Comparative Study of Style Transfer

1st Bibek Adhikari
Dept. of Computer and Electronic Engineering
Kantipur Engineering College, T.U
Lalitpur, Nepal
bibekadhikari97@gmail.com

2nd Sunil Ojha
Dept. of Computer and Electronic Engineering
Kantipur Engineering College, T.U
Lalitpur, Nepal
sunilojha525@gmail.com

3rd Sunil Shrestha
Dept. of Computer and Electronic Engineering
Kantipur Engineering College, T.U
Lalitpur, Nepal
sunilshrestha180@gmail.com

4th Er. Shayak Raj Giri(MSc)
Supervisor
Dept. of Computer and Electronic Engineering
Pulchowk Campus,Institute of Engineering, T.U
Lalitpur, Nepal
shayakraj@ntc.net.np

5th Er. Bishal Thapa
Project Assistant
Dept. of Computer and Electronic Engineering
Kantipur Engineering College, T.U
Lalitpur, Nepal
bishalthapa@kec.edu.np

***Abstract*—Style transfer is the optimization technique used to create to new image using style image (such as artwork by famous painter) and input image you want to style and blend together such that image is transformed to look like the content image, but painted in the style of the style image. The key technique that makes style transfer possible is convolutional neural network (CNN). This paper will survey major technique of doing style transfer on images, and briefly compare different model and result. We mainly compare one of pioneer technique introduced by the Leon Gatys in 2015 and the latest technique in 2019 in the paper High Resolution Network for Photorealistic Style Transfer. A Neural Algorithm of Artistic Style, introduced by Gatys generates great result, but the result so obtained didnt preserve the feature of the content image and also that paper hasnt briefly described about the inner working of gram matrix. Photorealistic style transfer is the improvement or enhancement to the neural style transfer by Gatys. It helps to conserve the structure and common feature in the content image. Both the techniques use VGG Network, which was trained on ImageNet Dataset for performing classification of images. We created our own dataset which consists of 5 classes or categories and perform transfer learning in VGG Network to perform a style transfer.**

***Index Terms*—Neuralstyle Transfer, CNN, Photorealistic**

I. INTRODUCTION

Style Transfer is the problem of taking a content image and a style image as input, and out putting an image that has the content of the content image and the style of the style image. The key technique that makes neural style transfer possible is convolutional neural network (CNN). Convolutional Neural Networks (CNN) helps to creates artistic fantastic imagery by separating and recombining the image content and style. In order to combine the content of the content image and the style of the style image, we have many ways to independently represent the semantic content of an image and the style in which the content is presented.

Recent advances in CNN, we are able to tackle this challenge with great success. By tackling this challenge, the method of style transfer provides new insights into the deep image representations learned by Convolutional Neural Networks and demonstrate their potential for high level image synthesis and manipulation. CNN is capable of extracting content information from an arbitrary photograph and style information from a well-known artwork. This process of using CNNs to render a content image in different styles is referred to as Neural Style Transfer (NST). We will also cover the datasets used to train CNNs to perform style transfer and evaluation metrics for this task. Evaluation metrices being just the comparison between different model used style transfer.

Art is an essential part of peoples life but majority of people lack a proper skill and technique to produce or make paint or introduce some style to their own photos. People mostly desire that they had a painting of their photos which is a complex task that takes more time and sometimes while manually painting by hand there may encounter mistakes which will be difficult to overcome. So, the idea of style transfer comes into existence which makes the painting type of image of a particular image within no time using its artificial intelligence.

The core objective of this research paper is to compare the result of Neural Style Transfer and Photo-

realistic Style Transfer with the use of our own dataset.

II. RELATED WORK

Style Transferring allows users to style image to any type of style they want to transfer their images. This system mainly allows users to create their own artistic images in the way they desire. The core idea behind this system development is that it would be very easy to transfer the images taken by the cameras into the painting form.

Previously style transfer was done using an app called Prisma which was launched in June 2016 by Alexey Moiseenkov in order to create amazing photo effects, transforming photos into paintings. Prisma uses artificial neural networks that enable users to make photos appear like they were painted by Picasso, Munch or even Salvador Dali himself.

