MJRET
OPEN ACCESS

Multidisciplinary Journal of Research in Engineering and Technology
www.mjret.in
eISSN:2348-6953

NEURAL STYLE TRANSFER: A COMPARATIVE STUDY USING VGG NETWORK

Animesh Singh, Ayush Gautam, Merin Meleet

Department of Information Science and Engineering
R V College of Engineering
Mysore Road, Bengaluru

Abstract: *Style transfer is an optimization technique used to create a new image out of a content image and a style image (as in an artistic work by a well-known painter) as a result of blending, the content image is altered to be painted in the same style as the content image but to resemble it of the style image. The essential method for creating style transfer is the convolutional neural network that allows for transfer (CNN). This paper will analyse key methods for performing style transfer photos and quickly compare multiple models and their outcomes. We primarily contrast one pioneering technique, developed by Leon Gatys in 2015, with the most recent technology in the High-Resolution Network from 2019 for the transfer of photorealistic style. Gatys's Neural Algorithm of Artistic Style produces excellent results, but the resulting data doesn't maintain the content image's characteristics and the fact that the paper gives no concise description of the internal mechanisms gram matrix. The enhancement is the conversion to a photorealistic style transfer from Gatys' neural style transfer. It aids in maintaining the structure and defining characteristics within the content image. Both methods make use of VGG, a network that was trained using the ImageNet dataset for conducting image categorization. We made our own data set with five types or groups and used transfer learning in a VGG network to do a style transfer using transfer learning.*

Keywords— *Neural Style Transfer, CNN, Photorealistic, VGG19, Convolutional Neural Network, Content Image, Style image*

1. INTRODUCTION

The process of using a content image and a style image as input and merging them to produce an image that has both the content of the content image and the style of the style image is known as style transfer. Convolutional neural network (CNN) is the primary method that enables neural style transfer. By separating and recombining the visual content and style, convolutional neural networks (CNN) aid in producing artistic spectacular imagery. We have a variety of ways to separately describe the semantic content of an image and the style in which the material is presented in order to mix the content of the content picture and the style of the style image.

Recent developments in CNN have allowed us to successfully take on this problem. By taking on this problem, the style transfer approach shows the convolutional neural networks' capacity for sophisticated picture synthesis and modification while also offering fresh insights into the deep image representations they have learnt. CNN has the ability to extract style information from a well-known piece of art and content information from any random image. Neural Style Transfer (NST) is the term for this method of employing CNNs to generate a content image in many styles. We will also discuss assessment measures and the datasets which we used in order train CNNs to transfer styles evaluation metrics are simply a comparison of several models that are used to convey style.

Although most individuals lack the necessary expertise and technique to manufacture or make paint or add some flair to their own images, art is a crucial component of peoples' lives. Most often, people wish they could have a painting made of one of their images, but this is a tough process that takes a lot of time, and when mistakes are made when manually painting by hand, they may be challenging to correct. Thus, the concept of "style transfer" is created, which quickly transforms a given image into a painting-like representation using artificial intelligence.

The main goal of this study is to use our own dataset to compare the outcomes of neural style transfer versus photorealistic style transfer, applicable to both the quality and type of speech, as well as the composition of the speaker.

2. LITERATURE SURVEY

Users of Style Transferring may style their photographs in whatever way they wish to transfer them. Users may primarily build their own creative images using this technology in the manner of their choice. The main concept driving this system's development is how simple it would be to convert camera-captured photos into paintings.

In the past, style transfer was carried out using the Prisma app, which Alexey Moiseenkov released in June 2016 in order to produce astonishing picture effects and turn photographs into artworks. Prisma employs artificial neural networks to give users the ability to transform images into works of art that resemble those by Salvador Dali, Munch, or even Picasso.

A Neural Algorithm of Artistic Style by Leon Gatys, Alexander Ecker, and Matthias Bethge is the research paper that forms the basis of the Prisma App technology. It was presented in 2015 at the prestigious machine learning conference, Neural Information Processing Systems (NIPS) [1], [2]. The institution and the business are not related, and this technology was created separately and before Prisma.

