A Brief Overview of Image Style Transfer Techniques

Rishi Aluri (33274797), Running Wang (31713831), Supritha Konaje (32864477),
Yiping Song (32798768), Yuji Zhao (33290962)

***Abstract*—With the increase in popularity of social media apps such as 'Snapchat', 'Instagram' and other photo-reconstruction related applications, there has been a considerable growth in the amount of research seen in image reconstruction and stylization methods. The transfer of style to an image requires not only the image to be visually pleasing but also should be done within a suitable amount of time to be able to use them. This paper produces a rich review of the history and definition of image style transfer to the technical overview of some of the interesting methods developed over time. Not only have we also showcased the noteworthy outcomes and achievements but also briefly evaluated a few of the key methods along with the unresolved problems.**

I. Introduction

The principle of image styling is to give the original image the body of the secondary image. In other words, we get an image that mimics the style of the second image. It was difficult to apply the texture or style of one image to another using the traditional approach of drawing because everyone is not as talented as famous artists. Instead, a better, automated solution was discovered through Image Transfer. As a result, an artist's paintings can be lent to new, unique artistic approaches for better use.

The quality of a picture should not be rendered for accurate image stylization, and it should fulfil the result by providing a decent quality image even after employing algorithms to change the feel of the image.

This paper talks about the growth and gives an overview of Image Stylization. Section II talks about the History of Image Stylization where it mentions the need and growth of the various techniques used so far. Section III is about applying neural networks in Image Style Transfer, technical improvements, loss functions, and applications of Image Style Transfer in real situations. Section IV shows the achievements of others on image style transfer based on the algorithm of Gatys et al., making it easier to obtain complex and less distorted images, which contribute significantly to the image direction. In Section V, we briefly evaluate and technically analyse some of the different methods involved in the image-stylization field. Section VI discusses that there are still some problems to be faced, such as high hardware requirements for processing high-pixel images, and the universality of technology is not strong enough.

II. History of image Stylization

The concept of Image Stylization began with a simple algorithm in which programmers wrote code for each pixel or change in the image's nature. It didn't seem to be an issue at first because the photos weren't particularly complicated. With the trend of applying filters to camera-clicked images, games, virtual reality, and artwork transformation, the demand for stylization expanded as the years went by.

Complex images of artists could not be used in a photo stylization since capturing the fine details of prominent artists would be challenging. Initially, picture brushstrokes were the key conceptual theme in stylization, where we could blur or eliminate the clarity of an image with a single stroke. This would give the original image a cartoon-like appearance.

Patch Based Synthesis provided the foundation for the second period of stylization [1]. The algorithm is given one image, and it makes a new image with the features of the second given image. It superimposes the original image by using a few features and patches of the exemplar image. Texture synthesis, which focuses on synthesizing regular or semi-regular textures from exemplar photos, was the source of inspiration for this notion.

The current era is entirely dedicated to Neural Network Image Stylization. The notion of a neural network was first presented to identify textual data features. Many people wondered if this method might be used to extract features from one image and transfer them to another. For example, we are given a snapshot of a lovely house to which we must incorporate elements from Van Gogh's work Starry Night. We'll take features from the paintings and apply them to the photos.

Fig. 1: Stylization using Neural Network Algorithm

Figure 1 is an example of the starry night; other approaches would do poorly due to their various limitations. When the Neural Network technique is utilized, the image is flawlessly styled.

Each method has its own set of disadvantages and benefits. Neural Network has been very promising so far and many improvisations in the algorithms are being done. But in some circumstances, patch-based stylization would yield superior results than a Neural Network Method algorithm.

