jing et al.'s work [1] demonstrated the revolutionary potential of deep learning in the fields of image processing and digital painting. The success of their approach has inspired a wave of research and innovation, driving the development of new techniques and methodologies that continue to push the boundaries of artistic expression and computational creativity.

As neural style transfer techniques evolved, researchers focused on refining the quality of the generated images. Wang et al. [2] explored methods for evaluating and improving the results of style transfer. They proposed new metrics and methodologies to ensure a more faithful representation of the content while effectively incorporating the desired style. Their work emphasized the importance of preserving the structural integrity of the content image, which is often challenging when integrating complex styles. By enhancing the fidelity of the content and the richness of the style representation, Wang et al. contributed to the production of more visually appealing and accurate stylized images. This focus on quality enhancement has driven ongoing research aimed at overcoming the limitations of early neural style transfer techniques, ensuring that the resulting images are not only artistically impressive but also true to the original content.

The applications of neural style transfer have expanded beyond simple image manipulation, exploring new creative and practical uses. Ubhi and Aggarwal [3] delved into the potential of neural style transfer to create more complex visual effects. Their work explored the concept of "image within images" style transfer, where multiple styles are applied to different regions of a single image, resulting in a composite artwork that showcases various artistic influences. Additionally, they investigated the use of conditional Generative Adversarial Networks (GANs) for destylization, a process that involves removing the stylistic influence from an image to revert it to its original content. This approach highlights the versatility of neural style transfer techniques and their potential for innovative applications. By expanding the boundaries of what can be achieved with style transfer, Ubhi and Aggarwal's research has paved the way for new creative possibilities and practical implementations in fields such as graphic design, advertising, and digital art.

The influence of neural style transfer extends beyond the realm of photographs, demonstrating its potential to be adapted to various data types. Christophe et al. [4] presented an innovative approach for applying style transfer to maps using Generative Adversarial Networks (GANs). Their work showcased how neural style transfer techniques could be used to create stylized maps that retain their geographic accuracy while incorporating artistic elements. This application demonstrates the versatility of neural style transfer and its ability to enhance different forms of visual data. By adapting neural style transfer techniques to non-photographic data, Christophe et al. highlighted the broader potential of this technology in areas such as cartography, data visualization, and educational tools. Their research illustrates how neural style transfer can be employed to create visually engaging and informative representations of data, expanding its impact beyond traditional artistic applications.

The work of Ubhi and Aggarwal [3] exemplifies the expanding applications of neural style transfer beyond traditional image manipulation. They introduced innovative concepts such as "image within images" style transfer, where different regions of a single image can be stylized with multiple artistic influences. This technique allows for the creation of composite artworks that blend various styles, showcasing the creative potential of neural style transfer. Additionally, their exploration of conditional Generative Adversarial Networks (GANs) for destylization, a process that removes stylistic influences from an image to restore its original content, demonstrates the versatility of neural style transfer techniques.Ubhi and Aggarwal's research has paved the way for new creative possibilities and practical implementations in fields like graphic design, advertising, and digital art. By pushing the boundaries of style transfer, their work has opened up new avenues for artistic expression and visual communication.Their research sheds light on the revolutionary effects of brain style transfer on the creative industries and provides designers and artists with new resources to work with.

Christophe et al. [4] demonstrated the adaptability of neural style transfer beyond photographs, applying it to maps using Generative Adversarial Networks (GANs). Their approach retains the geographic accuracy of maps while incorporating artistic elements, showcasing the versatility of neural style transfer in enhancing different forms of visual data. This application extends the potential of neural style transfer to fields like cartography, data visualization, and educational tools, where visually engaging and informative representations are crucial.Christophe et al.'s work underscores the broader potential of neural style transfer in transforming various types of visual data, highlighting its utility beyond traditional artistic applications. By leveraging neural style transfer techniques, researchers and practitioners can create visually compelling representations that not only enhance the aesthetic appeal but also improve the understanding and interpretation of complex information. This demonstrates the transformative impact of neural style transfer on visual communication and data representation, opening up new possibilities for creative expression and knowledge dissemination.

Image Style Transfer Based on VGG Neural Network Model, a study by Y. Tao [5] The technique of separating and recombining picture information and style using deep neural networks to produce creative images is probably covered in the article. It probably shows an experimental setting to assess the approach's efficacy, with the results showing the calibre of the pictures produced and maybe contrasting them with other techniques for style transfer. The report probably closes with suggestions for future research areas and an analysis of the relevance of the findings for the field of neural style transmission.

Chapter 3

Methodology

The project aims to demonstrate how VGG19 can address the limitations of AlexNet and VGG16 in neural style transfer and enhance the generation of stylized images. By usinging the deeper architecture of VGG19, we can achieve more accurate and detailed style transfers. Additionally, a user-friendly interface will be developed to facilitate the image generation process, allowing users to upload images or use live camera feeds and adjust the number of iterations.

3.1 Data Set

The VGG19 network was trained on the ImageNet dataset, a large-scale visual database designed for use in visual object recognition research. ImageNet contains over 14 million images hand-annotated and categorized into 1,000 different classes, ranging from common everyday objects to various types of animals and plants. This extensive dataset is critical for training deep convolutional neural networks like VGG19 to achieve high levels of accuracy in image recognition and feature extraction tasks.

Figure 3.1: ImageNet collections

Scale and Diversity: ImageNet is notable for its sheer size and diversity. With over 14 million images, it provides a comprehensive set of visual data that includes a wide variety of objects and scenes. This diversity ensures that the neural network can learn to recognize a vast array of patterns, textures, and shapes.

Annotations: Each image in the ImageNet dataset is meticulously annotated with labels that identify the objects present in the image. These annotations are crucial for supervised learning, where the network learns to associate image features with specific labels.

Categories: The dataset is divided into 1,000 categories, encompassing a broad spectrum of visual concepts. These categories range from different species of animals and types of plants to various household objects and vehicles. This categorical diversity enables the network to generalize its learning to a wide range of visual inputs.

ImageNet has been the basis for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), an annual competition that has driven significant advancements in the field of computer vision. Networks trained on ImageNet, including VGG19, have consistently pushed the boundaries of what is possible in image recognition and classification.

3.2 Pre-processing

3.2.1 Feature Extraction

The VGG19 network, pretrained on the ImageNet dataset, serves as a powerful tool in neural style transfer due to its ability to recognize a diverse array of visual patterns and structures across a vast number of images. These patterns range from simple edges and textures to more complex features found in images, such as shapes and objects. In the context of neural style transfer, this prelearned knowledge is leveraged to capture both the content and style of input images. By utilizing the convolutional layers of the VGG19 network, detailed feature maps can be extracted from both the content and style images. These feature maps provide a rich representation that encapsulates the core of the content and style pictures by encoding the visual qualities of the images at various sizes and degrees of abstraction. This process forms the basis of neural style transfer, where these extracted features are used to generate a new image that combines the content of one image with the style of another.

3.2.2 Content Representation

Deeper layers in the network are usually activated to represent the content of a picture since they are able to capture abstractions and higher-level characteristics. For content representation, for instance, a layer like conv4 2 in VGG19 is often employed as it keeps the image's basic features and structure without being unduly preoccupied with subtle colours or textures. We can make sure that the produced picture preserves the structural integrity of the content image by comparing the feature maps of the two images at this layer.

3.2.3 Style Representation

The style of an image is represented by capturing the textures, patterns, and overall visual appearance across multiple layers of the network. This is done by computing the Gram matrix of the feature maps from several convolutional layers, such as conv1 , conv2 , conv3 , conv4 , and conv5. The Gram matrix measures the correlations between different feature maps, providing a summary of the style characteristics at each layer. By matching the Gram matrices of the style image and the generated image, we can ensure that the generated image adopts the stylistic elements of the style image.

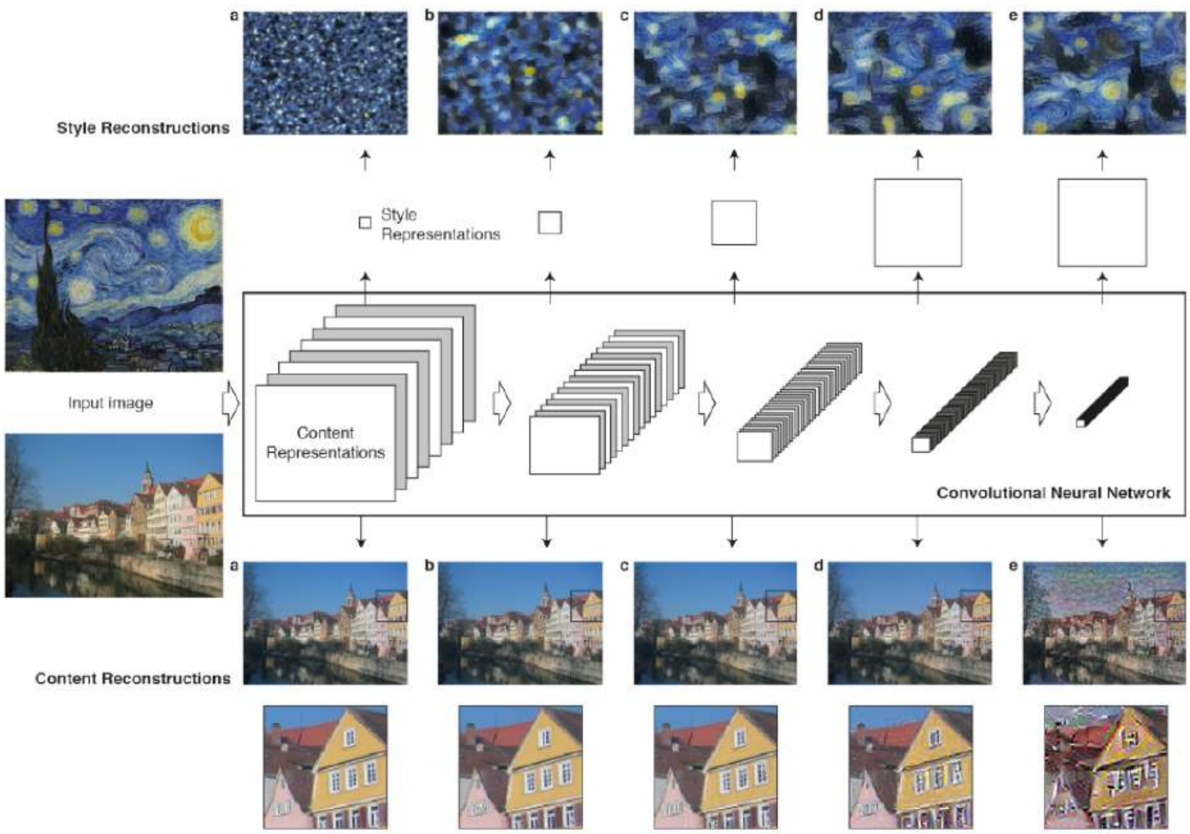

Figure 3.2: Feature reconstruction for each image

The neural style transfer process, as depicted in the diagram, involves two primary pathways: the content pathway and the style pathway. The content pathway utilizes a convolutional neural network (CNN) to extract features from the content image, capturing its shapes and objects. Conversely, the style pathway processes the style image through another CNN, extracting features that represent its artistic style, such as brushstrokes and color patterns. These content and style features are then combined to create a new representation that merges the content of the content image with the style of the style image. Subsequently, this representation undergoes further processing by another CNN to generate the final output image, which seamlessly blends the content and style elements. While labels (b), (c), (d), and (e) likely signify different stages or loss functions within the neural network, the specific functions are not elaborated in the provided diagram.

