# Combining CycleGAN and perceptual loss for image style transfer: Double training process

**Hanjun Li**

School of Management, Lanzhou University. Lanzhou, Gansu, 730030, China

lihj2021@lzu.edu.cn

**Abstract.** Image style transfer is an important field within the broader area of image processing. While existing style transfer models have addressed certain challenges, there remains room for improvement in areas such as color restoration, edge detection, and the ability to effectively transform lighting conditions. Additionally, there is a lack of research exploring the features and properties of these models. To address these limitations, this paper proposes a double training method that combines CycleGAN with the minimization of perceptual loss for style transfer. The proposed method is tested and compared to existing models, resulting in impressive performance improvements, particularly when the style image background is white. These comparative results not only validate the effectiveness of the double training method with CycleGAN, but also provide valuable qualitative insights into CycleGAN and perceptual loss. Drawing from these conclusions, this work further proposes a new model hypothesis that builds upon the double training method and includes background-object segmentation and background whitening. This novel approach aims to enhance the effectiveness of style transfer by allowing objects to maintain their original forms while transferring the style of the source image to the target image.

**Keywords:** CycleGAN, Image style transfer, Perceptual loss.

## 1. Introduction

Image style transfer has always been a popular and significant task in image processing. Style transfer refers to the processing of transforming the information the visual feature of one image to emulate another target image style, while preserving its original content [1]. Thus, this task results in a novel composition that combines the content of one image with the style characteristics of another. In image style transfer, the desired output image should meet two main requirements: content preservation and style transfer. Content preservation aims to maintain the semantic information and spatial structure of the input image in generated image. Style transfer aims to capture and incorporate the style of the style reference image, including its color, texture, brushstroke and other visual features in the output result. However, achieving these two goals both appears to be a significant challenge, as many methods of this task shows artifacts, distortions, and other quality problems in their generated image. The consideration of time efficiency and flexibility of the model is significant, alongside the evaluation of output quality. These three elements compose the basic principle of evaluation of style transfer model. To balance the trade-off, or even find a solution which turns it into a non-trade-off problem, and retaining visual fidelity and coherence in generated images becomes the main challenge. The key to tackle this challenge is how to extract sufficient good-quality content and style information while train the network to make good use of them. Therefore, many techniques developed under the configuration of neural network have achieved excellent breakthroughs [2,3].

Among them the Neural Style Transfer (NST) developed by Gatys et al. in 2015, underlay neural network pipelines based on feature extraction from pre-trained deep network [4]. The idea of feature extraction significantly improves the content and style information quality comparing to the Pix2Pix before [5]. Another remarkable model being able to implement image style transfer task is CycleGAN [6]. It provides a method of transforming content image into the domain of style image by constructing the mapping between two domains. The introduction of consistency loss in CycleGAN puts constrains on the mapping, thus preserving the semantic and structural consistency between origin image and the reconstructed image. It enables unpaired GAN image translation for the first time.

However, the NST method and CycleGAN both have problems, especially in visual expression in style characteristics of generated output. The NST method may cause artifacts, distortions and other unsatisfying rendering in brushstrokes and textures. And CycleGAN shows possibility of failing to construct appropriate mapping between content domain and target domain. It also lacks flexibility in some degree, as the construction of target domain requires feeding certain-style training sets to learning the mapping of two domains [7].

As a result, this work proposes a solution combining different loss functions from this two main method, aiming to take unique advantages from them in extracting feature information and preserving the overall coherence of visual representations. This work adopts the technique of Double-perceptual (DP) loss invented in Enhanced super-resolution generative adversarial network with adaptive dual perceptual loss (ESRGAN-DP) to improve extracted feature information quality, which shows better effect than the original single perceptual loss extracted from VGG-19 model only [8,9]. Moreover, CycleGAN module is leveraged to maintain its capability of constructing holistic mapping relationship, reducing the risk of incoherent content-style expression in the generated output. Therefore, adversarial loss and consistency loss are adopted in CycleGAN model, too. This work proposes two ways of integrating perceptual loss into CycleGAN style transfer. As the weight distribution of multi-loss function can pose critical influence on the output, this work also explores the best weight coefficient set from the geometrical/ mathematical perspective.

In conclusion, main contributions of this paper are listed as follows:

1. double training method of combining perceptual loss with CycleGAN model for style transfer task

2.A comparative analysis of the results for double training was conducted to obtain a qualitative analysis of the features and characteristics of CycleGAN and perceptual loss. Based on this analysis, potential ideas for improving the output quality are proposed.

