From Photo to Art - Style Transfer

Chengzhuo Wang, Wendi Kuang, Xuchen Wang, Zitong Zhu

Abstract

Transferring the style of an existed artwork to a selected image has been a hot topic for the past few years as it can be used in gaming, virtual reality, and image editing. In this project, we investigated the effects of the content similarity between the two source images on the transferring quality using a quantitative evaluation approach. The quantitative evaluation method utilized the content loss and style loss to describe the quality of a style transfer. It is discovered that when both the style and content images possess similar contents and patterns, a better style transfer is achievable. While, in contrast, when the two sources images possess distinct objects (or contents), the final image (i.e., the synthesized image) tends to be in bad quality. Therefore, it is concluded that the current style transfer model may not be suited to transfer the style of a given artwork to any images.

1 Introduction

The pursuit of spiritual wealth by the human can date back thousands of years, and painting, as a popular form of art, has always been appealing to people across the world (Jing et al., 2019). Ever since the mid-1990s, numerous studies have been proposed to understand the art theories behind famous artworks (Jing et al., 2019). With the advancement of machine learning and deep learning techniques, Gatys et al. (Gatys et al., 2015), for the first time, proposed the idea of using *Convolutional Neural Networks (CNNs)* to migrate popular painting styles to any natural images. Later, this classic work led to the foundation of a new work field called *Neural Style Transfer (NST)*. This style transfer method can render the content image in the style of the style image (Yeh et al., 2018).

So far, most studies focused on modifying the model to improve the quality of the image transfer. However, in practice, numerous factors could affect the final results, like the complexity of objects in the content image and the painting style of the style image. Moreover, the majority of the existed studies only used qualitative approaches, which normally involve the deployment of user studies, to evaluate the transferring quality. The quality of this kind of methods heavily depends on the amount of information collected. If the user responses are not sufficient, the results could be biased or lack of generality. Another drawback of this approach is the lack of objectiveness of the results. Not only each individual could have his/her own opinion about the transfer quality, but also that people tend to prefer a particular type of work over another during a specific period.

In this project, a similar *CNNs* model, as discussed in the work of Gatys et al. (2015), was reproduced. Then, a quantitative evaluation method was proposed to inspect the quality of a style transfer. Lastly, the effect of image content similarity on the quality of the final synthesized image was studied by checking the overall loss change due to image content variation. In our project, we used two losses, namely the content loss and the style loss, as our evaluation metrics. The final image with high content loss and high style loss has bad transfer quality; while the one with low content loss and low style loss has good transferring quality. By proposing a quantitative evaluation approach and studying the effect of content similarity, we believe this work will help us to gain more insights about neural style transfer using CNN model.

2 Problem Definition and Algorithms

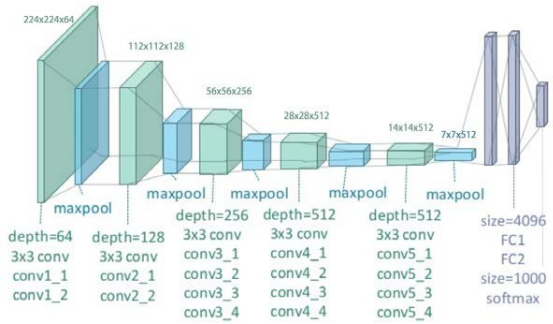


Figure 1: Illustration of the network architecture of VGG-19 model (Zheng et al., 2018)

2.1 Task Definition

Style transfer is a computer vision problem that takes in two individual images, namely a content image and a style reference image, and renders them together so that the resulting output image preserves the primary elements in the content image, but appears to be painted in the style of the style reference image. It is worth mentioning that the term 'style' only refers to the style of a single painting in the following context rather than the general painting style of an artist. The term 'content' is used to indicate the objects within a single style image, as an artist normally created many paintings that may have various objects.

