

Image Style Transfer Using CNN

Armaan Khan^[1], Ankit Kumar^[2], Dr. Nilima Kulkarni^[3], Anirudh Kamat^[4], Deepak Kumar^[5]

Department of Computer Science and Engineering, MIT School of Engineering,
MIT Arts Design and Technology University, Pune, 412201, India

Abstract— Image style transfer was a really difficult task earlier because image processing takes a lot of amounts of computational power. As the technology is improving day by day due to which machines with higher computations are available easily and thus the image processing task is made easier than before. And as for the image style transfer, we have to learn features from the image where the size of the images could be very high so computation increases significantly. But now various improvements are made on Convolutional Neural Networks (CNNs) and with the help of transfer learning we already have pretrained models like VGGnet which is a 19 layers deep neural network architecture to work on, with the help of which we are saved from writing everything from the scratch. Image style transfer is basically an algorithm which uses Convolutional Neural Networks to learn features from two different images and mixes them together in a new image which is basically a combination of the two input images.

Keywords— Convolutional Neural Network (CNN), Neural Style Transfer (NST), Deep Neural Network (DNN), Visual Geometry Group (VGG) net, Deep Learning

I. INTRODUCTION

The term "Neural Style Transfer" refers to a set of software algorithms that change the appearance and visual style of other digital photographs. Neural Style Transfer (NST) algorithms are defined by their use of convolutional neural networks (CNNs) for image transformation. NST is frequently used to create new works of art from photographs, such as converting the impression of famous paintings to user-supplied images.

CNNs are artificial neural networks that can be used to classify images. They trained on huge labeled datasets and learned end-to-end feature extraction and classification. Transferring the style of an image from one image to another image can be considered as image texture transfer problem. The objective of texture transfer is to extract the textures from the input images while keeping the content of the original input image as good as possible.

Neural style transfer uses Convolutional Neural Networks to combines two images, first - a content image (Fig1) (can be any picture whom art style you want to create) and second - a style image (Fig1) (such as a painting or a design) so that the output image (Fig2) appears to be the original content image but it would seem painted in the style of the style image[1].

Consider the following images:



Fig. 1. Style Image and content image

How would it look if we chose to paint this Turtle purely in this style? Something like this?

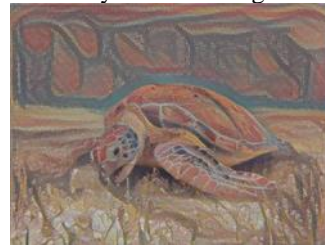


Fig. 2. Resultant image of Fig 1's content and Fig 2's style.

II. RELATED WORK

In [2] Leon A. Gatys used simple deep learning which allowed them to generate a new image using art style of style image and a content image. And They found that the style and content image can be separated and can be used to generate new art.

In [3] Selim used a method which is more generally built for applying for Image Style transfer only portraits of people, They Created their Function to capture local distribution on top of general Image style transfer algorithm for better capturing the painting texture and maintains the integrity of facial structures. And they found out that The general Image style transfer algorithm doesn't maintain the texture of the style painting which is very important in the case of portraits, with the help of their function they were able to maintain the texture and integrity of the portrait

III. PROCESS/METHOD

A. Model Architecture

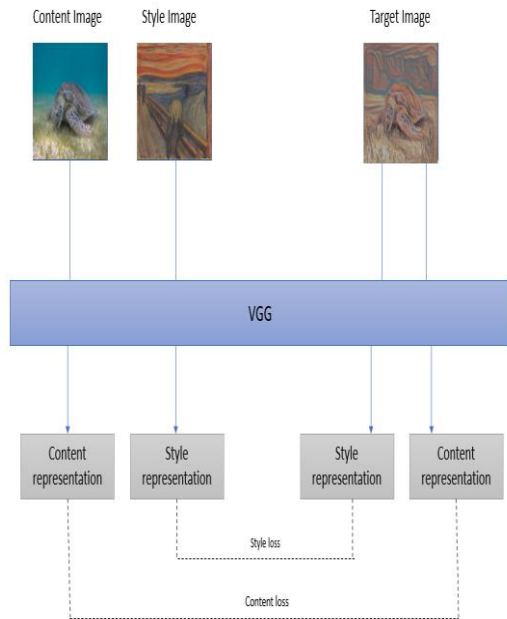


Fig. 3. Architecture of proposed system

We would be providing input as content image and style image to the system. then based on the feature mapping it will extract the features from Content Image as well as Style Image and those features extracted images would be called Content representation and Style representation respectively. The difference between original content image and target image is content loss and difference in original style image and target image is called style loss. When we combine these two we get total loss and we need to minimize it to get high quality target image

B. Proposed System

Convolutional neural networks are comparable to human performance in the common visual object recognition reference task. We utilized the feature space provided by sixteen convolutional layers and five grouped layers from the VGG Network's [4] nineteenth layer.