## 2. Method

This paper aims to enhance the effectiveness of style transfer through the integration of adversarial loss, consistency loss from CycleGAN, and perceptual loss. Moreover, the structure of CycleGAN model, the way of balancing different models' loss magnitude and how to choose hyper-parameters should be considered with much attention. Drawing upon the distinctive pipeline and training structure of CycleGAN and Gatys's method, this paper presents two hypothetical approaches for experimental evaluation.

### *2.1. Adversarial Network Structure*

Adversarial network creates a co-inspiring way in order to improve the performance by making the generator and discriminator competing with each other, as the generator needs to generate outputs real enough to deceive the discriminator, and the discriminator needs to distinguish whether outputs are real or not as correctly as it could. The CycleGAN model adopts a double-generator-discriminator module. The generator G aims to transform image from domain X to domain Y, and the discriminator DX should make adversarial discrimination between real image input Y and fake Y generated by G. For another pair of generator F and discriminator DY, the process of transforming image from domain X to domain Y is same.

CycleGAN's network focuses on describing the information between two domains through consistency loss, which limits its robustness when the target task image differs significantly (e.g., in spatial structure) from the images used for training. The failure example of CycleGAN’s paper has illustrated this problem [6].

Integrating perceptual loss into CycleGAN can help overcome this problem. Perceptual loss directly calculates the differences between content features, style features, and the image undergoing transfer, making it more capable of capturing specific information, such as spatial structure, texture, or brushstrokes present in the input content and style images.

Based on this hypothetical analysis, a double training pipeline is designed.

The double training method is substituting the generators of CycleGAN for the image transform network in neural style transfer. It means after the CycleGAN network finishes its training, it will play the role of image transform network, into which the input is put, and then minimize the perceptual loss, containing content loss and style loss calculated by content image X and style image Y.Double training may help solve this problem because instead of the information based on certain categories of domains, perceptual loss focuses on more concrete and specific information of input images, which has advantages in running more random tasks.

The datasets used in this paper for training and testing are “apple2orange” and “horse2zebra” from UC Berkeley group from paper and website. “apple2orange” and “horse2zebra” datasets are both divided into “trainA” (995), “trainB” (1019), “testA” (266), “testB” (248). These two datasets all have enough training images size, and their two domains are similar in the big picture, like apple and orange, horse and zebra, but they differentiate in color, texture and other details. This similarity and difference provide good conditions for style transfer tasks.

This work adopts the architecture for generative networks from ESRGAN-DP. The generator network consists of three essential parts: an encoder, a transformer, and a decoder. The encoder contains three convolutions, with inputs sized 256×256. Then 9 residual blocks are leveraged for 256×256 images. Decoder contains two transpose convolutions and a convolution for RGB output. Instance normalization has been used. The discriminator is 70×70 PatchGANs [9].

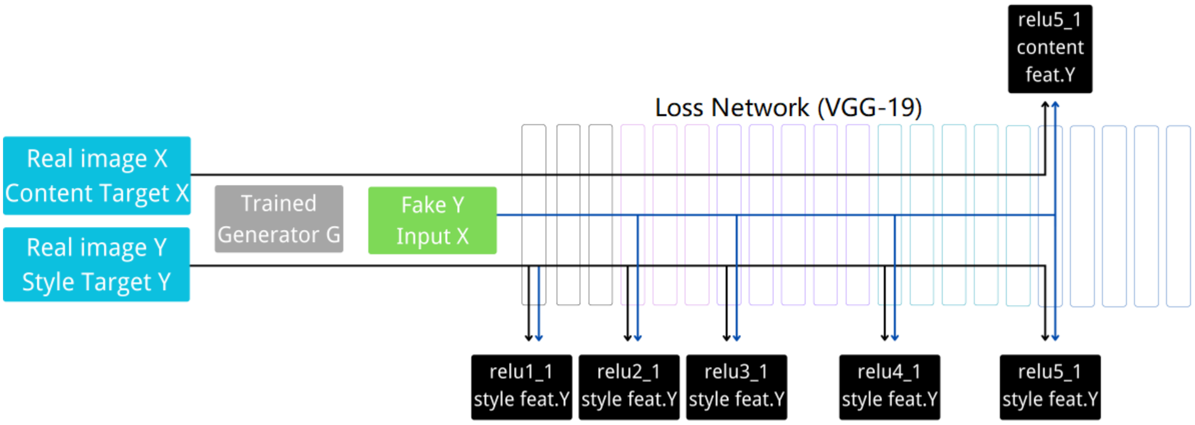

**Figure 1.** Pipeline of double training process (Figure Credits: Original).

### *2.2. Pipeline of Double Training Method*

Figure 1 illustrates the pipeline of the dual-training approach, taking the process of stylizing image X to image Y (e.g., transforming a horse to a zebra) as an example. In the given example, the trained generator G has completed the specified epoch training and can generate fake Y from certain domain X. After selecting the original images X and Y for style transfer, where X serves as the content image and Y as the style image, the perceptual training optimizes the generated fake Y, obtained by feeding X as input to the trained generator G, using the VGG19 pre-trained model for feature extraction and style transfer. Method 1 and method 2 adopt the same set of layers for feature extraction. To be more specific, content layer is block5_conv1 and style layers are block1_conv1, block2_conv1, block3_conv1, block4_conv1, and block5_conv1. Style target is calculated by gram matrix [10].