Many methods have been developed by computer scientists for the style transfer problems: neural style transfer (NST) model employs neural network to power the transformation; fast neural style transfer improves the efficiency by using a single feed-forward pass, and state-of-the-art style transfer models can learn multiple styles via one model. However, models and algorithms only play a partial role in the problem. The input is another important factor. Even if we use the same model, different input images may result in different rendering images. In this project, our primary goal is to figure out whether the content similarity between the source images would affect the generated image.

We used a neural style transfer model to transfer the style. One of the two input images of the model is the style reference painting, while the other one is a landscape image. The output of this model is a newly generated image that combines the two input images. We then quantitatively compared resulting images using different content images while keeping the style image the same. By doing so, we can develop a better understanding of how the content similarities of the two source images can affect the whole style transfer process.

2.2 Algorithm Definition

The style transfer problem requires two parts: the pre-trained feature extractor and the neural transfer network. As suggested by Gatys et al., we used VGG19 (Figure 1) as our feature extractor, which is pre-trained on the ImageNet (Gatys et al., 2015). Some layers of VGG19 extract the content of the image, while the others extract the style and the texture of the image. Figure 2 is a image transfer example, and Figure 3 presents two example intermediate layers generated during the transferring process. Figure 3a is the outputs of the content layer, detecting the edges of the wave; Figure 3b is the outputs of the style layer, indicating the brush stroke and the patterns. Similar outputs mean some degrees of similarity between the two images. In this way, the extractor enables us to compare the content and the style of the two images by producing outputs at content layers and style layers.

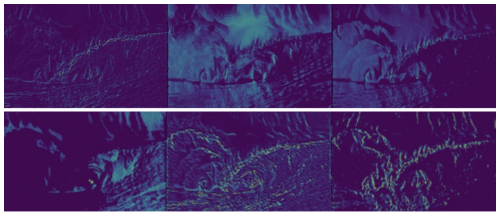
Figure 2: Example of style transfer from surfing image to grey, block-wise style (Note: left: style; middle: content; right: final)

The second part of the NST model generates a stylized image. The architecture of the whole modeling process is illustrated in Figure 4. At the beginning of training, a style image, a content image, and a randomly generated noise image are run through the pre-trained feature extractor, and the outputs at various layers are saved for later comparison. The content loss was measured by calculating the mean square error between the synthesized image and the original content image; while the style loss was measured by calculating the gram matrix between the stylized image and the original style image. The total loss is, therefore, a combination of the content loss and the style loss and has the expression as: $L_{total}(S, C, G) = \alpha L_{content}(C, G) + \beta L_{style}(S, G)$. By weighing

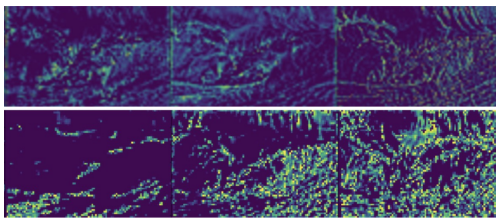
the two losses differently, the output image can render different degrees of stylization (NST). Figure 5 is an example showing the effect of weights. The larger the weight of style loss has, the synthesized image appears more like an art. However, at the same time, it preserves less content from the original content image. During each iteration, the VGG19 model freezes, and only the output image is optimized. (Ritul, 2018) After changing the stylized image according to the derivative of the loss function, the three images again ran through the feature extractor to compute the loss. This process was repeated several times until a certain prespecified threshold has been reached.

2.3 Expectations

Our hypothesis for this project is that the content similarity between the style image and the content image has a certain effect on the generated image. Higher content similarity leads to a better final image. In other words, the synthesized image will have the style of the style image and preserve the contents within the content image. For example, we believe that the style of *Sunflowers* by Vincent van Gogh can be transferred better to a flower photo than to a dog photo. Numerically speaking, we expect the stylized image from the flower image has lower content loss and style loss than the stylized image from the dog image. Intuitively, if the content image and the style image have higher content similarity, the model could copy the style of the object and apply directly to the content image. On the contrary, if the content similarity is low, the model would have actually to learn the texture and apply it to new elements accordingly.