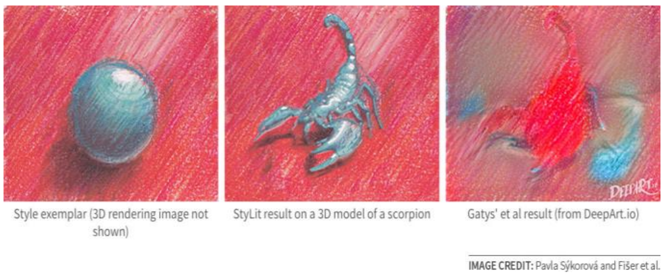

Fig. 2: Comparison of Image Stylization using Patch-Based and Neural Network Algorithm

In Figure 2, an object-geometry type image was used, and we got a better output from the patch-based algorithm as compared to the Neural Network based algorithm [2]. This is because the neural network algorithm so far is limited to specific cases where the photo is of scenery, portraits, etc. and the reference image is an artistic painting. But when we give the same algorithm with an image that is realistic and gives a 3-Dimensional effect and give a basic geometrical shaped image as a reference image then we did not get the desired output.

There have been many trials to overcome such drawbacks. One of the solutions was to work on creating an algorithm that identifies the features from the reference image using Patch-Based Synthesis and apply it to an image that needs to be transformed using the Neural Network Synthesis. This way there will not be any feature loss from the image and the result will also be of good quality.

Image stylization has come a long way and there are still many improvements and findings that need to be made to obtain the desired image with the desired quality.

III. PREVIOUS WORKS

A. Applying CNNs into Image Style Transfer

Before Deep Neural Networks were introduced to image style transfer, non-parametric algorithms were used to synthesize photorealistic natural textures by resampling the pixels of a texture [9]. All these methods can be categorized into two groups. One of them is global methods by matching means and variances of pixel or histogram. Another is local methods which is trying to fit dense correspondences between photos [4]. Unfortunately, they cannot be used in arbitrary scenarios.

In 2015, Gatys et al. first used CNNs to separate representations both of images' content and style and then combine them to reconstruct artistic images of high perceptual quality. This can be seen as the beginning of Neural Artistic Style Transfer which means using Deep Neural Networks in image style transfer [2].

B. Technical improvements

However, there are still some problems needed to be solved in this area. Here are some examples.

First, neural artistic style transfer cannot preserve photorealism. Therefore, to gain photorealistic images, post-processing techniques are required to refine the outputs, called photorealistic image stylization. The goal of photorealistic image stylization is to translate an image from one domain to another. It can be used not only in image translation but also in Style Transfer [4].

Second, neural style transfer is no longer limited to static scenarios, it is also used in dynamic situations, such as videos that require real-time image processing and consistency of a series of consecutive images. Many neural network techniques such as feed-forward CNN, Recurrent Convolutional Network are used in this scenario [8].

Third, some of these video processing methods have inspired X. Gong et al. for stereoscopic image style transfer. Because directly applying to style transfer on stereoscopic images will cause view inconsistency. To enhance view consistency, many stereoscopic image editing methods were developed. For example, extending single image seam carving, patch-based synthesis framework, layer-based stereoscopic image resizing, and so on. However, these methods have some defects, like specifying tasks or time-consuming. X. Gong et al. proposed a method by using dual path CNNs to both increase view consistency and solve the problems of previous methods and finally produce high-quality stereoscopic stylized images [8].

Last, a minor defect of Gatys et al.'s Neural Artistic Style Transfer is it cannot preserve colours. According to some knowledge, this can be achieved by colour histogram matching and luminance-only transfer [6].

C. Loss functions used in Image Style Transfer

A variety of losses are used in neural style transfer. One common loss function for image style transfer is per-pixel loss. But sometimes it does not meet our needs, hence some advanced methods were developed to gain higher quality images. For example, the per-pixel loss can be modified into per-pixel classification loss. Other loss functions such as per-pixel Euclidean loss, feature reconstruction loss, per-pixel reconstruction loss, or a combination of feature reconstruction loss and style reconstruction loss, are all applied in training a neural network for style transfer [5].