There were two tasks, first is to generate content image from the input content image and second is to generate style image from the input style image

This was accomplished through the use of feature mapping. Convolutional feature maps, in general, provide an excellent representation of the features in an input image. They preserve the spatial information contained in an image while omitting the style information (if a feature map is used as it is)

The loss that incurred during the generation of content image is called content loss [2] and it can be calculated as

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

For chosen content layer 'l' $L_{content}$ is the mean square error, F_{ij} is the content image and P_{ij} is the generated image (P_{ij}).

For the content image we don't have to use the deep layers of VGG architecture because we want most of its feature as it is. So we have used the features of convolutional layer 2, convolutional layer 4, convolutional layer 7.

Now for the generation of style image, to calculate the generated style image loss we require gram matrix [2].

For the creation of gram matrix, we need to find the correlation among the color channels, the color channels which share higher correlation would contribute more towards the generation of style image. The correlation of all the channels with respect to each other is given by Gram Matrix. We use dot product to find the correlation as dot product helps us find how similar two vectors actually are.

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \quad (2)$$

Here G_{ij} is the resultant gram matrix, and F_{ik} is the original matrix, F_{jk} is the transposed matrix.

Now to find the style loss [2], we need to find the mean square error between input image gram matrix and feature map of generated style image

$$L_{style} = \frac{1}{2} \sum_{i=0}^L (G_{ij}^l - A_{ij}^l)^2 \quad (3)$$

Here G_{ij} is the Gram Matrix, A_{ij} is the feature map of generated style image.

For the style image we can use the deep layers of VGG network so here the features convolutional layer 2, convolutional layer 4, convolutional layer 10, convolutional layer 13 are used.

As we have found out the content and style loss. We can now find the Total loss [2],

$$L_{total} = \alpha L_{content} + \beta L_{style} \quad (4)$$

Here we cannot add the two losses directly so α and β are the weights that we had to use to find the weighted sum of content loss and style loss.

Once the we calculate the total loss, then our job would be to minimize this loss to get the output image as similar as the content image and the total loss can be minimized by using backpropagation which in every iteration tries to decreases the total loss value and finally the output resultant image would look like a piece of art.

IV. RESULTS

The key finding of this paper is that in CNN the content and style are separate. We can extract the content or the style of an image and create a new unique image using it. The algorithm has allowed us to generate new images of high quality that is a combination of the content of normal image and the artworks. The results shed new light on how Convolutional Neural Networks learn deep image representations

Below images are the few result images generated by our model here first image is the content image, second is the style image and last one is the generated image.

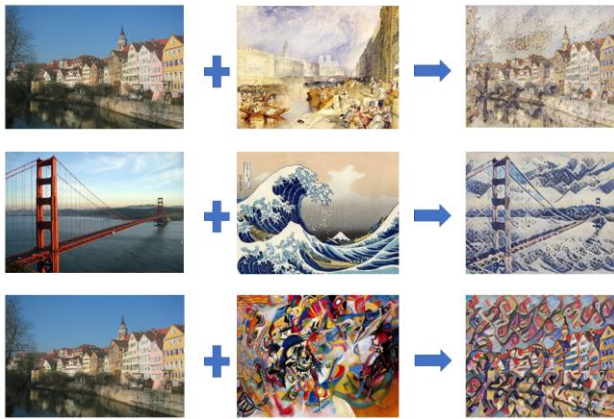


Fig. 4. Some of the sample images and their result we achieved

V. DISCUSSION AND FUTURE WORK

Recently there have been multiple researches done on the neural style transfer. and as the computational power is becoming cheaper and faster, and as more optimized and fast deep learning algorithms are discovered we are hoping to see more research on this topic. And its use in other related fields also

As for future work the performance can be improved by training the model with more data and working on advanced algorithms. And further can be converted into a mobile application

REFERENCES

- [1] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," ArXiv e-prints, Aug. 2015
- [2] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2414–2423
- [3] Selim, Ahmed, Mohamed Elgharib, and Linda Doyle. "Painting style transfer for head portraits using convolutional neural networks." *ACM Transactions on Graphics (ToG)* 35.4 (2016): 1-18.
- [4] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [5] P. Rosin and J. Collomosse, Image and video-based artistic stylisation. Springer Science & Business Media, 2012, vol. 42.
- [6] Ghiasi, Golnaz, et al. "Exploring the structure of a real-time, arbitrary neural artistic stylization network." *arXiv preprint arXiv:1705.06830* (2017).