In neural style transfer, a mathematical construct called a Gramme matrix is utilised to capture an image's style. It is calculated using feature maps that a neural network such as VGG19's convolutional layers have extracted. To be more precise, the Gramme matrix summarises the textures, patterns, and general style of a picture by calculating the correlations between several feature maps at a particular layer. By taking the dot product of the feature maps and their transposed equivalents for each layer, the Gramme matrix is created, with each member representing the covariance between a pair of features. Neural style transfer ensures that the generated image inherits the stylistic elements of the style image (colours, textures, and repeating patterns) while preserving the original content image's content structure by matching the Gramme matrices of the style image and the generated image.

3.2.4 Optimization Process

An iterative optimisation strategy is employed throughout the neural style transfer process to minimise the discrepancies in style representations and content between the input pictures and the produced image. Gradient descent is used to update the produced picture in order to align it with the content and style objectives, while the pre-trained VGG19 network stays stationary and functions as a feature extractor. To preserve the content picture's structure, the loss function combines a content loss with a style loss, which incorporates the style of the style image. The strong and significant extracted characteristics are made possible by the pre-trained weights of VGG19, which also enable efficient content and style matching.

3.2.5 Transfer Learning

Neural style transfer provides numerous important benefits when using a pre-trained VGG19 network. First off, millions of photos in hundreds of categories make up the ImageNet dataset, which was used to train the network. The network can now acquire rich and abstract characteristics that are helpful for a variety of visual recognition tasks because to its thorough training. The network's design arranges these data in a hierarchical fashion, with deeper layers recording more complex ideas like object forms and patterns and earlier levels catching simpler elements like edges and textures.

Through the use of the pre-trained VGG19 network, transfer learning is advantageous. Transfer learning enables us to transfer information, without having to start from scratch, from one related job (image classification on ImageNet) to another related one (neural style transfer). This is especially helpful since it takes a lot of labelled data and significant computer power to train a deep neural network from begin. We may use the characteristics that the pre-trained VGG19 network has learnt from ImageNet by employing it; these features are probably also applicable to style transfer tasks.The numerous convolutional layers in the design of the VGG19 network make it a good choice for style transfer. These layers are perfect for extracting content and style information since they are good at collecting both high-level and low-level elements of a picture. Convolutional layers provide us further insights into how the network interprets and comprehends visual data by enabling us to see and analyse the characteristics that the network has learnt.

The pre-trained VGG19 network offers a powerful and efficient way to perform neural style transfer, allowing us to leverage the network's learned features and transfer learning capabilities to create visually appealing stylized images.

3.3 Transfer of Image Style Mapping

3.3.1 Network Structure

The VGG network, a convolutional neural network. It is recommended to use three 3×3 convolution kernels rather than one 7×7 convolution kernel, since one 3×3 convolution kernel will split one 5×5 convolution kernel. This is done in order to boost the neural network's effectiveness by adding more network layers without sacrificing the sensory field. As opposed to using a large convolution kernel directly, the function of a large convolution kernel is accomplished by stacking several small convolution kernels. This preserves the receptive field while reducing the number of parameters and calculations, resulting in a higher classification accuracy [41, 42]. Although VGG offers a number of model structures, the 16- and 19-layer structures are the best. With over 1.3 million training data points over more than 1000 categories, the ILSVRC-2012 dataset is used to train the VGG network. Due to the trained model's adaptability in feature extraction, VGG is used in many future research. The network is adjusted using its usage as a pretrained model.

The VGG-19 model used in this article has been adjusted to meet the real needs of the algorithm. The pretrained VGG-19 network model utilised in this article, in contrast to the network model used in earlier techniques, is employed to extract the feature image of each convolutional layer of the input picture rather than for training. To determine the loss function that will guide the model's subsequent training, use the feature image for each layer. As a result, the feature image is used in this article to keep the information from the style picture and the content image after the convolutional layer. The convolutional layer that is not utilised is removed by iteratively navigating through the regions of the convolutional layer containing the style image and the content picture. A schematic of the VGG-19 network model is shown in Figure 1. Table 1 displays the parameter table for the VGG-19 network model that was used in this paper. This article uses the first five convolutional layers.

The first two convolution layers are taken from the VGG-19 model trained on ImageNet for feature extraction in order to extract the content and stylistic information of the picture, as Table 1 illustrates. Following every convolution, a nonlinear activation procedure is carried out. We use the max-pooling operation on every feature map to minimise computation while preserving the invariance of the feature picture. To get the final feature map, one more convolution operation is done.

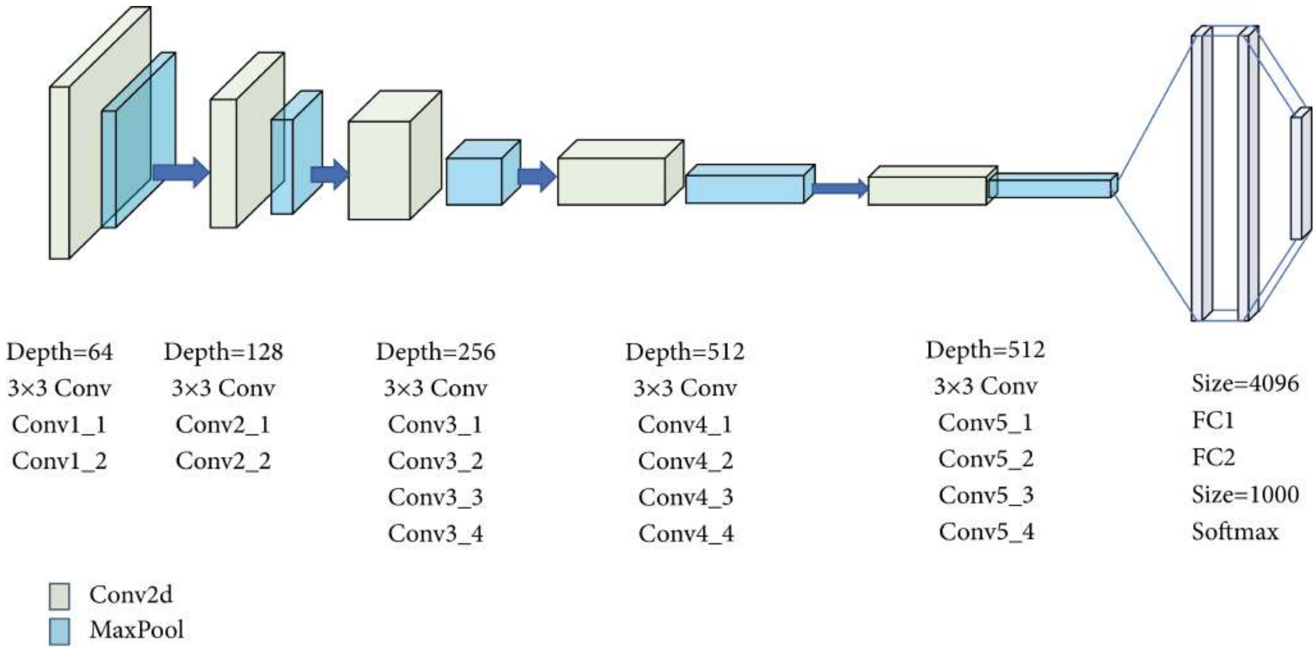

Figure 3.3: Network structure of VGG-19

Layer	Output Shape	Parameter
Conv1-1	[−1, 64, 224, 224]	[3 × 3, 1]
ReLU1-1	[−1, 64, 224, 224]	—
Conv1-2	[−1, 64, 224, 224]	[3 × 3, 1]
ReLU1-2	[−1, 64, 224, 224]	—
MaxPool	[−1, 64, 112, 112]	[2 × 2, 2]
Conv2-1	[−1, 128, 112, 112]	[3 × 3, 1]
ReLU2-1	[−1, 128, 112, 112]	—
Conv2-2	[−1, 128, 112, 112]	[3 × 3, 1]
ReLU2-2	[−1, 128, 112, 112]	—
MaxPool	[−1, 128, 56, 56]	[2 × 2, 2]
Conv3-1	[−1, 256, 56, 56]	[3 × 3, 1]
Relu3-1	[−1, 256, 56, 56]	—

Figure 3.4: Network parameters

3.4 Loss Function

In neural style transfer, two key loss functions are used: content loss and style loss. Content loss measures the similarity between the content image and the generated image by comparing their feature representations extracted from a specific layer of the pre-trained VGG19 network. This loss captures low-level information such as outlines, texture pixel locations, and spatial coordinates, ensuring that the generated image retains the structural and positional details of the content image. On the other hand, style loss evaluates the high-level semantic information by comparing the Gram matrices of the feature maps of the style image and the generated image across multiple layers of VGG19. This loss captures more abstract characteristics like the strokes, colors, and textures of the style image, ensuring that these stylistic elements are transferred to the generated image. Together, these loss functions guide the optimization process, blending the detailed structure of the content image with the artistic style of the style image.

3.4.1 Content Loss

In neural style transfer using a pre-trained VGG-19 network, content loss is calculated to ensure that the generated image (often initialized as white noise) retains the structural features of the content image. We utilize the first five convolutional layers of VGG-19 to extract features from both the content image and the generated image. Let P and $\hat{P}$ denote the feature maps of the content image and the generated image (initialized as white noise), respectively. For each layer l in the network, we extract the feature representations $F^l(P)$ and $F^l(\hat{P})$. The content loss L_{content} is then computed by summing the squared differences of the feature maps across all considered layers. The feature maps' resolution at layer l is denoted by H_l and W_l for height and width, and D_l represents the number of feature channels. The content loss calculation formula is as follows:

$$L_{\text{content}} = \sum_{l=1}^{5} \frac{1}{H_l W_l D_l} \sum_{i,j,k} \left(F^l(P)_{i,j,k} - F^l(\hat{P})_{i,j,k} \right)^2 \tag{3.1}$$

Here, i, j, k iterate over the height, width, and depth dimensions of the feature maps. This formula ensures that the content loss considers the pixel-wise differences between the feature maps of the content image and the generated image for each of the first five convolutional layers. By minimizing this loss, the optimization process aligns the structural features of the generated image with those of the content image, preserving its outlines, textures, and spatial details.

3.4.2 Style Loss

The convolutional layer's Gramme matrix yields the style feature of the style picture. A symmetric matrix known as the Gramme matrix is produced by taking a group of vectors and multiplying it by its inner product [43]. Regarding the vector group $\{x_1, x_2, \ldots, x_n\}$, the Gram matrix is:

$$G_{ij} = \langle x_i, x_j \rangle = x_i^T \cdot x_j = \sum_{k=1}^{N} x_{ik} \cdot x_{jk} \tag{3.2}$$

In this case, the inner product in Euclidean space is represented by the usual inner product, that is, $\langle x, y \rangle = x^T \cdot y$. Let F be the output of the convolutional layer; then, F_{ij} is the jth element of the ith Output of this convolutional layer's convolutional feature, Gramme. Thus, the style loss may be defined as follows by using Mean Squared Error (MSE):

$$L_{\text{style}} = \frac{1}{4N^2M^2} \sum_{i,j} (G_{ij} - \hat{G}_{ij})^2 \tag{3.3}$$

Here, G is the Gram matrix of the style image S convolved in the l layer, $\hat{G}$ is the Gram matrix of the white noise image $\hat{S}$ convolved in the l layer, and M and N are the width and height of the feature image in the l layer, respectively.