D. Applications of Image Style Transfer

Applications of neural style transfer are everywhere. Neural texture synthesis is to combines textures gained from different images [11]. Photo-realistic facial texture transfer is to transfer one's facial features to another face [13]. Clothes can be also designed by style transfer recombining styles and shapes from different clothes [14].

IV. OUTCOMES AND ACHIEVEMENTS

Image style transfer is a creative method with an impact that is difficult to define. We can find paintings that appear quite like the styles of some well-known artists but measuring a painting's style mathematically is challenging. The conclusion is quite personal. However, the algorithm's design reason is that the low-level Feature Map of CNN is near to the picture texture, while the high-level Feature Map is close to the natural qualities of the image content, which also explains the neural network's black box from a different perspective.

The first is the paper [3] published by Gatys et al. in 2015, which is the pioneering work of image style transfer.

The algorithm is based on a deep neural network, which can transform any picture according to the style of any artist and provides a way to understand how humans create and perceive artistic imagery. The most critical discovery of this work is find that the expression of content and style can be separated in the neural network. So, what is the basis of separation? The article borrows the network structure of VGG19, which has 16 convolutional layers and five pooling layers (using average pooling instead of max pooling).

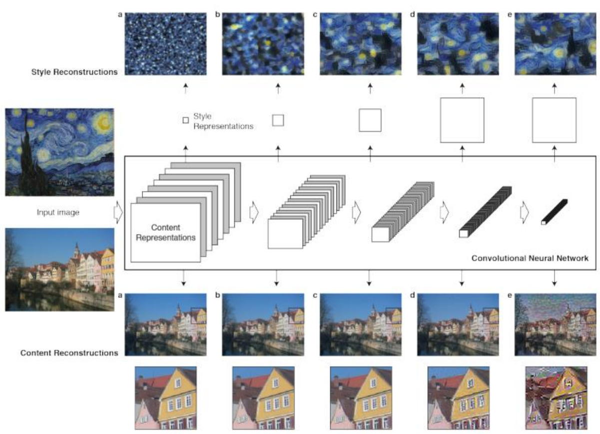

Fig.3.A Neural Algorithm of Artistic Style

The article found that although VGG and other networks are used for discrimination tasks, after multi-layer convolution, the features of the input image are well extracted, and more of the features on the content are retained.

Then Gates published the second paper [7] in CVPR 2016 defined Content Loss (also known as Feature Reconstruction Loss) and Style Loss, and the two are summarized as Perceptual Loss. And it visualizes the extracted image style representation and content representation and defines a successful style transfer- if the created image resembles a style image but reveals the content image's aim and scene.

Deep Photo Style Transfer, Unpaired image-to-image Translation using Cycle-Consistent Adversarial Networks, Stable and Controllable Neural Texture Synthesis, and Style Transfer Using Histogram Losses are improvements in the methods proposed by Gatys et al. in recent years.

Deep Photo Style Transfer [9] focuses on the scope of style transfer being extended to transfer the style of the photo, suppress the distortion of the picture, and turn the transfer process into a local affine transformation of the *colour* space. And solve the problem of style overflow through semantic segmentation.

The second paper [10] point that there is no need for sample pairs of source and target domains, only the images of the source and target domains are needed in the image domain migration task. A very practical place is that the two input pictures can be any two pictures, that is unpaired.

Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses mainly explains why the Gram matrix is unstable in the Style Transfer task and proposes adding histogram statistics to strengthen the constraints.

The above are Slow Transfer based on image optimization, many others are based on model-optimized fast transfer.

Perceptual Losses for Real-Time Style Transfer and Super-Resolution [5] uses perceptual loss to train the neural network. The enhanced characteristics retrieved from the pre-trained network are used in this perceptual loss equation. The approach is more reliable than per-pixel loss, and it can also achieve real-time testing.