## 3. Result and Discussion

In Experiment of double training method, based on the horse2zebra dataset, three representative images are selected as the source content images. Their representativeness lies in the fact that the quality of the generated fake Y images, produced by the original CycleGAN with these images as inputs, varies. The first image exhibits a good transfer effect for fake Y, although it has slight flaws in terms of color deviation. The black stripes of the zebra are transformed into more of a brownish color in the fake Y. The second image shows a moderate transfer effect for fake Y, achieving a certain degree of zebra pattern migration, but there are still noticeable areas that were not successfully transferred. The third image demonstrates a poor transfer effect for fake Y, with minimal changes observed.

Among Figure 2, 3 and 4, the style images were manually selected from the testB set of the horse2zebra dataset, consisting of four images. The style images all depict individual zebras, but they exhibit variations in terms of background color, zebra size, and spatial structure. Each of the three different content images was paired with the four different style images, resulting in a total of 12 distinct content-style combinations.

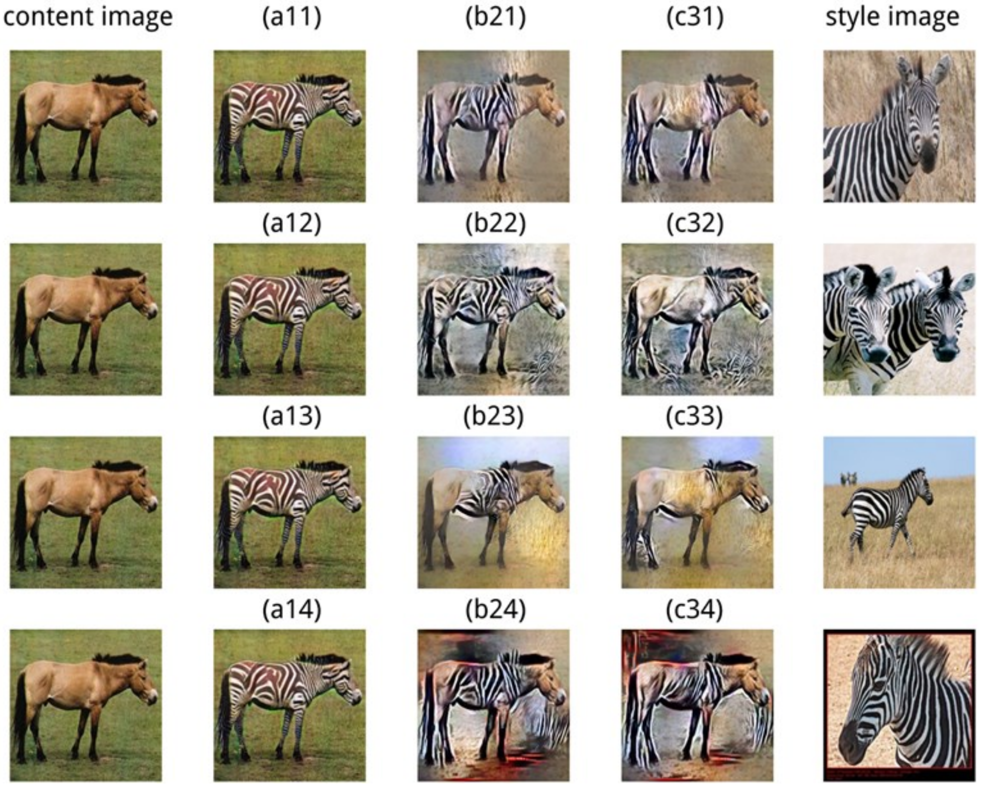

**Figure 2.** Results of double training process of good transfer effect (Figure Credits: Original)

The (a), (b), and (c) sets of each illustrative figure represent the output results of CycleGAN only, CycleGAN+perceptual loss, and perceptual loss only. In the (a) set, the outputs are generated by CycleGAN after certain epochs of training with the input of corresponding content images. In the (b) set, the optimization objective is (a), the CycleGAN output. And then implementing style transfer based on perceptual loss on the objective according to the corresponding content-style combination. In the (c) set, the optimization objective is the content image, leading to style-transferred output for the corresponding content-style combination. These results are analyzed to examine the effectiveness of double training method and to reveal the potential characteristics, properties, and other qualitative hypotheses and conclusions underlying CycleGAN and perceptual style transfer methods.