Although earlier algorithms [1], [2] before Neural Algorithm of Artistic Style produced impressive results, they were all bound by the same fundamental flaw: they relied solely on low-level picture attributes of the target image to guide the texture transfer. Strong computer vision systems that learn to extract high-level semantic information from real images have been created by convolutional neural networks and are utilised in Gatys Paper [1], [2]. For object recognition and localisation, utilise the VGG16 Network.

The Neural Algorithm of Artistic Style generates excellent results, however the neural style transfer concept, particularly why the Gram matrices might reflect style, is yet not fully understood. Additionally, the correct preservation of content picture attributes is lacking.

Later, research on the photorealistic style transfer was conducted with the goal of transferring the style of one picture to another while maintaining the original contour and structure of the content image, making the content image continue to seem as a genuine shot after the style transfer. High Resolution Network for Photorealistic Style Transfer, which was inspired by the network proposed by Johnson et al [3], offers a solution that has a generation network to generate the output image and a pre-trained network to calculate the content loss and style loss, but our generation network's architecture is different from the network in Johnson et al [3]. For object location and recognition, VGG19 Network is employed.

[5] With the development of computer graphics, deep learning has been greatly developed because it can be used to train object recognition models. CNN can extract features. In addition to image recognition and image classification, CNN is also used for style transfer. [4] introduced the high-resolution network for posture estimation and updated the record of the COCO pose estimation data set, which served as the basis for the generation network in the High-Resolution Network. The goal of a high-resolution network is to continually accept data from low-resolution networks while maintaining high resolution representations throughout the process. In compared to other networks, the high-resolution network offers two advantages.

- In contrast to most current networks, the high-resolution network connects both high- and low-resolution subnets in tandem.
- To improve high resolution representation, do recurrent multi-scale fusion using low-resolution representations of the same depth and related levels.

3. METHODOLOGY

To create a new image known as the final image, two input images, the content image and the style image, are needed. The generated picture is stylistically comparable to the style image and has the same information as the content image. We must concentrate on the actual operation of a convolution neural network in order to produce an output image.

The style loss, content loss, and other characteristics are computed as the pictures pass through the VGG network, and the overall loss is determined, which is then utilised to create a new output image.

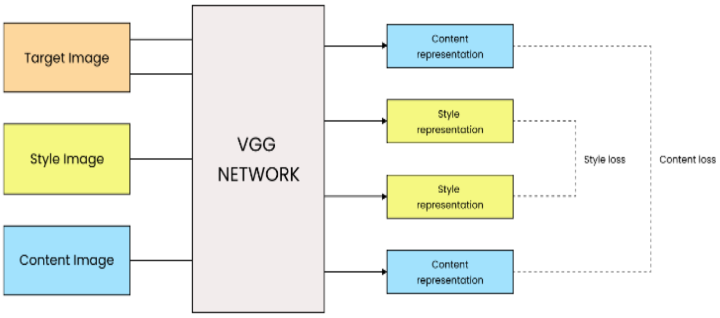


Fig. 1: Flow diagram of NST

1) **Content Loss:** Content loss refers to the resemblance between a noisy picture(G) produced at random and the content image (C). Let P and F represent the original picture and the produced image, respectively, to determine content loss. Additionally, we pick layer L as a hidden layer in a network to compute the loss, and F[l] and P[l] are feature representations of the corresponding pictures in layer L. The loss of content is now described as follows:

$$L_{content}\ (\vec{p},\ \vec{q},\ l)=\frac{1}{2}\sum_{i,j}\left(F_{ij}^{l}-P_{ij}^{l}\right) \qquad (1)$$

2) **Style Loss:** A style image's distance from an output picture is referred to as "style loss." The degree of correlation between feature maps in a particular layer [l] is used to quantify style information. We compute the dot product between the vectors of the activations of the two filters in order to determine the correlation between various filters or channels. The resulting matrix is known as the **Gram Matrix.**

Two channels are considered to be correlated if the dot-product across the activation of the two filters is substantial; otherwise, the pictures are said to be uncorrelated.