High-Resolution Network for Photorealistic Style Transfer [12] uses a high-resolution network as the image generation network. Unlike other systems that lower resolution and subsequently recover it, our generation network keeps the same high resolution throughout the process. High-resolution subnets can continually receive information from the low-resolution network by connecting high-resolution networks to low-resolution networks in parallel and frequently multi-scale fusion. This allows our network to reject less information from the picture, resulting in images with a more complex structure and less distortion, both of which are important for visual quality.

V. COMPARISON AND EVALUATION OF METHODS

Over the years, there have been a lot of methods proposed for the transfer of style to the desired image. Here we are going to briefly compare and evaluate a few of the interesting ones.

A. CNN for Image Styling

Gatys et al. in 2016 proposed the use of a pre-trained Convolutional Neural Network, specifically the VGG-19 network, for downscaling the input image and then reconstructing the same image with the features of the style image along with keeping hold of the contents of the input image [7]. The content image or the input image coupled with the style image is fed as inputs for the VGG network. The content representation at each stage of the convolution network is encoded at each layer based on the filter responses to the image. The feature map that matches the original image from one of the layers (mostly the lower layer of the VGG network) is used to compute the content loss. The stylized image is fed to the VGG network where the content representation from the stylized image is used to determine the content loss by computing their mean squared error. This process is repeated until the content loss is minimized.

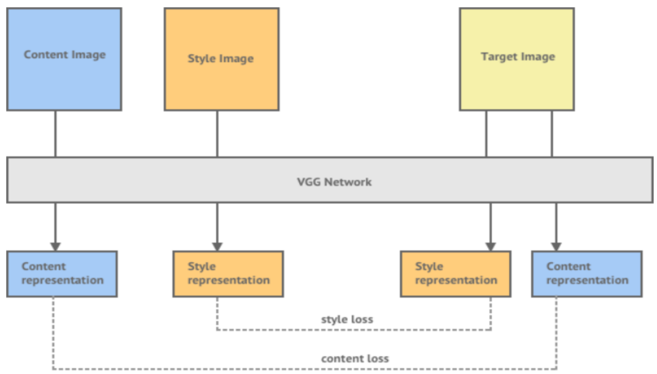

Fig. 4. CNN for image style transfer

The style loss is computed in a similar fashion as that of the content loss wherein the style representation of the style image from the VGG network is compared to the style representation of the stylized image (Target image) on basis of the mean squared error of their gram matrices thus ensuring the styles are same in both the images.

B. *Perceptual Losses for Style Transfer*

Johnson et al. in 2016 proposed a time-efficient technique of image style transfer with the use of a feed-forward network approach [5]. In this paper, an image transformation network that uses encoder-decoder architecture is used to reduce the image, extract objects, and reconstruct using up sampling with the desired style. When the image transformation network gives an output image, it is fed to the loss network which here is the VGG-16 pre-trained network that also uses the content image and the style image as inputs. The loss network uses the perceptual loss function to determine the feature reconstruction loss for measuring the differences in the content image to the output image using the Euclidean distance between them. Similarly, the perceptual loss function also computes the style reconstruction loss using the squared Frobenius norm of the difference between the Gram matrices of the output image and the style image. The total loss is then used to update the weights of the image transformation network.

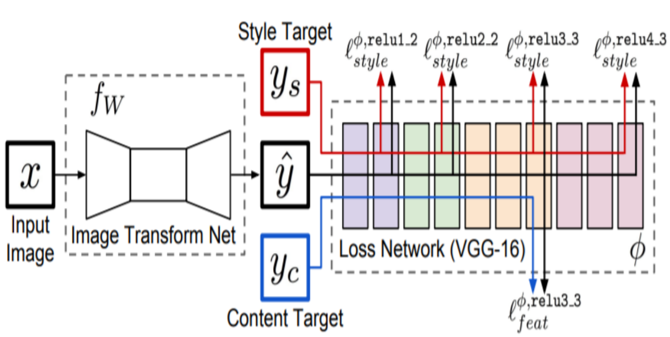

Fig. 5. Feed forward Network for image style transfer

C. *Photorealistic style transfer using high-res network*

Ming Li et al. in 2019 proposed a style transfer technique like the Johnson et al. paper but instead of using a normal encoder-decoder architecture, the paper proposed the use of a high-resolution generation network that stylizes the content image and produces the output image [12].