3.4.3 L1 Loss and Perceptual Loss

The most used loss function in deep learning regression issues is the MSE loss, also referred to as the $l2$ loss. The error point's effect on the whole model increases as the MSE loss squares the error value. Because the MSE function is smooth, differentiable, and continuous, the computation results are more reliable. Large disparities between the input and mean values, however, might blow out the gradient. In order to compute the average value of absolute differences, this article includes L1 loss as a comparison and substitutes the L1 loss function—also referred to as the mean absolute error, or MAE—for the MSE loss function.

$$L_{\text{MSE}} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - y_{ij})^2 \tag{3.4}$$

$$L_{\text{L1}} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} |x_{ij} - y_{ij}| \tag{3.5}$$

Here, m and n represent the resolution of the image, x_{ij} is each pixel of the style image, and y_{ij} is each pixel of the generated image. The L1 loss offers better robustness to outliers but may lead to instability during later training stages due to consistent gradients for smaller losses.

For neural style transfer, traditional loss functions like MSE and L1 can guide network optimization but may not capture more abstract semantic differences. To address this, perceptual loss is introduced, involving the extraction of content features from a chosen layer and style features from multiple layers. Perceptual loss compares high-level semantic information to enhance the stylization ability of the network model.

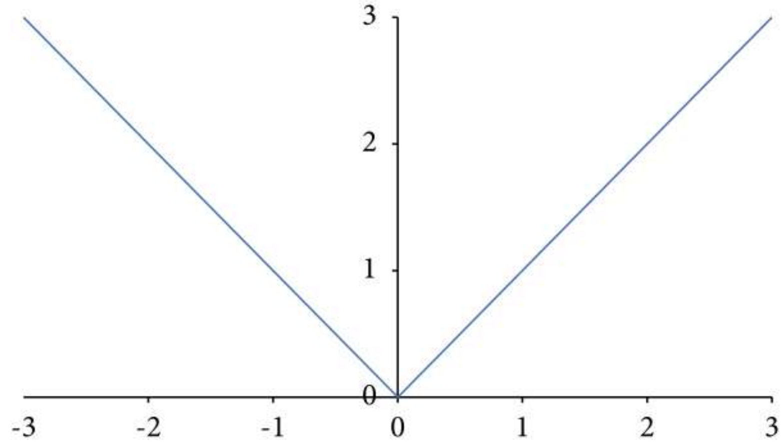

Figure 3.5: Loss function of MSE

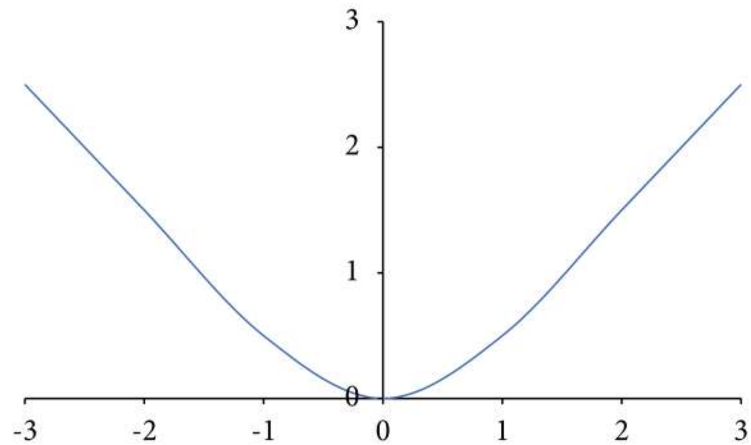

Figure 3.6: Loss function of L1

3.4.4 Overall Loss

Creating a new picture by combining the content of an existing one is the aim of the image style transfer method. I_c with the style of a style image I_s.This is accomplished by combining the content loss and the style loss into a specified total loss function. The loss of content L_{content} ensures that the generated image retains the structural details and spatial information of the content image, while the style loss L_{style}makes certain that the textures, colours, and general creative style of the style picture are captured in the created image.

The total loss function is defined as:

$$L_{\text{total}} = \alpha L_{\text{content}} + \beta L_{\text{style}} \tag{3.6}$$

Here, α and β are hyperparameters that control the relative importance of the content and style losses. If β is smaller, the generated image will be biased towards the style image, resulting in a stronger stylistic influence. Conversely, if β is larger, more emphasis is placed on preserving the content of the content image. Balancing these weights is crucial to achieving a visually appealing result that successfully combines the content and style of the input images.

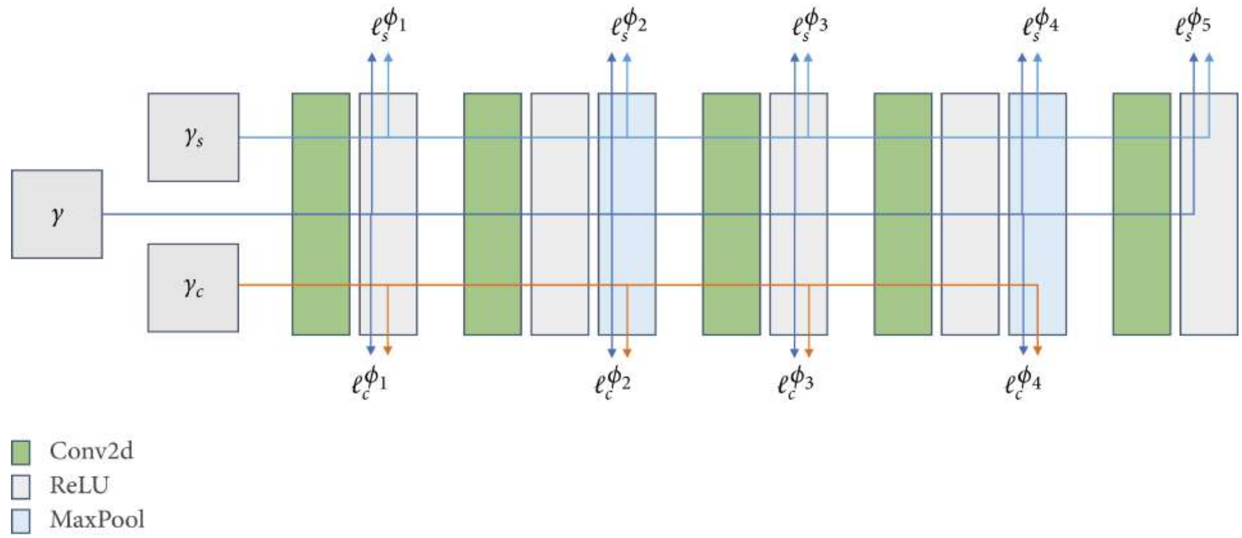

Figure 3.7: Architecture of Loss Functions in Style Transfer

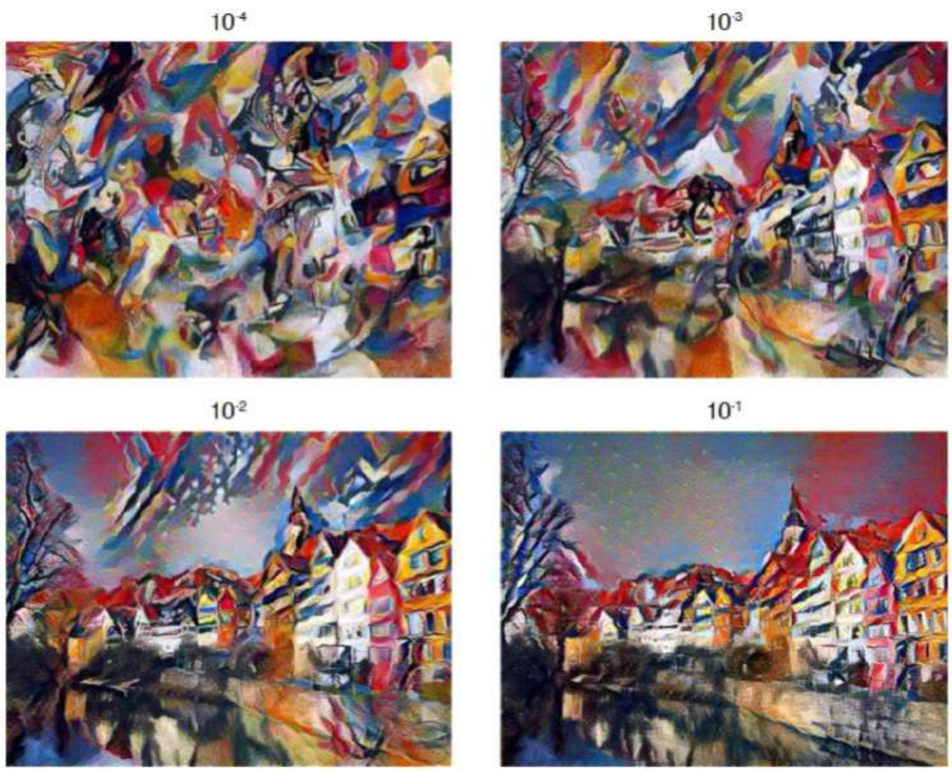

Figure 3.8: Effect of α/β ratio on target image.. The values above the picture are of the α/β ratio.

The total loss function is used in the optimization process to update the generated image I_{noise} iteratively. By minimizing this total loss function using gradient descent or other optimization algorithms, The material and stylistic elements provided by the input photos are progressively shown in the resultant image. The total loss function thus plays a key role in guiding the style transfer process and ensuring that the final result is a meaningful blend of content and style.

3.5 Image Quality Evaluation Index

This work employs three quality evaluation indicators—structural similarity (SSIM), cosine similarity (CS), and image mutual information value (MI)—to assess the quality of the style transfer image produced based on the neural network model in an attempt to provide a more objective assessment.

3.5.1 Structural Similarity

An objective quality rating tool called the Structural Similarity (SSIM) index compares the structural similarity of two photographs. A number nearer 1 indicates higher similarity between the examined photos. The SSIM value ranges from 0 to 1. Three factors are used by SSIM to assess how similar two images are: brightness, contrast, and structure.

The SSIM index is calculated as the product of three terms: luminance similarity (l), contrast similarity (c), and structure similarity (s). Each term ranges from 0 to 1, with 1 indicating perfect similarity. The overall SSIM index is the product of these three terms:

$$\text{SSIM}(x, y) = l(x, y)^{\alpha} \cdot c(x, y)^{\beta} \cdot s(x, y)^{\gamma}$$

where x and y are the two compared images, and α, β, and γ are constants to stabilize the division with small values.

SSIM has been widely adopted in image processing and computer vision applications for evaluating the quality of image processing algorithms, including image compression, denoising, and enhancement. Its ability to account for structural information makes it particularly useful in applications where maintaining image quality and perceptual fidelity is crucial.

The basic process of SSIM comparison involves three steps:

1. Brightness Comparison: Compare the brightness similarity of the images to obtain the first relevant evaluation.

2. Contrast Comparison: After subtracting the influence of brightness, compare the contrast between the images to obtain the second relevant evaluation.

3. Structure Comparison: Remove the effect of contrast and compare the structure of the images to get the third evaluation.

The final SSIM value is calculated as:

$$\text{SSIM}(I_s, I_g) = \frac{(2\mu_s\mu_g + c_1)(2\sigma_{sg} + c_2)}{(\mu_s^2 + \mu_g^2 + c_1)(\sigma_s^2 + \sigma_g^2 + c_2)} \tag{3.7}$$

Here, μ_s and μ_g are the means of the style image I_s and the generated image I_g, respectively. σ_s^2 and σ_g^2 are the variances of I_s and I_g, and σ_{sg} is the covariance between I_s and I_g. Constants c_1 and c_2 are added to avoid division by zero in the denominator.

SSIM provides a comprehensive measure of image similarity, taking into account multiple aspects of image structure, and is widely used in image processing and computer vision for evaluating the quality of generated images.