Mathematically:

Gram Matrix of Style Image(S): Here, the letters k and k stand for various channels or filters inside the layer L. Let's name this G_{kk}[l][S].

$$Gkk[l][s] = \sum_{i}^{H} \sum_{j}^{W} \left(A_{ijk}[l][G] - A_{ijk'}[l][G]\right) \quad (2)$$

Gram Matrix of Generated Image(G): Here, the letters k and k stand for various channels or filters inside the layer L. Let's name this G_{kk}[l][G].

$$Gkk[l][s] = \sum_{i}^{H} \sum_{j}^{W} \left(A_{ijk}[l][G] - A_{ijk'}[l][G]\right) \quad (3)$$

The square of the difference between the Gram Matrix of the style image and the Gram Matrix of the generated image represents the cost function between style and generated images.

$$L_{style} = \frac{1}{\left(2*H^{l}*W^{l}*C^{l}\right)^{2}} \sum_{K} \sum_{K'} (G_{kk'}\ [l][S] - G_{kk'}\ [l][G]) \quad (4)$$

Total Loss Function :

$L_{total} = \alpha\ L_{content} + \beta\ L_{style}$ (5)

α is content weight and β is style weight.

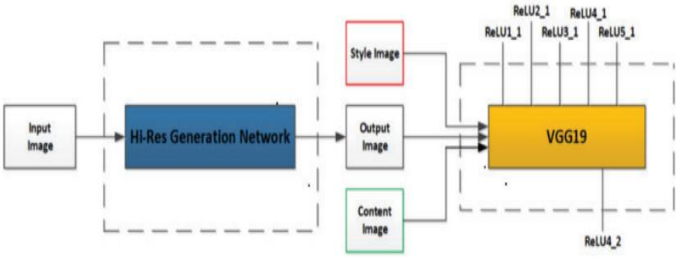
Our randomly generated image will be optimised into a meaningful work of art after the loss has been determined and can be reduced via back propagation.

3) **Photorealistic Style Transfer**: The goal of photorealistic style transfer is to transfer the look of one picture to another while maintaining the content image's original form and structure. As a result, the content image continues to appear authentic after the style transfer. Some previously suggested realistic picture style techniques run the risk of losing the content image's fine details and producing strange distortion structures [5]. The primary goal of photorealistic image stylization, sometimes referred to as colour style transfer, is to transfer the color-distribution style from the content loss text. When merely modifying the style without altering the structure of the picture, the feature representation computed by the loss network VGG for the content image and the output image should be identical. This paper's two primary contributions are as follows: In the beginning, a high-resolution network is suggested as the generation network to convey the style with a cleaner structure and fewer distortion. Second, a novel option for photorealistic style transfer is achieved by the effective use of the conventional natural image style transfer technique.

We utilise Euclidean distance for the content loss, as demonstrated by the expression:

$$I^{\varnothing,j}_{content}(y,y') = \frac{1}{C_jH_jW_j}\left\|\varnothing_j(y') - \varnothing_j(y)\right\|^2 \qquad (6)$$

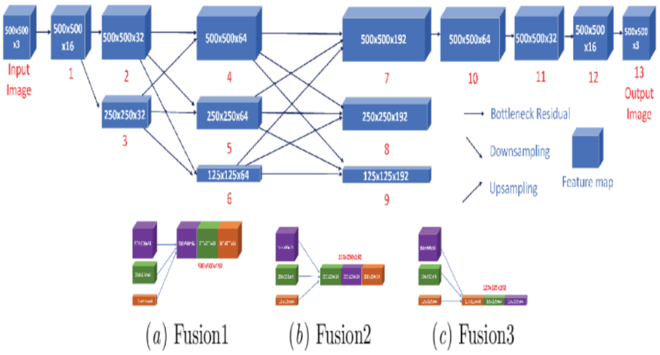
Fig. 2: Photorealistic Style Transfer



Style Loss*:* The squared Frobenius norm of the difference between the Gram matrices of the output and content pictures is used to measure style loss:

$$I^{\varnothing,j}_{style}(y,y') = \| G^{\varnothing}_j(y) - G^{\varnothing}_j(y')\|^2 \qquad (7)$$