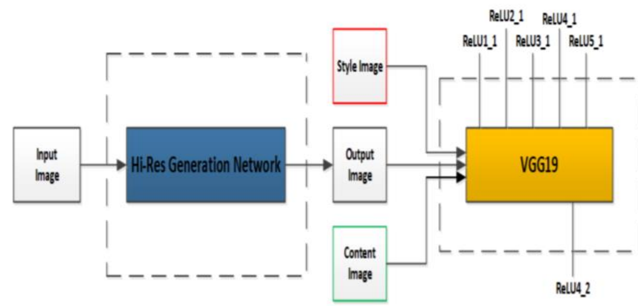

Fig. 6. High-resolution network for photorealistic Style transfer

The loss network used is the pre-trained VGG19 network, which performs better than the VGG16 network, calculates the total loss. The total loss is calculated from the content loss of the Euclidian distance of the content image and the output image and the style loss from the Euclidian distance of the gram matrices of the style image and the output image. This loss is then fed to the high-resolution network to update its weights.

D. *Style transfer in stereoscopic image*

Xinyu Gong et al. in 2018 published a paper that explained the method of style transfer for stereoscopic images. The method not only styled the images but also provided view-consistent results [8]. The proposed model consisted of a dual-path style transformation network and a loss network. The transformation network took both the stereoscopic images and with the help of an encoder and a feature aggregation block, produced the feature aggregated images. The decoder then upscales the images and produces the output stereoscopic styled images. The loss network computes the perceptual loss like the Johnson et al. and a multi-layer view loss is computed which helps the network to maintain the view consistency. The multi-layer view loss consists of an image-level view loss which constrains the outputs to be view-consistent, and the feature-level view loss that maintains the feature maps in the stylizing network to be consistent.

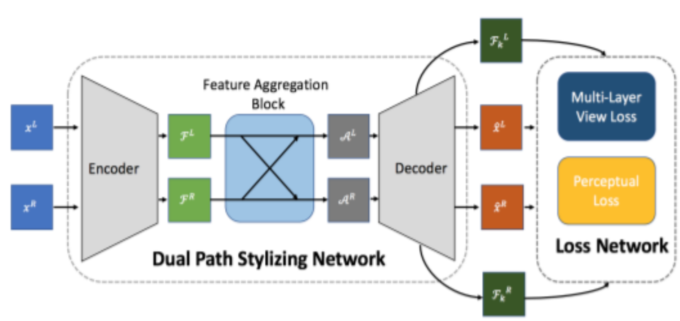

Fig. 7. Stereoscopic Image Style Transfer

VI. INDICATION OF OUTSTANDING OR UNSOLVED ISSUES AND PROBLEMS

Along with the development of machine learning and the progress of neural network technology, the technology of image style transfer is also becoming more and more mature. The image which is processed even makes people think it is a real painting by a master. However, the traditional algorithm transforms the colour of the original photo at the same time, which may change the appearance of some scenes in undesirable ways. Nowadays, Gatys et al.'s technology [6] allows transferring only the texture of the style image, leaving the background colour of the content image.

Fig.8: Keeping the content color's style transfer

Furthermore, Style transformation neural network module through now mature technology, constantly reducing some of the weight of the results have little impact, from the size of 100-200M down to the current size of 20-100K, greatly reducing the CPU processing time and retain most of the image effect [15].

Image style transformations are also widely used in photos. This technique [4] is not the same as the traditional image transfer that simulates a famous painter's style, because the output style must maintain the photorealistic, without distortion parts. For example, transferring a day photo into a night photo (Figure 8).