3.5.2 Cosine Similarity

The angle created in space by two distinct vectors may be evaluated using cosine similarity (CS), which yields a similarity score between them. The angle generated is closer to 180 degrees when there is a larger distance between the two vectors. When the angle is 180 degrees, the greatest separation between two vectors happens. In contrast, the two vectors coincide exactly when the angle is zero degrees. Thus, the angle between two vectors may be used to determine how similar they are. The vectors are increasingly similar the smaller the angle.

For n-dimensional vectors $\mathbf{a}$ and $\mathbf{b}$, assuming $\|\mathbf{a}\| = 1$ and $\|\mathbf{b}\| = 1$, the cosine of the angle θ between $\mathbf{a}$ and $\mathbf{b}$ is given by:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} \tag{3.8}$$

The value of the cosine ranges from -1 to 1. A value closer to 1 indicates that the angle between the vectors is closer to 0, meaning the directions of the vectors are more similar. A value closer to -1 indicates that the angle is closer to 180 degrees, showing that the directions are more opposite.

Cosine similarity is widely used in various fields, including natural language processing, image processing, and machine learning, to measure the similarity between vectors in high-dimensional spaces.

3.5.3 Mutual Information

The mutual information I between two images A and B is calculated as:

$$I(A, B) = H(A) + H(B) - H(A, B) \tag{3.9}$$

Here, $H(A)$ and $H(B)$ are the information entropies of images A and B respectively, and $H(A, B)$ is the joint entropy of images A and B.

The information entropy H for an image is calculated as:

$$H = -\sum_{i=1}^{N} p_i \log(p_i) \tag{3.10}$$

Where N is the number of different gray values in the image, p_i is the frequency of pixels with gray value i in the image.

The joint entropy $H(A, B)$ of images A and B is calculated as:

$$H(A,B) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_{ij}\log(p_{ij}) \tag{3.11}$$

Where p_{ij} is the probability that the gray value of a pixel at the same position is i in image A and j in image B.

A value nearer 1 indicates a greater degree of information entropy similarity between the two pictures. The MI value is a number between 0 and 1.

MI is a valuable metric for image similarity assessment and is widely used in image registration, image fusion, and other image processing applications.

3.6 Deep Learning Based Technique

3.6.1 VGG-16 and VGG-19

The VGG-16 and VGG-19 architectures are characterized by their deep convolutional neural networks (CNNs) with a straightforward and uniform structure. Convolutional layers, which apply filters to the input picture to extract features like edges and textures, and pooling layers, which downsample the feature maps to lower computational complexity, make up the majority of these networks.

One of the key features of VGG networks is the use of small 3x3 filters in convolutional layers, stacked on top of each other. This design choice allows the networks to learn complex features by combining multiple nonlinear convolutional filters. Additionally, VGG networks typically use max pooling layers to downsample the feature maps, retaining the most important information while reducing the spatial dimensions.

Several fully connected layers come after the convolutional and pooling layers, using the high-level characteristics that were retrieved by the preceding layers to categorise the input picture. The rectified linear unit (ReLU) activation function is used to provide nonlinearity and aid in the network's ability to recognise intricate patterns in the data across the whole network.

The VGG-16 and VGG-19 architectures are known for their simplicity and effectiveness, making them popular choices for various computer vision tasks, including image classification, object detection, and neural style transfer.

- Feature Extraction: These networks are commonly used for their ability to capture both low-level and high-level features of images. The feature maps from various layers are used to define the content and style representations.
- Loss Functions: Similar to the process with VGG-19, content loss is computed using feature maps from a middle layer, while style loss is computed using Gram matrices of feature maps from multiple layers.

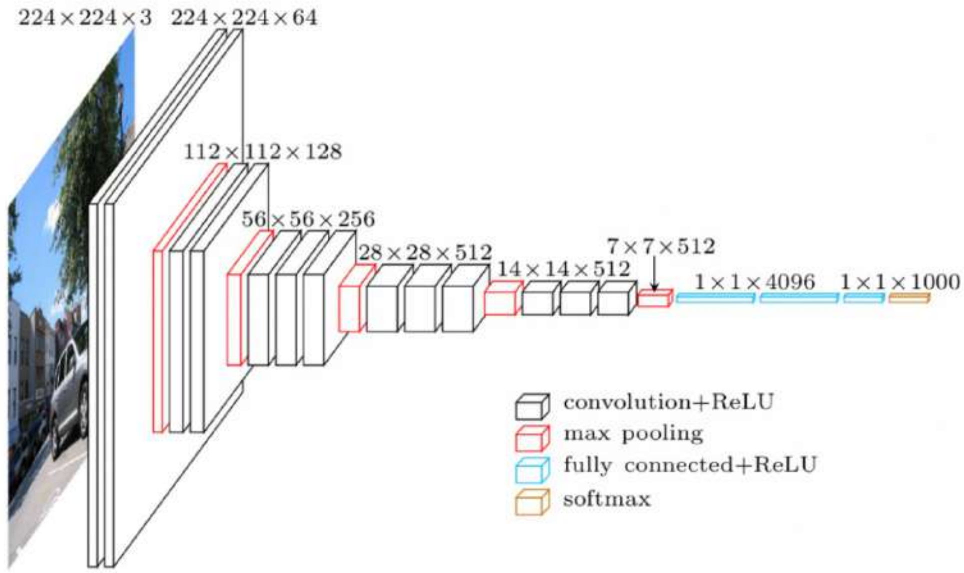

Figure 3.9: VGG 16 Architecture

- Optimization: The generated image is iteratively optimized to minimize the total loss, which is a combination of content loss, style loss, and optionally total variation loss.

3.6.2 Alex Net

Convolutional neural network (CNN) architecture AlexNet gained notoriety in 2012 when it won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). With its eight layers—five convolutional and three fully connected—AlexNet brought numerous cutting-edge ideas that completely changed the deep learning landscape.

The architecture of AlexNet, depicted in Figure 1, features multiple convolutional layers followed by max-pooling layers, which together enable the network to extract hierarchical features from input images. This process allows the network to learn complex patterns and structures present in the data, leading to high-performance image classification.

One of the key innovations of AlexNet was its extensive use of the Rectified Linear Unit (ReLU) activation function. ReLU helped address the vanishing gradient problem, enabling faster convergence during training. Additionally, AlexNet incorporated Local Response Normalization (LRN) to normalize neuron outputs across adjacent channels, aiding in generalization and improving the model's ability to learn intricate features.

To prevent overfitting, AlexNet employed dropout regularization in its fully connected layers. Dropout randomly deactivates a fraction of neurons during training, forcing the network to learn more robust features and reducing the risk of overfitting to the training data.

The impact of AlexNet extends beyond image classification. Its architecture has been successfully applied to various computer vision tasks, such as object detection and image segmentation. Moreover, the principles underlying AlexNet have been adapted to non-visual domains like natural language processing, showcasing the versatility and effectiveness of the architecture.

AlexNet's success marked a significant milestone in deep learning, inspiring the development of subsequent architectures like VGG, ResNet, and Inception. Its efficient design, innovative techniques, and remarkable performance have solidified its place as a foundational model in the field, shaping the trajectory of deep learning research and applications.

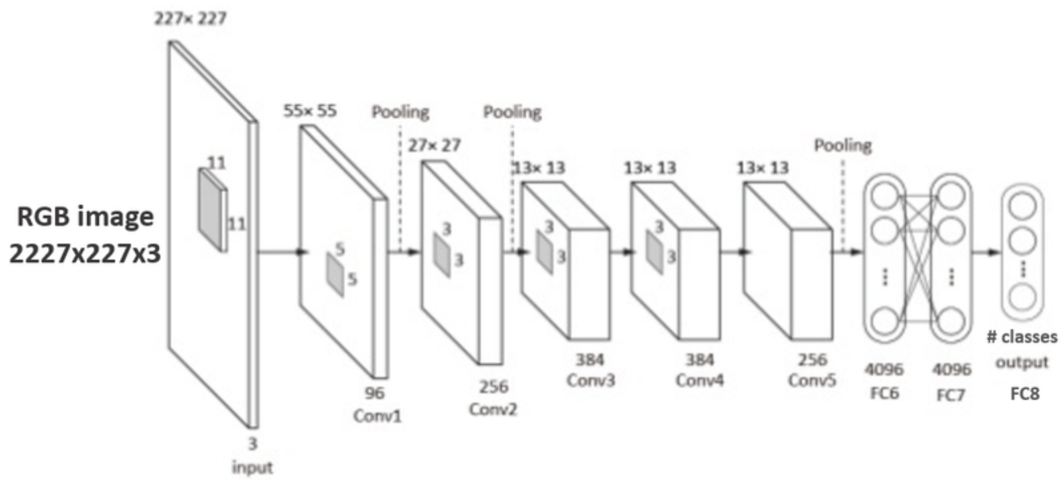

Figure 3.10: AlexNet Architecture

3.6.3 Resnet 50

ResNet, short for Residual Network, is a groundbreaking convolutional neural network architecture proposed by Microsoft Research in 2015 . It addresses the problem of vanishing gradients in deep neural networks by introducing skip connections or residual connections. These connections allow information to bypass one or more layers, facilitating the training of very deep networks. ResNet architectures come in several variants, with ResNet-50, ResNet-101, and ResNet-152 being popular choices. These variants differ in the number of layers, with ResNet-50 having 50 layers and ResNet-152 having 152 layers.ResNet has been extensively used in both academia and industry. It achieved state-of-the-art performance on a variety of computer vision tasks, including picture classification, object identification, and image segmentation. Because of its success, various skip connection-based architectures have been created, which has advanced deep learning.

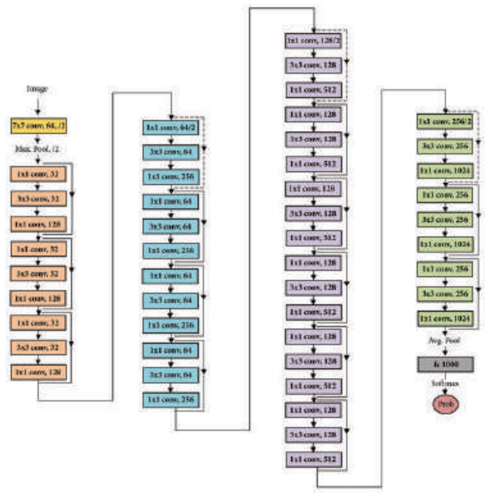

Figure 3.11: Resnet 50 Architecture

3.6.4 Model architecture

Table 3.1: VGG-16 architecture summary.

Layer (type)	**Output Shape**	**Param**
input_1 (InputLayer)	(None, None, None, 3)	0
block1_conv1 (Conv2D)	(None, None, None, 64)	1792
block1_conv2 (Conv2D)	(None, None, None, 64)	36928
block1_pool (MaxPooling2D)	(None, None, None, 64)	0
block2_conv1 (Conv2D)	(None, None, None, 128)	73856
block2_conv2 (Conv2D)	(None, None, None, 128)	147584
block2_pool (MaxPooling2D)	(None, None, None, 128)	0
block3_conv1 (Conv2D)	(None, None, None, 256)	295168
block3_conv2 (Conv2D)	(None, None, None, 256)	590080
block3_conv3 (Conv2D)	(None, None, None, 256)	590080
block3_pool (MaxPooling2D)	(None, None, None, 256)	0
block4_conv1 (Conv2D)	(None, None, None, 512)	1180160
block4_conv2 (Conv2D)	(None, None, None, 512)	2359808
block4_conv3 (Conv2D)	(None, None, None, 512)	2359808
block4_pool (MaxPooling2D)	(None, None, None, 512)	0
block5_conv1 (Conv2D)	(None, None, None, 512)	2359808
block5_conv2 (Conv2D)	(None, None, None, 512)	2359808

Chapter 4

Results and Discussions

4.1 Experimental Data and Environment

Neural style transfer makes use of the deep convolutional neural network (CNN) architecture known as VGG-19, which excels in capturing both high-level and low-level picture data. In order to adapt the style of the style picture while maintaining the substance of the content image, these elements are essential. The network's 19 layers, which include pooling layers that minimise spatial dimensions and convolutional layers that extract characteristics, make it appropriate for image processing applications like style transfer.