The research paper behind the Prisma App technology is called A Neural Algorithm of Artistic Style by Leon Gatys, Alexander Ecker and Matthias Bethge and was presented at the premier machine learning conference: Neural Information Processing Systems (NIPS) in 2015 [1], [2]. This technology was developed independently and before Prisma, and both the university and the company have no affiliation with one another.

Previous algorithms before Neural Algorithm of Artistic Style [1], [2] achieve remarkable results, but they all suffer from the same fundamental limitation: they use only low-level image features of the target image to inform the texture transfer. Convolutional Neural Network has produced powerful computer vision systems that learn to extract high-level semantic information from natural image which is used in Gatys Paper [1], [2]. VGG16 Network is used for object recognition and localization.

Neural Algorithm of Artistic Style produces a great result, but the principle of neural style transfer, especially why the Gram matrices could represent style remains unclear. It also lacks the proper preservation of features of content image.

Later the photorealistic style transfer research was done which aim to transfer the style of one image to another, but preserves the original structure and detail outline of the content image, which makes the content image still look like a real shot after the style transfer. Inspired by the network proposed by Johnson et al [3], High Resolution Network for Photorealistic Style Transfer provides a solution which has a generation network to generate the output image, and a pre-trained network to calculate the content loss and style loss, but the architecture of our generation network is different from the network in Johnson et al [3].VGG19 Network is used for object recognition and localization.

Generation network in High Resolution Network is inspired by Sun et al. (2019) [4] , who proposed the high-resolution network for pose estimation and refreshed the record of the COCO pose estimation data set. Most networks have the high-to-low and low-to-high processes. The high-to-low process aim to produce lower resolutions and higher channel counts, while the low-to-high process is designed to produce high-resolution representations and reduce the resolution of the feature maps. High-resolution network is designed to maintain high resolution representations through the whole process and continuously receive information from low-resolution networks. The high-resolution network has two benefits in comparison to other networks.

- The high-resolution network connects both high- and low-resolution subnets in parallel, rather than connecting in series like most existing networks.
- Perform repeated multi-scale fusion with the help of low-resolution representations of the same depth and similar levels to enhance high resolution representation.

III. METHOD

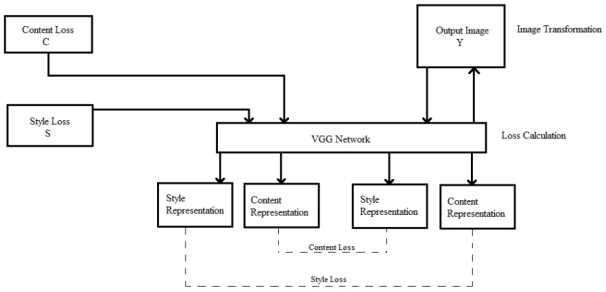


Fig. 1: system flow diagram

As you can see in Figure above, there are two input images namely content image and style image that are used to generate a new image called stylized image. A few things to notice about this image is that it has the same content as the content image and has a style similar to that of the style image. It looks good and we are pretty sure its not achieved by overlapping these two images so how do we get here what is the math behind this idea? To answer these question we need to take a step back and focus on what does a convolution neural network actually work.

As the images go under the VGG network the style loss, content loss and other features are calculated and total loss is calculated which is then used to generate a new output image.

1) Content loss: Calculating content loss means how similar is the randomly generated noisy image(G) to the content image(C).In order to calculate content loss :

Assume that we choose a hidden layer (L) in a network to compute the loss.Therefore, let P and F be the original image and the image that is generated.And, F[l] and P[l] be feature representation of the respective

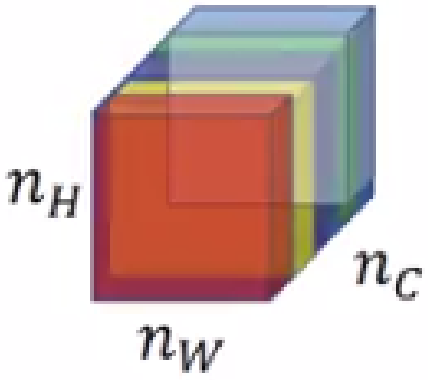


Fig. 2: Different channels or Feature maps in layer l

images in layer L.Now,the content loss is defined as follows:

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

(1)

2) Style loss: Before calculating style loss, lets see what is the meaning of **style of a image** or how we capture style of an image. This image shows different channels or feature maps or filters at a particular chosen layer l.Now, in order to capture the style of an image we would calculate how **correlated** these filters are to each other meaning how similar are these feature maps.