(a) Outputs from the content layers



(b) Outputs from the style layers

Figure 3: Intermediate layers of VGG19 model of style transfer for surfing image

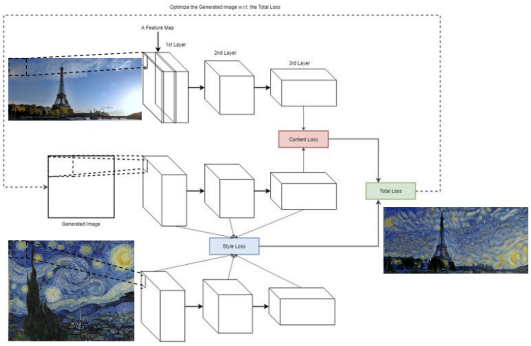


Figure 4: Architecture of transfer network that computes content loss and style loss and optimize the output image

In order to test our hypothesis, the same style reference image was used to transfer the styles for two categories of content images. This experiment was conducted to see whether the content loss and the style loss between the two categories are distinct. Then, a different style image, whose content is not close to any of the content images, was used to transfer the styles for the same two categories of content images to see whether the results of the two categories are mixed up. Another experiment was performed to show that it is the content similarity of the style image to the content image rather than the content image itself that causes the segmentation of the data points. If the results of both experiments meet the expectations, we could

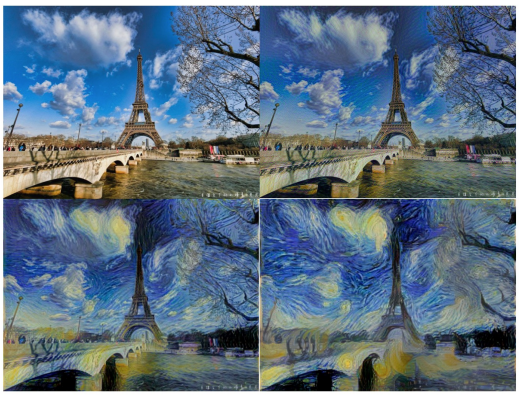


Figure 5: Transferring Eiffel Towel to Starry Night painting style with different ratio of the content loss weight, α, to the style loss, β. The upper left image used 1:1 ratio; the upper right image used 1:100; the lower left image used 1:10000; and the lower right image used 1:1000000

then conclude that the content similarity can be a critical factor in the style transferring process.

3 Experimental Evaluation

3.1 Methodology and Data Description

As discussed in Introduction, one of the goals is to propose a quantitative measure for the quality of the final synthesized image. In order to do so, both the style and content losses were extracted from the loss function used to update the neural network and used for the evaluation. The final style loss after all iterations was marked as 'x,' and the final content loss was marked as the 'y.' In this setting, each transferred image will be a single dot within a scatter plot for the losses. As indicated by the double-sided arrow shown in Figure 6, the images lie in the bottom-left region, where both style and content losses are low, are classified as the ones with high quality; while images stay in the top-right corner are the ones with low quality.

As for the data, three image sets were used. The first image set included several artworks/paintings and was used as the style images. The second image set includes multiple images that are identified to have similar contents as the style image. The last image set includes more images that have very distinct contents as the style image. These image sets were explicitly selected such that a systematical study on the effect of content similarity on the final synthesized image can be conducted.

In order to prove that this content-style loss characterization is reasonable, in section 3.2, some transferred images with different distances to the origin point in the scatter plot (e.g., Figure 6) would be shown. By observing the details and performances of transferring, the expectation in section 2.3 can be analyzed and evaluated through the content-style loss characterization. Then, with the help of the control variate method, the influences of the style images would be shown.