Fig. 3: High-Resolution generation network structure



Total Loss: The formula for total loss is:

$$\hat{y} = \underset{y}{\operatorname{argmin}}\, \lambda_c l^{(\varnothing,j)}_{content}(y,y_c) + l^{(\varnothing,j)}_{content}(y,y_s) + \lambda_{TV} l_{TV}(y) \qquad (8)$$

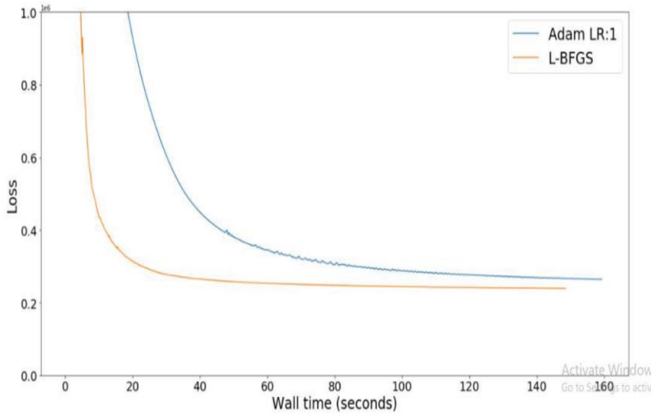
A. *Optimization Technique*

An objective function, also known as an error function, is a mathematical function dependent on the model's internal learnable parameters that are used to compute the target values (Y) from the set of predictors (X) used in the model. Optimization algorithms help us minimise (or maximise) an objective function, or error function, E(x). For instance, we refer to the Weights(W) and Bias(b) values of the neural network as its internal learnable parameters. These values are used in computing the output values and are learned and updated in the direction of the optimal solution, i.e. minimising the Loss, by the networks training process[6].

According to Gyats paper [2], L-BFGS is the optimization method used in neural style transfer, while Adam is the optimization algorithm used in photorealistic style transfer. Below is a comparison graph between Adam and L-BGFS. We employed the Adam optimization technique in our model, which is quicker than L-BFGS.

Fig. 4: L-GBFS vs Graph of Adam

4. RESULTS



A. Dataset

We have gathered roughly 8000 photos for our dataset from various sources. There are 5 classes in our dataset. Each class has between 1900 and 2000 photos total. An approximate 8000 picture training set and 1000 image validation set make up the dataset.

The dataset has four classes, which are explained below:

Deity: It is an assortment of pictures of different gods and goddesses from different faiths.

It is a collection of pictures of different stupas and temples called the Holy.

Travel and adventure: The collection includes views and landscapes from the Terai to the Himalayan areas.

Artworks: It consists of a variety of aesthetic and special artworks by different artists.

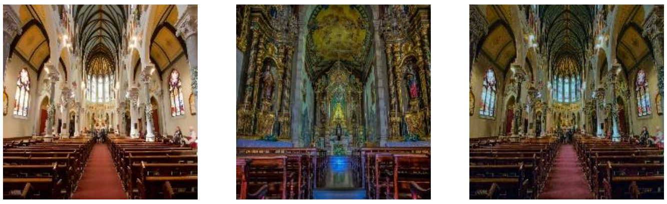
Architecture: This is a collection of images of beautiful museums, temples & churches.

Anime: It consists of different images of anime characters and animated scenery.

Given that we are utilising a VGG Network for transfer learning, we freeze up to 13 layers and employ dropout layers in the classification section after each dense layer. Since we are using the right model and sufficient data to do transfer learning, the issue of under fitting doesn't arise. To prevent overfitting, we also used data augmentation approaches. The user interface built to use the model had a smooth flow with user friendly and proxy avoiding aspects in the attendance management system. The overall system including the interface, storage and the model worked successfully in coherence when tested in real time.

