Image Style Transfer Using CNNs

Nguyen Ba Thanh
Institution1 address
20224291

Le Thanh Long
Institution2 address
20224325

Abstract

Image style transfer, the process of rendering the semantic content of an image in the style of another, presents a significant challenge in image processing. A key limitation of prior approaches has been the inability to effectively disentangle image content and style, as well as to manipulate them independently. Leveraging image representations derived from Convolutional Neural Networks (CNNs) trained for object recognition, which explicitly encode high-level semantic information, we propose a novel algorithm for Image Style Transfer Using CNNs. This method enables the separation and recombination of image content and style, allowing the synthesis of new images that preserve the semantic structure of a content image while adopting the visual characteristics of a style image. By employing a loss function that simultaneously minimizes content and style discrepancies, our algorithm achieves results of high perceptual quality. This approach not only advances the state-of-the-art in artistic image synthesis but also offers deeper insights into the representational power of CNNs for high-level image analysis and manipulation.act

1. Introduction

The term "Neural Style Transfer" (NST) refers to a class of computational techniques that modify the visual appearance and artistic style of digital photographs. NST algorithms leverage the capabilities of Convolutional Neural Networks (CNNs) to transform images by combining the visual content of one image with the stylistic features of another. This technique has become widely recognized for its ability to generate new artistic creations, such as applying the style of famous paintings to user-provided photographs. CNNs, a specialized type of artificial neural network, are commonly used for image recognition tasks. These networks are trained on vast labeled datasets, learning to extract features and perform image classification in an end-to-end manner. The process of transferring style between images can be framed as a texture transfer problem, where the goal is to extract and apply the textures from a style image while preserving the semantic content of the original content image. In Neural Style Transfer, CNNs are used to merge two images: a content image (e.g., a photograph or any image whose structure is to be preserved) and a style image (e.g., a painting or artistic design). The objective is to generate an output image that retains the content and spatial arrangement of the content image while adopting the artistic textures and patterns of the style image. For instance, as shown in Figure 1, a user may supply a photograph as the content image and select an iconic painting as the style image. The resulting output, as illustrated in Figure 2, resembles the original photograph but appears to be rendered in the style of the chosen painting [1]. Neural Style Transfer has significantly expanded the creative possibilities of image synthesis, offering a powerful tool for artistic expression and blending the boundaries between art and technology. By harnessing the hierarchical feature extraction capabilities of CNNs, NST enables users to produce visually compelling and stylistically unique results.

(a) Content Image (b) Style Image

Figure 1. Style Image and content image

2. Related works

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 trans-

Figure 2. Resultant image of Fig 1

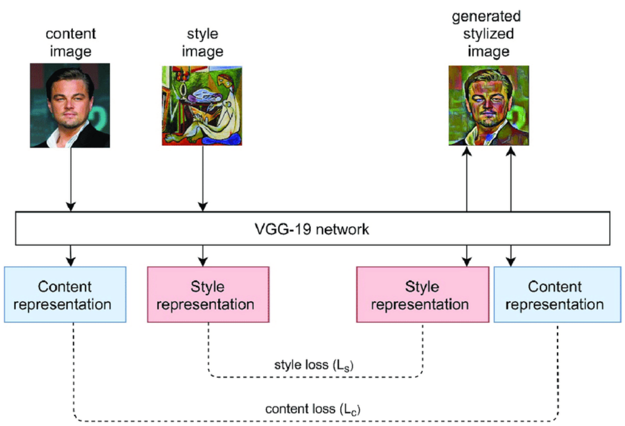

Figure 3. Architecture of proposed system

fer 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

3. Proposed Methods

A.Model Architecture

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 For a chosen content layer l, $\mathscr{L}_{\text{content}}$ is the mean square er-

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} \left(F_{ij}^{l} - P_{ij}^{l}\right)^2$$

ror, F_{ij} is the content image, and P_{ij} is the generated image.

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.

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}$$

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

$$\mathcal{L}_{style} = \frac{1}{2} \sum_{l=0}^{L} (G_{ij}^{l} - A_{ij}^{l})^2$$

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],

$$\mathcal{L}_{total}(\vec{p},\vec{a},\vec{x}) = \alpha\mathcal{L}_{content}(\vec{p},\vec{x}) + \beta\mathcal{L}_{style}(\vec{a},\vec{x})$$

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.

4. 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.

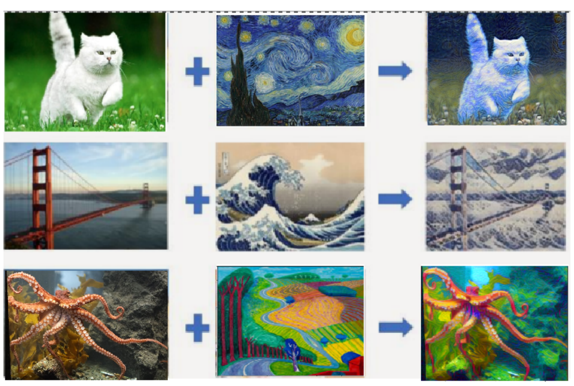

Figure 4. Some of the sample images and their result we achieved

5. 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] Ahmed Selim, Mohamed Elgharib, and Linda Doyle, "Painting style transfer for head portraits using convolutional neural networks," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 4, 2016, pp. 1–18.

[4] Karen Simonyan 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, vol. 42, 2012.

[6] Golnaz Ghiasi, et al., "Exploring the structure of a real-time, arbitrary neural artistic stylization network," *arXiv preprint arXiv:1705.06830*, 2017.