3.2 Results and Discussion

For this study, two distinct, but interesting, cases were investigated. In case one, the content images were selected such that they all possess similar contents as the style image (i.e., building). In total, 20 content images were used in this part, and the results for this case are shown in Figure 6. According to Figure 6, the two categories of content images generated very distinct behaviors in terms of losses. Most building content images, which are very similar to the style image, are associated with relative smaller losses compared to the glacier content images. This result clearly indicates that the content of the content images could significantly impact the quality of the final image.

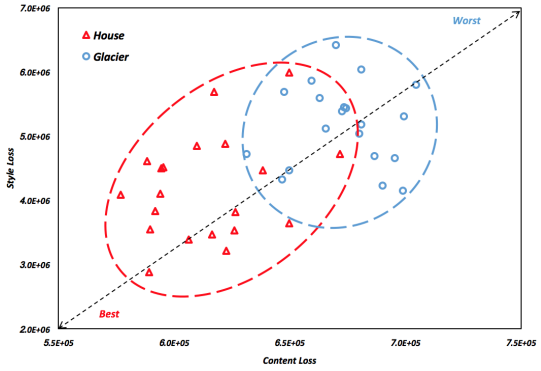


Figure 6: Losses for style transfer using similar content and style images

In order to better illustrate the results, two sets of images for building and glacier content images are displayed in Figures 7 and 8, respectively. For the synthesized building image, the building walls all possessed uniform and similar color as in the style image. However, in the synthesized glacier image, the glacier image rendered some block-shaped color patterns, which makes the whole image distorted.

Figure 7: Similar content and style images (Note: left: style; middle: content; right: final)

Figure 8: Distinct content and style images (Note: left: style; middle: content; right: final)

More images were selected and shown in Figures 9 and 10. This time both content images are building images with one that has much more

complex structures. As shown in Figure 10, the synthesized skyscraper, despite also being classified as building like the style image, lost a significant portion of details, especially for the windows. Therefore, apart from the content, image complexity may also impact the final result potentially. Overall, visually speaking, the style transfer is more successful using the building content images than using the glacier content images, which is in line with the general trends observed in Figure 6.

Figure 9: Style transfer of building images: simple building (Note: left: content; right: final)

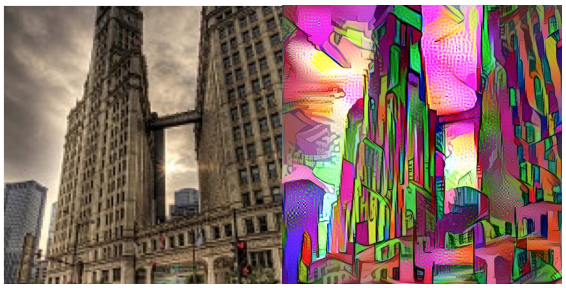


Figure 10: Style transfer of building images: complex building (Note: left: content; right: final)

It is worth mentioning that the image examples presented above (i.e., Figures 7, 8, 9, and 10) are ranked 1^{st} and 20^{th}, and 3^{rd} and 10^{th}, respectively, in terms of the Euclidean distance to the origin point in Figure 6. Therefore, through visual inspection, it can be concluded that the one that is closer to the origin has better performance, and the transferring performance can potentially be evaluated for each image using the Euclidean distance to the Origin point.

Furthermore, another experiment using a style image that has a very distinct content (i.e., forest) from the two content image categories (i.e., buildings and glaciers) was performed to provide more information. The results are shown in Figure 11.

According to Figure 11, the points are not segmented well in this case, which is due to the content difference between style and content images. Therefore, different style images cannot ensure the performance of style transfer.

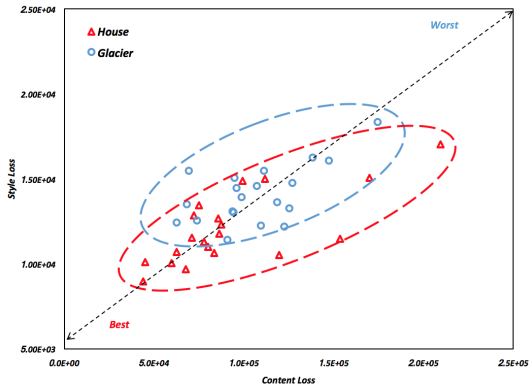


Figure 11: Losses for style transfer using distinct content and style images

Overall, by analyzing the two cases, the content similarity is proved to be a crucial factor during a style transfer process. Using content-wise similar style and content images can lead to better transferring performance than using distinct ones.