Fig. 9: The original photo and the converted photo

However, this technology [4] needs to smooth the image to ensure the photo feel, so some surface detail and patterns cannot be well converted (Figure 9).

Fig 10: The resulting photo loses the detail of style

There are also some technical limitations of image style transfer. The dimensionality and the number of units in the Convolutional Neural Network grow linearly with the number of pixels, which means that transferring the high-resolution photo will cost much time [7].

The image style transfer stability of Gatys et al. methods cannot be guaranteed. The image does not get a good effect after conversions, such as contrast and sharpness. In many cases, it is necessary to continuously adjust the parameters to get a better effect [11].

VII. Conclusion

To conclude, image stylization has seen many phases. The first phase is the traditional method of being inspired by brush strokes in a picture. Then came patch-based synthesis, which creates a new image using the textures from the exemplar image and the primary image. It could not use an abstract image as its exemplar image, hence, to accommodate this drawback, the current era focuses on Neural Network Image Stylization. This allows us to take an abstract image as an exemplar image and transfer this style to the scenery. Papers in recent years have made it possible to generate images with finer structure and less distortion, which is essential for visual quality.

Gatys et al. first used CNNs to separate representations of image’s content and style and combine them to reconstruct artistic images. This paper proposed the use of a pre-trained Convolutional Neural Network, specifically the VGG-19 network, for downscaling the input image and then reconstructing the same image with the features of the style image along with keeping hold of the contents of the input image.

But there are still some problems left. The image style transfer stability of Gatys et al.’s methods cannot be guaranteed. The image does not get a good effect after conversions, such as contrast and sharpness. The dimensionality and the number of units in the Convolutional Neural Network grow rapidly with the number of pixels. This means that computers with stronger computing power are required, which is a big test for current hardware devices.

Image Style Transfer is certainly very interesting and has a lot of potentials since it is not now also used in many fields such as Neural Texture Synthesis, Photo-realistic Facial Texture Photo Transfer and clothes designing.

References

[1] Efros, Alexei A., and Thomas K. Leung. "Texture synthesis by non-parametric sampling." Proceedings of the seventh IEEE international conference on computer vision. Vol. 2. IEEE, 1999.

[2] Gatys, Leon, Alexander S. Ecker, and Matthias Bethge. "Texture synthesis using convolutional neural networks." Advances in neural information processing systems 28 (2015): 262-270.

[3] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." arXiv preprint arXiv:1508.06576 (2015).

[4] Li, Yijun, et al. "A closed-form solution to photorealistic image stylization." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[5] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European conference on computer vision. Springer, Cham, 2016.

[6] Gatys, Leon A., et al. "Preserving color in neural artistic style transfer." arXiv preprint arXiv:1606.05897 (2016).

[7] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[8] Gong, Xinyu, et al. "Neural stereoscopic image style transfer." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[9] Luan, Fujun, et al. "Deep photo style transfer." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[10] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

[11] Risser, Eric, Pierre Wilmot, and Connelly Barnes. "Stable and controllable neural texture synthesis and style transfer using histogram losses." arXiv preprint arXiv:1701.08893 (2017).

[12] Li, Ming, Chunyang Ye, and Wei Li. "High-resolution network for photorealistic style transfer." arXiv preprint arXiv:1904.11617 (2019).

[13] Kaur, Parneet, Hang Zhang, and Kristin Dana. "Photo-realistic facial texture transfer." 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019.

[14] Ganesan, Ashwinkumar, and Tim Oates. "Fashioning with networks: Neural style transfer to design clothes." arXiv preprint arXiv:1707.09899 (2017).

[15] Medium. 2022. Creating a 17 KB style transfer model with layer pruning and quantization. [online] Available at: https://medium.com/@jamesonthecrow/creating-a-17kb-style-transfer-model-with-layer-pruning-and-quantization-864d7cc53693 [Accessed 12 January 2022].