The first two channel in the above image be Red and Yellow.Suppose, the red channel captures some simple feature (say, vertical lines) and if these two channels were correlated then whenever in the image there is a vertical lines that is detected by Red channel then there will be a Yellow-ish effect of the second channel. Now,lets look at how to calculate these correlations (mathematically).

In-order to calculate a correlation between different filters or channels we calculate the dot-product between the vectors of the activations of the two filters.The matrix thus obtained is called **Gram Matrix**.

If the dot-product across the activation of two filters is large then two channels are said to be correlated and if it is small then the images are un-correlated.Putting it mathematically :

Gram Matrix of Style Image(S): Here k and k represents different filters or channels of the layer L. Lets call this Gkk[l][S].

$$Gkk\,[l]\,[S] = \sum_{i}^{H} \sum_{j}^{W} (A_{ijk}{}^{[l][S]} - A_{ijk'}{}^{[l][S]})$$

(2)

Gram Matrix for Generated Image(G): Here k and k represents different filters or channels of the layer L.Lets call this Gkk[l][G].

$$Gkk\,[l]\,[G] = \sum_{i}^{H} \sum_{j}^{W} (A_{ijk}{}^{[l][G]} - A_{ijk'}{}^{[l][G]})$$

(3)

Now,we are in the position to define Style loss:

Cost function between Style and Generated Image is the square of difference between the Gram Matrix of the style Image with the Gram Matrix of generated Image.

$$L_{style} = \frac{1}{(2 * H^l * W^l * C^l)^2} \sum_{K} \sum_{K'} (Gkk'^{[l][S]} - Gkk'^{[l][G]})$$

(4)

Total Loss Function :

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

(5)

Alpha and beta in the above equation are used for weighing Content and Style cost respectively.In general,they define the weightage of each cost in the Generated output image.

Once the loss is calculated,then this loss can be minimized using backpropagation which in turn will optimize our randomly generated image into a meaningful piece of art.

3) Photorealistic Style Transfer: Photorealistic style transfer aims to transfer the style of one image to another, but preserves the original structure and detail outline of the content image, which makes the content image still look like a real shot after the style transfer. Although some realistic image styling methods have been proposed, these methods are vulnerable to lose the details of the content image and produce some irregular distortion structures [5].The main purpose of photorealistic image stylization (also known as color style transfer) is to transfer the style of color distributions text of Content Loss :Content image and the output image should have a similar feature representation as computed by loss network VGG. Because for only changing the style without any changes to the structure of the image. The main contributions of this paper are two-fold: First, high-resolution network as the generation network to transfer the style with a ner structure and less distortion is proposed. Second, the photorealistic style transfer successfully using traditional natural image style transfer algorithm, which

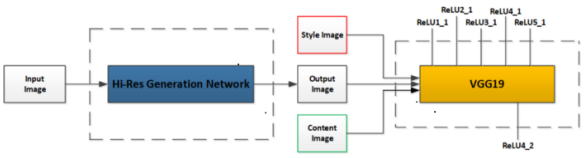
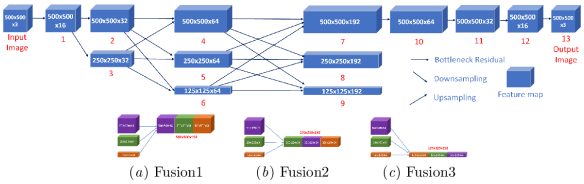


Fig. 3: Overview of Photorealistic Style Transfer



(a) Fusion1 (b) Fusion2 (c) Fusion3

Fig. 4: Structure of high-resolution generation network

provides a new choice for photorealistic style transfer.