3.3 Data Analysis

Through the experiments conducted in sections 3.1 and 3.2, the images in the data set have clear and simple categories, which allows the model to detect the features in the style image at ease. Transferring human portraits was also investigated. When the style image is complex, the transferring performance is not as good as when the style image is simple. Although the two style images are both human (see Figure 12), the famous *"Mona Lisa"* style was not migrated successfully.

For instance, when an abstract human painting is chosen as the style image, the performance is better than the famous painting *"Mona Lisa"* is selected when applying to another portrait. According to Figure 12, it is evident that the Mona Lisa style has bad performance since the significant portion of the details within the content image was lost in the Mona Lisa style. This information loss arises from the fact that facial features are very difficult to capture and analyze from the style image. In the Picasso abstract painting, since it is simpler, the outline and facial details are detected better. Despite this, it only superimposed the color onto the facial area rather than making her face as stereoscopic as the style image.

Figure 12: Portrait style-transfer example (Note: left: style; middle: content; right: final)

4 Related Work

The majority of the existed studies used user studies when evaluating the quality of the style transfer. Though being widely accepted and could be directly perceived through the senses, this evaluation method is highly subjective, and the bias during the information collection process is also inevitable. The quantitative method involving the two loss values in this paper is designed to compare the final image with both the content and style images. For each transferred image, the quantitative method can specify the final content loss and style loss. It transfers the difference between two images from abstract feelings to absolute quantity values and, therefore, can avoid subjective bias from human beings.

5 Future Work

In the future, we plan to research more evaluation methods for image processing. We could combine the quantitative evaluation method proposed in this project with the qualitative user evaluation widely used in the literature. We could design a survey letting users pick the best-transferred picture in their mind among 20 transferred pictures and check whether the results match with the quantitative characterization (i.e., the loss measure used in this study). We can obtain preference data from Amazon Mechanical Turk (AMT) workers to generalize the results. We are also interested in understanding and improving the quantitative evaluation method proposed by Yeh et al. (2018). In their work, the transferred images were evaluated quantitatively using a self-defined "effectiveness/coherence" score (Yeh et al., 2018). This parameter possesses the similar idea of measuring the distances between the synthesized image and the two input images.

6 Conclusion

In this project, a neural style transfer model was built to synthesize a stylized image from a content image and a style image. The final loss for each transferred image was used to evaluate the transfer quality. The results of the experiments meet our expectations, and three key findings based on the experiments were summarized as below.

- The content of the style image can significantly affect the quality of the final image. The synthesized image, which used similar style and content images, is better than its counterpart using object-wise distinct style and content images.
- When using images with rich details (e.g., portraiture) as style images, the synthesized image could still be in low quality. This is because the algorithm is not efficient to capture some subtle but noticeable features.
- Even if using the same model with identical settings and input images, there are still many other factors that may impact the synthesized image. Two possibilities that were discovered in this study include the choices of the hyperparameters and the complexity of the content image.

References

Style transfer guide.

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. 2019. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*.

Ritul. 2018. Style transfer using deep neural network and pytorch.

Mao-Chuang Yeh, Shuai Tang, Anand Bhattad, and David A Forsyth. 2018. Quantitative evaluation of style transfer. *arXiv preprint arXiv:1804.00118*.

Yufeng Zheng, Clifford Yang, and Alex Merkulov. 2018. Breast cancer screening using convolutional neural network and follow-up digital mammography. In *Computational Imaging III*, volume 10669, page 1066905. International Society for Optics and Photonics.