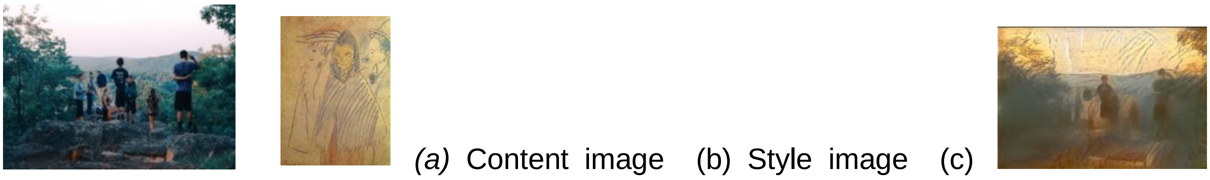
B. Output

The images after this paragraph show the outcomes of the Neural Style Transfer Algorithm and Photorealistic Style Transfer.



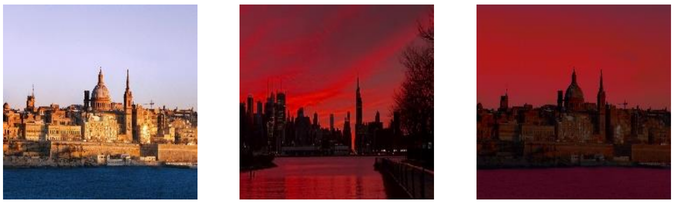
(a) Content image (b) Style image (c) Output Image

Fig. 5: Photorealistic Style Transfer

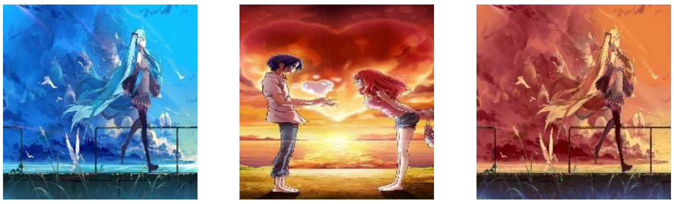


(a) Content image (b) Style image (c) Output Image

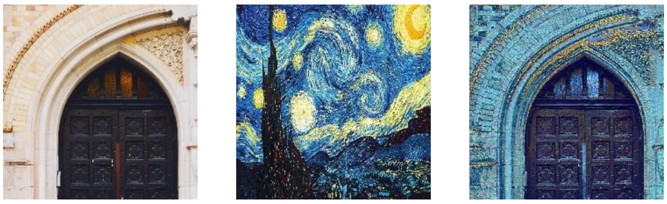
Fig. 6: Neural Style Transfer



(a) Content image (b) Style image (c) Output Image



(a) Content image (b) Style image (c) Output Image



(a) Content image (b) Style image (c) Output Image

Even while the neural algorithm for style transfer delivered fantastic results, it also entirely ignores the semantic information included in the content image, such as the weather, the colour of the sky, and the ability to tell the difference between day and night. Additionally, it does not properly preserve characteristics from content images.

The output picture has the same structure as the content image, and photorealistic style transfer preserves the curves and structure of the source image without distorting it. Additionally, it has a more complex structure and a more accurate colour scheme.

5. CONCLUSION

We found that the output of the Gatys and Photorealistic style transfer was different, as indicated in the graphs (Adam versus L-BFGS) above, and we came to the conclusion that the more classes in the dataset, the more accurate the results may be. Due to the utilisation of the approximately 20,000 class ImageNet dataset, the VGG model's accuracy was 92.3 percent. Our own dataset was used for transfer learning in the VGG model, and the accuracy resulting from this was 89.8 percent.

REFERENCES

[1] L. Gatys, A. Ecker, and M. Bethge, “A Neural Algorithm of Artistic Style,” Journal of Vision, vol. 16, no. 12, p. 326, 2016.

[2] G. Leon A, A. S. Ecker, and M. Bethge, “Image Style Transfer Using Convolutional Neural Networks Leon,” Arabian Journal of Geosciences, vol. 11, no. 21, pp. 2414–2423, 2018.

[3] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9906 LNCS, pp. 694–711, 2016.

[4] Chengsi Yao, Yuanhao Li,Yali Qi ”Research on neural style transfer algorithm”, AMIMA [2019].

[5] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep High-Resolution Representation Learning for Human Pose Estimation.”

[6] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, “On optimization methods for deep learning,” Proceedings of the 28th International Conference on Machine Learning, ICML 2011, pp. 265–272, 2011