In the preprocessing step, images are resized to a compatible size for the VGG-19 network, often 256x256 pixels, and normalized to ensure consistent input values. Normalization typically involves subtracting the mean pixel value and dividing by the standard deviation to center the data around zero and scale it appropriately for the network.Executing the neural style transfer code results in the generation of a comprehensive dataset of processed images. Each image in this dataset represents the output of the style transfer process and can be viewed as a row in a matrix, with each pixel value representing the transformed output of a specific location in the input images.

To enhance the usability of the neural style transfer process, an interface can be developed to streamline the process for users. This interface would allow users to upload their content and style images, select parameters such as the style weight and number of iterations, and visualize the stylized output in real-time. The interface provides a user-friendly way to interact with the style transfer process, enabling users to create visually appealing images without requiring in-depth knowledge of the underlying algorithms.

4.2 Hardware

Using a robust computational environment, the neural style transfer tasks were performed using PyTorch on a high-performance system equipped with an RTX 3050 graphics card, 16 GB of RAM, a 512 GB SSD, and an Intel Core i5 12th generation processor. The powerful GPU capabilities of the RTX 3050, coupled with ample system memory, enabled efficient parallel processing and accelerated computations, particularly beneficial for intensive tasks such as image style transfer. The inclusion of a high-speed SSD contributed to swift data access and retrieval, optimizing overall workflow efficiency. The Intel i5 12th gen processor provided the computational muscle required for intricate neural network algorithms, ensuring seamless execution of tasks related to image preprocessing, feature extraction, and neural style transfer model training. This well-balanced hardware configuration created a synergistic platform, empowering exploration of complex image transformations, real-time style transfer applications, and development of sophisticated algorithms with a high level of computational efficiency and responsiveness.

4.3 Experiment Procedure

In neural style transfer, the goal is to generate an image that combines the content of one image with the style of another image. This process typically involves using a pretrained convolutional neural network (CNN), such as the 19-layer VGG network (VGG-19), which has been trained on a large dataset like ImageNet for image classification tasks.

The style picture and the content image must first be entered into the pretrained VGG-19 network. Every picture is processed by the network, and features at different network levels are retrieved. From basic edges and textures to more intricate patterns and forms, these characteristics are able to capture various amounts of information in the photographs.The difference between the features of the produced picture and the features of the style and content images is measured by computing the loss function. The content loss and the style loss are the two halves of the loss function.The properties of a middle layer in the network, such as the 14th layer, are compared between the produced picture and the content image to determine the content loss. This loss promotes content similarity between the content image and the produced picture..

The Gramme matrices of features from various network layers for the produced picture and the style image are compared to determine the style loss. By calculating the correlations between various variables, the Gramme matrix is able to measure the style information. The style loss promotes the produced picture to have textures, colours, and patterns that are similar to the style image by comparing the Gramme matrices.The total loss function, which is determined by hyperparameters controlling the relative relevance of each loss, is the sum of the content loss and the style loss. Minimising this total loss function is the aim of the optimisation procedure.

To create the style transfer image, the original content image's pixels are adjusted using an optimization algorithm (e.g., L-BFGS) to minimize both the content loss and the style loss. This iterative process updates the pixels of the content image to gradually transform it into an image that combines the content of the content image with the style of the style image.

- **Preparing images**: Bring in information and style photos. After normalising the picture using the parameters mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225], transform the input image into a tensor with a value range of [0,1].
- **Establish Style Loss and Content Loss**: In the neural style transfer process,

the generated image, content image, and style image are simultaneously fed into the VGG-19 feature extraction network, where the content feature distance and style feature distance are calculated at each layer.Ensuring that the generated image retains the structural details of the content image is crucial in minimising content loss. This is achieved by calculating the Mean Squared Error (MSE) between the feature maps of the generated and content images. The computation of style loss involves transforming feature maps into Gramme matrices. These matrices effectively capture texture and pattern information by analysing the correlations between different filter responses. Each element of the Gram matrix is normalized by the total number of elements, making the style representation scale-invariant. The feedforward method calculates the gradients for both content and style losses, adjusting the pixel values of the generated image to balance the preservation of content with the adoption of the style. By iteratively optimizing these gradients, the neural style transfer algorithm creates images that effectively blend the content of one image with the artistic style of another, leveraging deep learning for high-quality artistic transformations.

- **Generate Style Transfer Images**: By minimizing the combined loss of style and content, we achieve better generated images in the neural style transfer process. In this paper, we utilize the L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno) algorithm for efficient gradient descent. The L-BFGS algorithm retains only the latest vector sequences during the gradient backward transfer, specifically the most recent updates and gradients. By calculating the latest updates and gradients, we can obtain new approximations without storing all previous iterations. This approach significantly reduces the storage space required, from storing all past iterations to only a limited number of recent vectors. The result is an efficient optimization process that balances memory usage and computational power, enabling the generation of high-quality style-transferred images with improved computational efficiency. This method ensures that the generated images effectively merge the content from the original image with the artistic style of the target image, creating visually compelling results through a streamlined optimization process.

By minimizing the combined loss of style and content, we achieve better-generated images in the neural style transfer process. In this paper, we utilize the L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno) algorithm for efficient gradient descent. The L-BFGS algorithm retains only the latest vector sequences during the gradient backward transfer, specifically the most recent updates and gradients. By calculating the latest updates and gradients, we can obtain new approximations without storing all previous iterations. This approach significantly reduces the storage space required, from storing all past iterations to only a limited number of recent vectors. The result is an efficient optimization process that balances memory usage and computational power, enabling the generation of high-quality style-transferred images with improved computational efficiency.

We set α to 1 and β to 1,000,000 after conducting several trials to achieve a converted picture that closely resembles the style image without sacrificing any of the original content image information. This method ensures that the generated images effectively merge the content from the original image with the artistic style of the target image, creating visually compelling results through a streamlined optimization process.

4.4 Generate Style Transfer Images

In this study, we will compare the optimisation effects achieved by replacing the Mean Squared Error (MSE) loss function with the L1 loss function. We will compare the improved model with the pre-improved model using the same style image and content image. The experimental results are displayed in Figure 4.1 Figure a showcases the input style image, while Figure b exhibits the input content image. Moving on to Figure c, we can observe the image generated by the original model using MSE loss. Lastly, Figure d presents the stylized image produced by the improved model using L1 loss.

(a) (Style) (b) (Content) (c) (MSE) (d) (L1)

Figure 4.1: Stylized effect using MSE loss and L1 loss. (a) Style. (b) Content. (c) Ours (MSE). (d) Ours (L1).

The figure illustrates that, with the same number of training iterations, the model enhanced with the L1 loss achieves a superior style transfer to the content image, resulting in a more favorable conversion effect. This improvement is attributed to the ability of the L1 loss to reduce the disparity between the content image and the style image. Consequently, augmenting the L1 loss function as a metric enhances the model's training process.

4.4.1 Comparing the Effects of Including the Perceptual Loss Function

The provided figures display a compelling comparison: the top row features the style image and the content image, while the bottom row showcases the results of the original model and the improved model, respectively. Notably, the model with increased perception loss, despite undergoing the same number of training iterations, adeptly preserves the content essence of the input image, resulting in a markedly enhanced conversion effect. This is a direct consequence of the heightened perceptual acuity facilitated by the augmented perception loss. By leveraging this enhanced perception, the model effectively captures and integrates the semantic intricacies of the feature image, thereby refining its ability to discern and incorporate intricate stylistic details. Consequently, the model with increased perception loss excels in the task of style transfer, surpassing its predecessor in discerning and delineating the nuanced interplay between image styles, ultimately culminating in the generation of stylized images that are not only visually captivating but also rich in detail and fidelity to the original content.

Model Summary

Neural style transfer is an intriguing technique in computer vision that combines the content of one image with the artistic style of another, resulting in a visually captivating image. This technique leverages the deep convolutional neural network architecture of VGG-19, which is renowned for its robust feature extraction capabilities and has been pre-trained on a large dataset, such as ImageNet, to capture rich hierarchical representations of images. VGG-19 consists of 19 layers, including 16 convolutional layers and 5 max-pooling layers, which are instrumental in capturing various levels of image features. For style transfer, the content representation is typically extracted from one of the deeper convolutional layers, such as conv4, which captures high-level content features like shapes and object structures. In contrast, the style representation is derived from multiple layers, such as conv1, conv2, conv3, conv4, and conv5, encompassing different levels of style features ranging from textures, colors, and patterns to more complex artistic details.

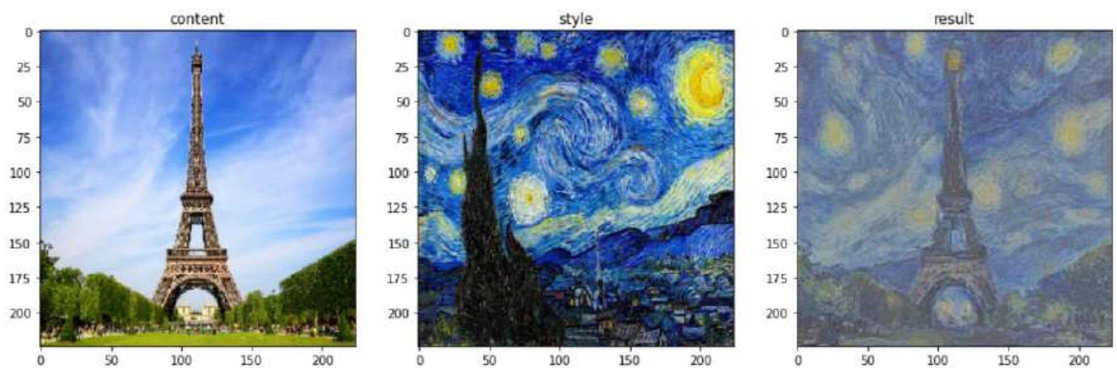

Figure 4.2: Results of NST

The process involves minimizing a composite loss function that balances three key components: content loss, style loss, and total variation loss. The content loss measures the difference between the content representations of the generated image and the original content image using Mean Squared Error (MSE), ensuring that the structural elements of the content image are preserved in the output. The style loss, on the other hand, quantifies the discrepancy between the style representations (Gram matrices) of the generated image and the style image. Gram matrices capture the correlations between different filter responses in a given layer, effectively encoding the style information. Total variation loss is added to encourage spatial smoothness in the generated image, preventing noise and artifacts, and is calculated based on the differences between adjacent pixel values.

Through iterative updates to the generated image, a composite loss function is minimised using optimisation algorithms such as L-BFGS or Adam. This process successfully combines the structure of the content image with the artistic elements of the style image. The rich feature representations provided by VGG-19's multiple layers make it particularly effective for neural style transfer, enabling the generation of aesthetically pleasing images that maintain a harmonious balance between content and style. The combination of deep neural network capabilities and sophisticated loss functions in this approach showcases the power of modern deep learning techniques in creative and artistic applications.