For the content loss, we use Euclidean distance as shown by the formula:

$$l^{\emptyset,j}_{content}(y\ ,y') = \frac{1}{C_j H_j W_j}||\emptyset_j(y') - \emptyset_j(y)||^2 \tag{6}$$

Style Loss: The style loss is dened as the squared Frobenius norm of the dierence between the Gram matrices of the output and content images:

$$l^{\emptyset,j}_{style}(y\ ,y') = \left|\left|G^{\emptyset}_j(y) - G^{\emptyset}_j(y')\right|\right|^2 \tag{7}$$

Total loss: The total loss is given by :

$$\hat{y} = \underset{y}{arg\min}\ \lambda_c l^{(\emptyset,j)}_{content}(y, y_c) + \lambda_s l^{(\emptyset,J)}_{content}(y, y_s) + \lambda_{TV} l_{TV}(y)$$

A. Dataset

For our dataset, we have collected about 7300 images from different sources. Our Dataset consists of 5 classes. Images are divided between 1200-1500 images per class. The dataset contains a training set of about 5300 images, a validation set of 1500 images. Four classes included in dataset are described below:

Deity: It is a collection of images of various gods and goddess from various religions.

Holy: It is a collection of images of various temples and Stupas.

Scenery: It is a collection of scenes and landscape from Himalayan region to Terai regions.

Perception: It is a collection of special effects and artistic effects in various images.

People: It is the collection of the pictures of human beings.

B. Optimization Method

Optimization algorithms helps us to minimize(or maximize) an Objective function(another name for Error function) E(x) which is simply a mathematical function dependent on the Models internal learnable parameters which are used in computing the target values(Y) from the set of predictors(X) used in the model. For example we call the Weights(W) and the Bias(b) values of the neural network as its internal learnable parameters which are used in computing the output values and are learned and updated in the direction of optimal solution i.e minimizing the Loss by the networks training process and also play a major role in the training process of the Neural Network Model [6] .

The optimzation algorithm which is used in Neural style Transfer according to Gyats paper [2] is L-BFGS and in Photorealistic Style Transfer [5] the algorithm used is Adam . The comparative graph between Adam and L-BGFS is shown in below. In our model we have used Adam optimization algorithm which is fast than L-BFGS.

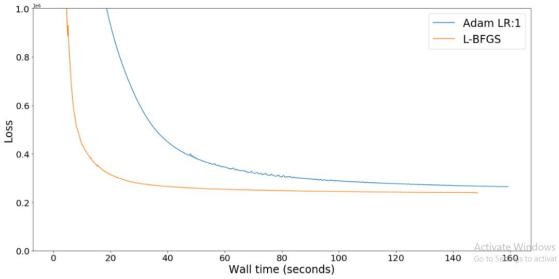


Fig. 5: Graph of Adam vs L-BFGS

IV. Result and Conslusion

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

Fig. 6: Neural style Transfer

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

Fig. 7: Photorealistic Style Transfer

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

Fig. 8: Our Output

Results from Neural Algorithm Style Transfer and Photorealistic Style Transfer can be seen above.

Effect Comparison: Although, Neural Algorithm of Style Transfer produced amazing result, also completely ignores the semantic information of content image like weather, coloring of the sky and also couldnt differentiate between day and night. It also lacks the proper preservation of features from content image. Photorealistic Style Transfer maintains the curves and structure of the content image and are not distorted and the output image has the same structure as content image. It also has a more elaborate structure and a more realistic color distribution.

V. Conclusion

As we did the comparative study of result of Gatys and Photorealistic style transfer we observed different output as shown in graphs(Adam vs L-BFGS) above and concluded that as the number of classes of the dataset increases the greater accuracy can be obtained. The accuracy of VGG model was 92.3% as it has used ImageNet dataset which has for about 20,000 classes.

Transfer learning was performed in the VGG model with our own dataset and the accuracy so obtained was 88.1%.

VI. Acknowledgment

We express our gratitude towards the Computer and Electronic Department of Kantipur Engineering College for providing us the learning and working environment which was helpful to work for our team. Thanks to our Head of Department Er. Rabindra Khati sir for continuous support and guidance. We are very grateful to our project supervisor Shayak Raj Giri for his continuous supervision and Er.Bishal Thapa for his support and regular feedback. And finally we are grateful to our family and friends for their encouragement in getting us complete this project within the stipulated time.

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. [Online]. Available: http://jov.arvojournals.org/article.aspx?doi=10.1167/16.12.326

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

[5] M. Li, C. Ye, and W. Li, "High-Resolution Network for Photorealistic Style Transfer," no. 2001, pp. 1–14, 2019. [Online]. Available: http://arxiv.org/abs/1904.11617

[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. [Online]. Available: http://www.cs.stanford.edu/ acoates/papers/LeNgiCoaLahProNg11.pdf