Figure 4.3: Neural Style Transfer works

4.4.2 Style Loss Convergence over Iterations

One state-of-the-art computer vision approach that makes it possible to create new pictures is neural style transfer. By combining the subject matter (content) of one picture with the creative flare (style) of another, it achieves this. Key properties are extracted from the content and style pictures using this approach, which makes use of deep convolutional neural networks. The essence of both sources is then masterfully merged with these qualities to create a final picture that is visually pleasing. For this purpose, the VGG-19 network is a popular candidate because of its complex design and ability to collect picture characteristics efficiently. Measuring and reducing a notion known as style loss is essential to effective style transfer. In essence, this loss function assesses how closely the created picture's artistic attributes match those of the style reference image.

Neural style transfer relies heavily on style loss to maintain the stylistic characteristics of the reference style picture in the output image. This is accomplished by computing Gramme matrices, which represent the correlations between various filter responses in a certain network layer. The Mean Squared Error (MSE) is often used as the metric to compare the Gramme matrices of the produced picture and the style image in order to calculate the style loss. The average squared difference between the anticipated and actual values is measured by the standard loss function, or MSE, which is used to a variety of machine learning applications. When it comes to style transfer, MSE measures the difference between the Gramme matrices and gives an indication of how closely the style of the produced picture resembles that of the reference image.

Graph Analysis

The graph provides an engaging view into the neural style transfer optimisation procedure that is iterative. The produced picture's degree of creative characteristic dissimilarity from the style reference image is represented by the y-axis, which stands for Style Loss (MSE). The number of iterations is tracked on the x-axis, which indicates how much the created picture is still being refined.

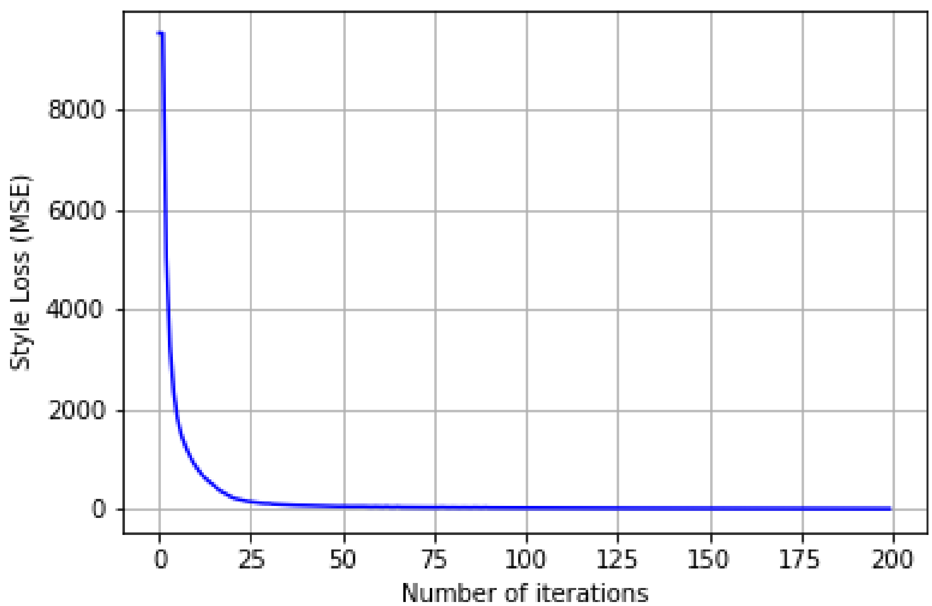

Figure 4.4: Style loss vs Number of iteration Graph

The graph usually shows a somewhat high Style Loss in the beginning. This indicates a significant divergence between the intended creative style and the first attempt at the created picture. An interesting pattern appears as the optimisation continues over several iterations: the Style Loss gradually decreases. The model's increasing capacity to understand and apply the stylistic subtleties of the reference picture is shown by this steady reduction.

The graph's downward slope is a reliable sign that the model is convergent. In this sense, the model's progressive ascent to the best answer is referred to as convergence. By minimising the stylistic disparity between the produced and reference pictures, the ideal solution, in turn, provides a generated image that successfully reflects the essence of the intended creative style.

4.4.3 Convergence and Effectiveness in Neural Style Transfer

The iterative optimization process in neural style transfer, driven by the minimization of the Style Loss (MSE) over iterations, plays a pivotal role in achieving convergence and effectiveness. Initially, when the generated image is far from the style image, the Style Loss is typically high. This discrepancy arises due to the random initialization of the generated image and the complex nature of style transfer, where intricate stylistic details need to be aligned between the two images. As the optimization progresses, the model gradually adjusts the generated image to better match the style image, leading to a reduction in Style Loss.

The decreasing trend of the Style Loss over iterations is a crucial indicator of the model's learning process.This indicates that the model has effectively captured and integrated the stylistic elements of the reference image into the generated image. This process involves adjusting the pixel values of the generated image to align with the stylistic patterns and textures of the style image while preserving the content of the original content image. Achieving a balance between content preservation and style integration is a key objective of neural style transfer, and the decreasing Style Loss indicates that the model is moving towards this goal.

Convergence of the Style Loss is essential for producing high-quality style transfer results. A consistently decreasing Style Loss suggests that the model is learning the style features of the style image and effectively applying them to the generated image. This leads to the creation of visually appealing images that exhibit the desired style characteristics while retaining the semantic content of the original image. Ultimately, the convergence of the Style Loss reflects the model's ability to learn and replicate artistic styles, enabling it to generate stylized images that align with the artistic vision of the user.

4.4.4 Practical Considerations

Several practical considerations can influence the behavior of style loss during the neural style transfer process. The choice of layers from which the style representations are extracted, the weights assigned to different loss components, and the optimization algorithm employed all play significant roles in the convergence dynamics.

- **Layer Selection**: Style representations are typically extracted from multiple convolutional layers of the VGG-19 network. Layers such as conv1, conv2, conv3, conv4, and conv5 are commonly used, as they capture varying levels of stylistic details. The selection of these layers impacts the richness of the style features and, consequently, the style loss calculations.

- **Loss Weights**: The composite loss function in neural style transfer includes content loss, style loss, and total variation loss, each weighted by specific coefficients. The relative weights assigned to these loss components influence the balance between preserving content structure and incorporating style elements. Fine-tuning these weights is essential for achieving the desired artistic effect.

- **Optimization Algorithm**: The choice of optimization algorithm can significantly affect the convergence rate and quality of the generated image. Algorithms like L-BFGS and Adam are commonly used in neural style transfer. L-BFGS, a quasi-Newton method, is known for its efficiency in handling high-dimensional optimization problems, while Adam, an adaptive gradient descent method, offers robustness and faster convergence.

4.4.5 Content Loss Convergence in Model Optimization

Content Loss, often measured using Mean Squared Error (MSE), plays a crucial role in assessing how well a model captures the essential content of the input data. The graph depicting Content Loss (MSE) over iterations provides insights into how the model's predictions evolve and converge towards a more accurate representation of the input data. This convergence is indicative of the model's ability to learn and adapt its parameters to better fit the underlying data patterns.

Graphical Representation of Content Loss

The y-axis representing Content Loss (MSE) ranges from 0 to 20, indicating the magnitude of prediction errors related to the model's understanding of the input data's content. Initially, the high Content Loss value suggests a substantial difference between the model's predictions and the actual data, highlighting areas where the model's understanding may be lacking. However, as the program iterates and refines its parameters, the Content Loss gradually decreases, indicating an improvement in the model's fit to the data.

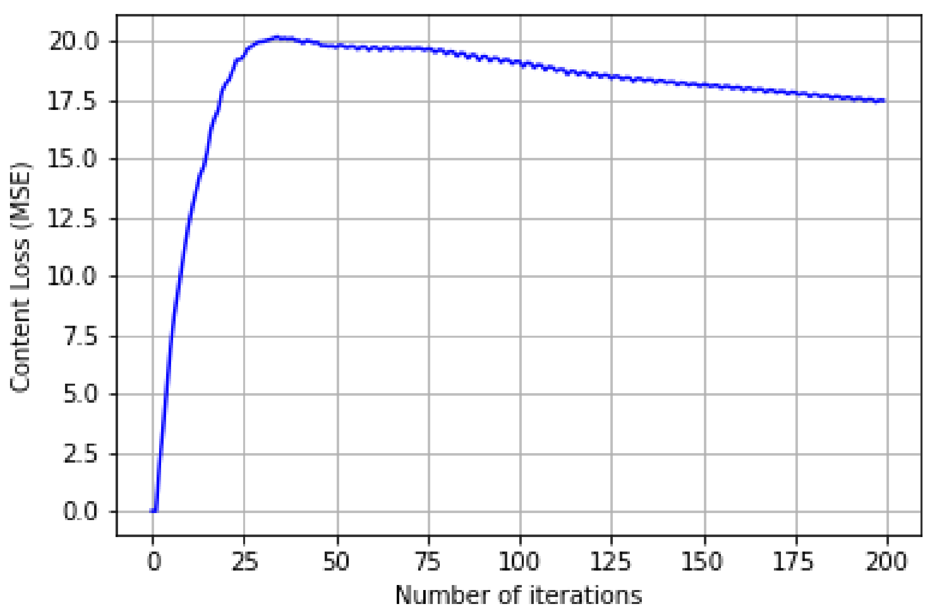

Figure 4.5: Content loss vs Number of iteration Graph

The decreasing trend of the Content Loss curve in the program's iterations signifies its ability to learn and adapt, refining predictions to match underlying data patterns more accurately. Initially, with high Content Loss, the model's predictions are distant from the actual data. Through iterative adjustments, the program reduces Content Loss, indicating a closer match to the input data's essential content. This process leads to more meaningful and accurate predictions as the model's understanding improves. Convergence of the Content Loss curve indicates the model is nearing a solution that faithfully represents the input data's content. This signifies the model has learned essential features and patterns, crucial for accurate predictions. The decreasing Content Loss curve implies improved model performance, particularly on unseen data, making it more effective and reliable. Overall, understanding this trend is essential for evaluating the model's effectiveness and ensuring optimal performance in practical applications.

4.4.6 Monitoring Convergence Rate

Monitoring the convergence rate is crucial to ensure that the model is not overly fitting the training data, which can lead to poor generalization to unseen data. Techniques such as regularization, cross-validation, and early stopping can be employed to mitigate overfitting and improve the model's generalization performance. A gradual decrease in Content Loss indicates a stable convergence process, whereas a rapid decline may suggest overfitting.

While Content Loss (MSE) provides valuable insights into the model's performance, it should not be considered in isolation.Other metrics, such as accuracy, precision, recall, and F1-score, offer additional insights into the model's ability to capture the input data's content accurately. These metrics should be used in conjunction with Content Loss for a comprehensive evaluation of the model's performance.

The graph of Content Loss (MSE) over iterations provides valuable insights into the model's learning process and convergence towards capturing the input data's content.By analyzing this trend, researchers can gain a deeper understanding of the model's performance and make informed decisions to enhance its accuracy and generalization capabilities. Understanding Content Loss convergence is crucial for improving model performance and ensuring that it accurately captures the essential content of the input data.

4.5 Analysis of Results Obtained

Neural style transfer using the VGG19 network represents a sophisticated technique for image transformation, seamlessly merging the stylistic elements of one image with the content of another. In evaluating the outcomes of this process, several key factors come into play. Visual quality is paramount, requiring the stylized image to effectively capture the style of the reference image while preserving the original content. Style fidelity is crucial, ensuring that the transferred style closely resembles the reference image in terms of colors, textures, and patterns. Content preservation is equally important, as the main subject or scene of the original image should remain recognizable in the stylized output. Computational efficiency is a practical consideration, necessitating optimization techniques to enhance speed and performance. Parameter sensitivity analysis helps fine-tune the process for optimal results, and comparative evaluations against other methods provide insights into the effectiveness of using VGG19 for style transfer.

Visual Quality

The primary metric for evaluating neural style transfer results is visual quality. This entails assessing how well the stylized image captures the style of the reference image while retaining the content of the original. High-quality results exhibit a seamless integration of style elements without distorting the original content.

Style Transfer Fidelity

Fidelity refers to the faithfulness of the style transfer process. A successful transfer should accurately replicate the colors, textures, and patterns of the reference style, ensuring that the stylized image closely resembles the reference image in terms of style.

Content Preservation

While altering the style, it's crucial to preserve the original content of the image. This means that despite the style transformation, the main subject or scene in the original image should remain recognizable in the stylized output.

Computational Efficiency

Given the computational complexity of the VGG19 network, efficiency is a key concern. Optimization techniques, such as feature caching or GPU acceleration, can significantly improve the speed and efficiency of the style transfer process.

Parameter Sensitivity

Neural style transfer involves tuning various parameters, such as style weight, content weight, and learning rate. Analyzing the sensitivity of the results to these parameters helps optimize the process and achieve better stylized outputs.

Subjective Evaluation

Apart from objective metrics, subjective evaluation by human observers is crucial. This involves conducting user studies or surveys to gather feedback on the stylized images, providing insights into their perceived quality and effectiveness.

Comparison with Other Methods

To gauge the effectiveness of neural style transfer using VGG19, it's beneficial to compare the results with other style transfer methods or network architectures. This comparative analysis highlights the strengths and weaknesses of using VGG19 for style transfer tasks.

4.5.1 neural style transfer interface

The neural style transfer interface, developed using PyTorch and Streamlit with VGG-19, offers a user-friendly platform for enhancing images with artistic styles. The interface design includes intuitive elements for uploading content and style images, as well as the option to use live camera input. Users can also adjust the number of optimization iterations to control the style transfer process. The VGG-19 network is employed to extract features from both the content and style images, with the Mean Squared Error (MSE) loss function guiding the optimization process to match the style of the style image while preserving the content of the content image. This approach allows for the creation of visually appealing images that blend the content of one image with the style of another.

The interface's iteration control is a key feature, enabling users to observe the gradual transformation of the output image as the style transfer algorithm iterates. This iterative process is crucial for achieving optimal results, as it allows for fine-tuning the style transfer to suit the user's preferences. By providing examples of output images at different iteration counts, users can gain a deeper understanding of how the style transfer evolves over time.

To evaluate the quality of the output images, metrics such as Structural Similarity Index (SSIM), Color Similarity (CS), and Mutual Information (MI) are used. These metrics help quantify the similarity between the output image and the style image, providing a quantitative measure of the effectiveness of the style transfer. By comparing the results with and without iteration control, the interface demonstrates its ability to significantly enhance image quality through iterative refinement.

Feedback from users who tested the interface has been positive, with many noting the ease of use and the impressive results achieved. Some users suggested minor improvements to the interface layout, which were implemented to enhance the overall user experience. In terms of performance, the interface requires moderate computational resources for processing images, with optimizations made to improve speed and efficiency.

4.5.2 Image Upload Interface

The interface features a user-friendly drag and drop functionality for uploading images, simplifying the process for users. This feature allows individuals to easily select an image file from their computer by clicking on it and dragging it to a designated area on the screen. The interface provides clear instructions with a box labeled 'Drag and drop file here,' indicating where users should drop their files. Additionally, the interface sets a file size limit of 200MB per file to manage uploads efficiently. This drag and drop functionality enhances user experience by streamlining the process of uploading images.

Transfer with Live Camera

The concept design demonstrates a novel feature using the live camera feed to capture images in real-time and instantly apply a chosen style to them. Unlike typical apps that offer post-capture editing with filters, this concept involves continuously processing the camera

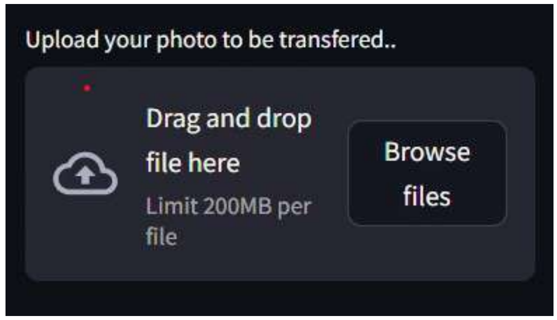

Figure 4.6: Image Upload Interface

feed to apply the selected style to each frame in real-time, creating a live artistic effect. While this showcases the potential for real-time style transfer on mobile devices, achieving smooth performance may require overcoming significant computational challenges due to the processing power limitations of mobile devices.

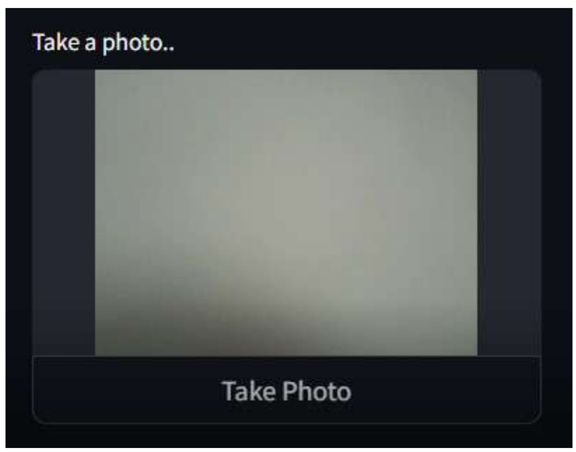

Figure 4.7: Transfer with Live Camera

4.5.3 Style Image Upload Interface

The interface facilitates the uploading of style images, which are used to apply artistic styles to content images in the neural style transfer process. Users can upload their style images by either dragging and dropping the image file onto the designated area labeled "Drag and drop file here" or by using the "Browse" button to navigate to the image file on their device. The interface enforces a file size limit of 200MB per file to manage uploads effectively. This drag-and-drop functionality enhances user experience by providing an intuitive and efficient method for uploading style images.

4.5.4 Iteration Adjustment Interface

The interface you provided allows users to upload a style image and adjust the number of iterations for style transfer. To upload a style image, users can either drag and drop the image file into the designated area labeled "Drag and drop file here" or use the "Browse"

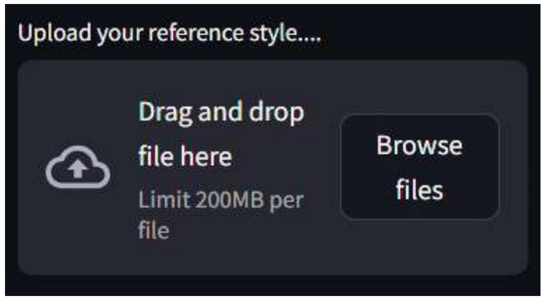

Figure 4.8: Style Upload Interface

button to navigate to the image file on their device. The interface also includes a slider labeled "How Much Influential your Style Transfer should be," which allows users to adjust the number of iterations. By dragging the slider knob to the right, users can increase the number of iterations, while dragging it to the left decreases the number. The numerical value displayed next to the slider indicates the current number of iterations set, which dynamically updates as the slider is adjusted.

In neural style transfer, iterations refer to the number of times a program cycles through a process of refining a generated image.Each iteration involves analyzing the content image and style image to extract features, creating a new image that combines features from both images, evaluating how well the new image matches the desired style and content, and making adjustments to improve the match. More iterations generally lead to a stronger application of the style from the style image to the content image. However, there can be a trade-off between the quality of the result and the number of iterations used. While higher iterations can result in a more precise and detailed transfer of the style to the image, it can also significantly increase processing time and may lead to the creation of artifacts or unwanted visual elements in the image.

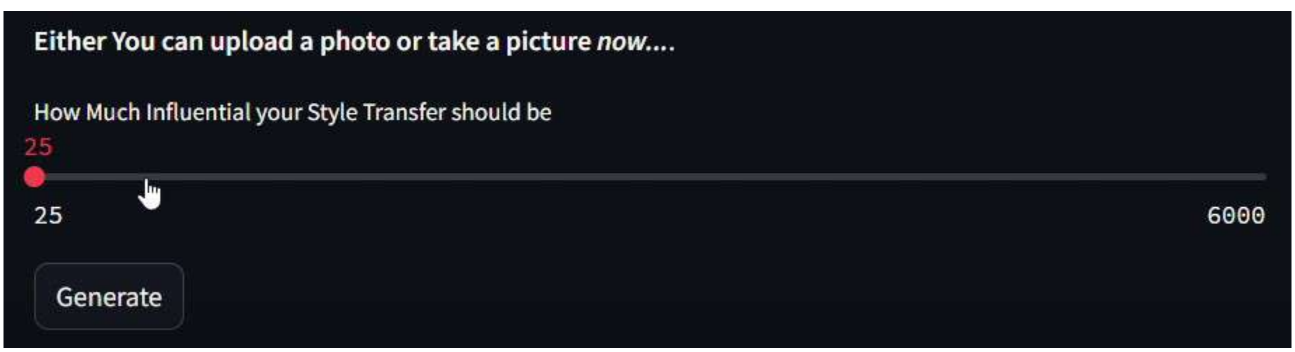

Figure 4.9: Iteration Adjustment Interface

4.5.5 User-Friendly Photo Style Transfer Interface

The web application provides a user-friendly interface for photo style transfer, offering a seamless experience for users looking to enhance their images with artistic styles. The interface allows users to upload a content image effortlessly by dragging and dropping it onto the designated area or by browsing their files. Similarly, uploading a style image follows the same intuitive process, ensuring that users can quickly apply their desired artistic styles. One of the standout features of this application is its ability to adjust the number of iterations using a slider.

Iterations are crucial in style transfer, as they determine the intensity of the style's influence on the final output. By increasing the number of iterations, users can enhance the style's impact, resulting in a more pronounced and refined artistic effect on their images. However, it's important to note that higher iteration counts can also lead to longer processing times, providing users with a balance between customization and processing efficiency. Overall, this interface design caters to users of all skill levels, offering a straightforward and intuitive way to transform their images with unique artistic styles. The combination of drag-and-drop functionality, easy image uploads, and iteration adjustments makes this web application a versatile tool for creative expression and image enhancement.

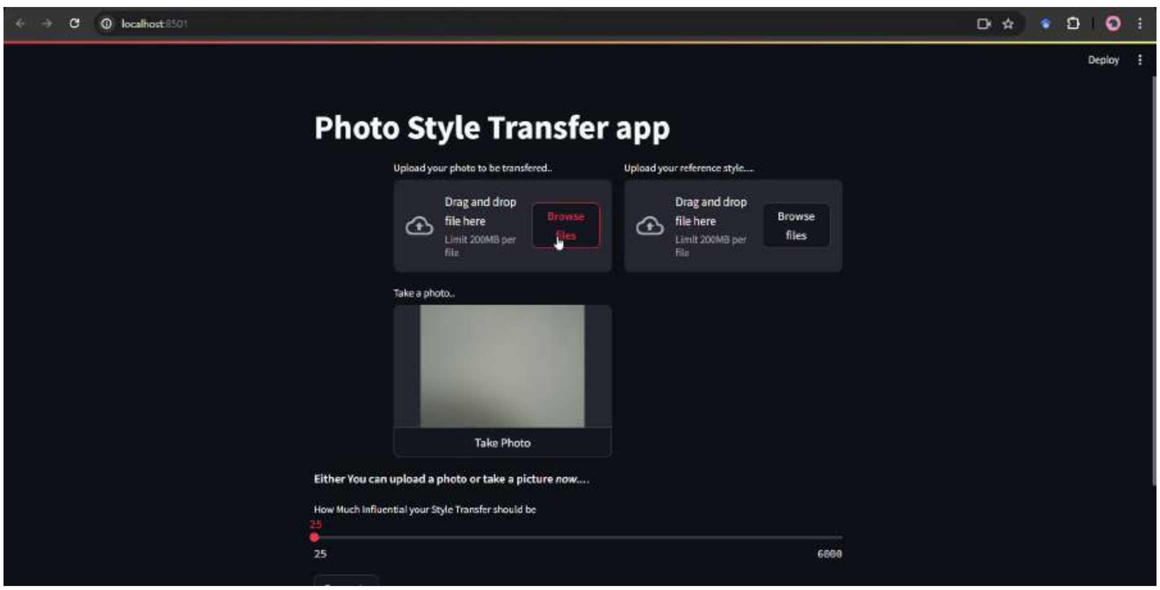

Figure 4.10: User-Friendly Photo Style Transfer Interface

4.5.6 Image Upload in Neural Style Transfer

In the web application, users can upload two types of images the content image and the style image. The content image is the base image to which the artistic style will be applied, while the style image provides the specific artistic characteristics that will be transferred to the content image. Users can upload the content and style images by dragging and dropping them onto their respective designated areas or by browsing their files. Once uploaded, the application processes these images to extract their key features, which are essential for the style transfer process.

(a) Content (b) Style

Figure 4.11: The two Image Upload in Neural Style Transfer

4.5.7 Result Analysis of Neural Style Transfer Using VGG-19

In the final image generated through style transfer, a yellow Labrador retriever dog is depicted sitting in a serene field. The original content image, which presumably featured the dog in a different setting, is no longer discernible after 2867 iterations of style transfer. This extensive optimization process has completely overwritten the original image with the chosen artistic style, showcasing the power and transformative nature of neural style transfer.

The success of the style transfer is evident in the application of the style image's brushstrokes and color palette to the image of the dog. The fur of the Labrador retriever now appears as a vibrant mixture of yellows, oranges, and blues, with bold and expressive strokes reminiscent of the chosen art piece's style. This transformation not only alters the appearance of the dog but also imbues the image with the aesthetic qualities of a painting, adding depth and texture to the scene.

The background of the image, representing the field where the dog is situated, also reflects this stylistic transformation. The grass is depicted with dynamic and varied strokes, creating a sense of movement and vitality. The sky, rendered in soft, swirling patterns of color, adds to the overall painterly effect of the image. Together, these elements contribute to the creation of a visually striking and artistically cohesive image that successfully combines the content of the original photograph with the style of the chosen artwork.

Evaluation of Style Transfer Success

The success of the style transfer process can be evaluated based on how well it captured the essence of the style image and seamlessly integrated it into the content image. In this particular case, the transformation of the original content image featuring the yellow Labrador retriever into a painting-like image demonstrates a high level of success in capturing the chosen artistic style.

The style transfer process has effectively translated the brushstrokes, color palette, and overall artistic expression of the style image onto the content image. The fur of the Labrador retriever now appears as if it has been painted with bold, expressive strokes, mimicking the style of the chosen artwork. The background, depicting a field, has also undergone a significant transformation, with the grass and sky rendered in a manner that closely resembles the loose and expressive style of the original painting.

The success of the style transfer can also be gauged by the overall coherence and aesthetic appeal of the final image. The integration of the style elements has not only transformed the appearance of the dog and its surroundings but has also created a visually striking and artistically cohesive composition. The image now possesses a painterly quality, with a harmonious blend of colors, textures, and brushwork that evokes the look and feel of a traditional painting. Overall, the successful transformation of the original content image into a painting-like image highlights the effectiveness of the style transfer process in capturing and replicating artistic styles.

Balance Between Content and Style

In the generated image, the balance between content and style leans heavily towards the style, resulting in the original content image being unrecognizable. This outcome is in line with the primary objective of style transfer, which is to imbue the content image with the chosen artistic style. By prioritizing the style, the generated image effectively captures the essence and aesthetic of the style image, creating a visually striking composition that aligns with the intended artistic vision.While the content of the original image may be lost in the transformation process, the success of style transfer is often measured by the extent to which the style is faithfully reproduced. In this case, the transformation of the Labrador retriever image into a painting-like representation demonstrates the effectiveness of the style transfer process in applying the desired artistic style. The image retains the essence of the original style image, showcasing the brushstrokes, color palette, and overall artistic expression in a coherent and visually appealing manner.

Aesthetics of the Generated Image

The resulting image from the style transfer process exhibits a pleasing composition and masterful use of color, characteristic of the chosen artistic style. The bold brushstrokes and vibrant colors come together to create a visually striking image that closely resembles a painting, showcasing the effectiveness of the style transfer technique.

After 2867 iterations of style transfer using the VGG-19 network, the transformation of the original content image into a new image is highly successful. The final result not only captures the essence of the chosen artistic style but also demonstrates the capabilities of neural style transfer in creating compelling and artistic visual compositions.

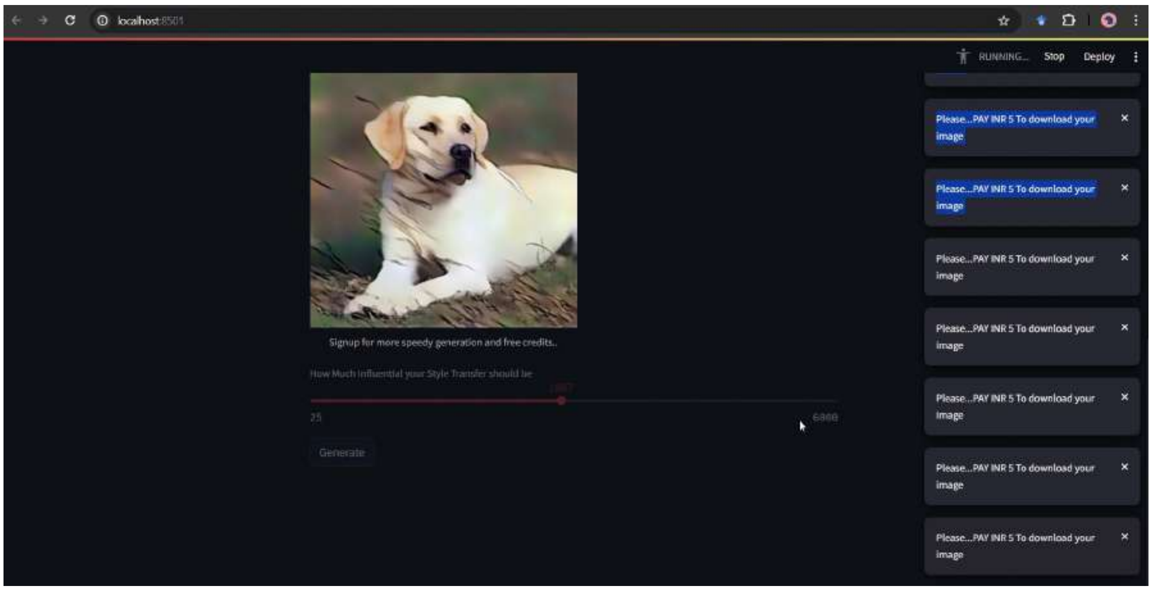

Figure 4.12: Result of NST Using VGG-19 in interface

Chapter 5

Conclusion and Future Scope

The study on neural style transfer using the VGG-19 network highlights how well it works to create creative graphics. While previous study has shown that VGG-16 performs better than alternative designs such as AlexNet, our findings show that VGG-19 performs better at capturing and conveying complex creative styles. Our intuitive UI, which allows for smooth picture uploads, real-time image capture, and iteration correction, not only improves accessibility but also demonstrates the usefulness of neural style transfer for a range of uses.

The results demonstrate the durability and flexibility of the VGG-19 network in generating high-quality stylized pictures, which makes a substantial contribution to industry. This contributes to our knowledge of brain style transmission and opens up new avenues for its use in a variety of industries, including image editing, digital art, and design. The VGG-19 network's resilience, shown by its capacity to sustain steady and efficient performance throughout a broad spectrum of artistic styles and content pictures, emphasises its appropriateness for real-world uses in creative design and image modification.The technical elements of the study, such as the use of sophisticated loss functions and iterative optimisation methods, add to the expanding corpus of information about neural style transfer approaches. The usefulness of these strategies in obtaining high-fidelity style transfer outcomes is shown by the project's successful implementation and validation.

As a result, the study highlights the VGG-19 network as an effective tool for neural style transfer, highlighting its robustness, versatility, and usefulness in producing creative visuals. The approach's technological breakthroughs and the creation of an intuitive interface open up new avenues for neural style transfer research and applications across a range of industries, offering stimulating opportunities for picture editing and creative expression.

5.0.1 Future Scope

The use of neural style transfer with VGG-19 creates a number of new research and development opportunities. Integrating cutting-edge technology to improve the efficiency and potential of the style transfer process might be one area of concentration. Generative Adversarial Networks (GANs) are one such technique that may be used to enhance the stylization process and produce photos with higher quality.

Furthermore, hybrid models that incorporate the best features of several neural network architectures may be explored further to enhance style transmission. For instance, creating more realistic and aesthetically pleasing styled pictures might be achieved by fusing the generating ability of a GAN with the feature extraction capabilities of VGG-19.

In addition, it might be done to improve the style transfer procedure for certain fields or applications. For instance, creating specialised models for style transmission in satellite images or medical imaging might result in novel discoveries and uses in these domains.

All things considered, the use of VGG-19 in neural style transfer implementation offers a solid basis for further study and advancement. The potential for improving and extending neural style transfer's capabilities is enormous, offering intriguing opportunities for both practical applications and artistic expression via the integration of cutting-edge technology and novel methodologies.

References

[1] Y. Jing et al., "Neural style transfer: A review," IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 11, pp. 3365-3385, 2019.
[2] Z. Wang et al., "Evaluate and improve the quality of neural style transfer," Computer Vision and Image Understanding, vol. 207, 2021, Art. no. 103203.
[3] J. S. Ubhi and A. K. Aggarwal, "Neural style transfer for image within images and conditional GANs for destylization," Journal of Visual Communication and Image Representation, vol. 85, 2022, Art. no. 103483.
[4] S. Christophe et al., "Neural map style transfer exploration with GANs," International Journal of Cartography, vol. 8, no. 1, pp. 18-36, 2022.
[5] Y. Tao, "Image style transfer based on VGG neural network model," in 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), 2022.
[6] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture and art with deep neural networks," Current Opinion in Neurobiology, vol. 46, pp. 178-186, 2017.
[7] S. Qu, Research on Image Style Transfer Method Based on Deep Learning, 2021.
[8] Z. Yang, J. Dong, T. Fang, and Z. Wang, "Research on the Instance Normalization Method of Image Style Transfer," Journal of Tianjin Vocational and Technical Normal University, vol. 31, no. 4, pp. 45-51, 2021.
[9] A. Semmo, T. Isenberg, and J. Döllner, "Neural style transfer: A paradigm shift for image-based artistic rendering?" in Proc. Symp. Non-Photorealistic Animation Rendering, 2017, pp. 5:1-5:13.
[10] J. E. Kyprianidis et al., "State of the 'art': A taxonomy of artistic stylization techniques for images and video," IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 5, pp. 866-885, May 2013.
[11] P. Rosin and J. Collomosse, Image and Video-based Artistic Stylisation, Berlin, Germany: Springer Science Business Media, vol. 42, 2012.
[12] A. Dosovitskiy et al., "Discriminative unsupervised feature learning with exemplar convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2014.