Analyzing the results in Figure 2, it could be observed that the style transfer effect in the (c) set is unsatisfactory, with only partial regions showing noticeable style migration. However, the fidelity in reproducing the zebra pattern color is relatively high. The (b) set inherits the favorable color characteristics from the (c) set and achieves a larger area of successful style transfer. The extent of successful zebra pattern migration in the (b) set is strongly correlated with the specific characteristics of the style images. For instance, in the case of (b22), which is based on a zebra style image with a white background, almost the entire body of the horse undergoes successful migration while maintaining good color reproduction. Comparatively, the results of (b21), (b23), and (b24), based on style images with non-white background colors, exhibit satisfactory color characteristics but yield effects superior to the (c) set and inferior to the (a) set. The rank of outputs visual quality in Figure 2 could be: (a11)>(b21)>(c21); (b22)>(a12)>(c32); (a13)>(b23)>(c33); (a14)>b(24)>(c24).

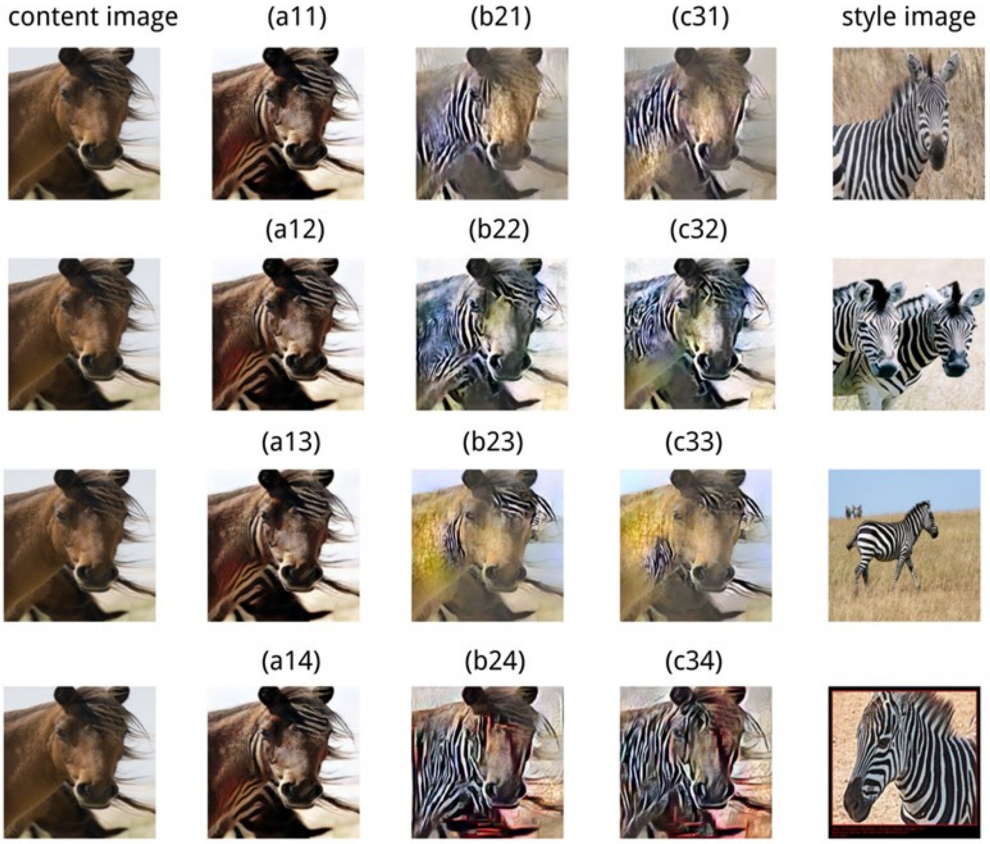

**Figure 3.** Result of double training process of moderate transfer effect (Figure Credits: Original).

Analyzing the results in Figure 3, it can be observed that there are similarities and differences compared to the results in Figure 2. The similarity lies in the fact that the effect of (b22), which is based on a zebra image with a white background as the style image, is superior to (a12) and (c32). However, in the results where the style image has a non-white background, there is no significant difference in the effects between the (b) and (c) sets. The area of successful migration is comparable to or slightly inferior to that of the (a) set. Moreover, the (b) set performs better than the (a) set in color. The rank of outputs visual quality in Figure 3 could be: (a11)=(b21)=(c21); (b22)>(a12)=(c32); (a13)>(b23)=(c33); (b24)=(c34)⩾(a14).

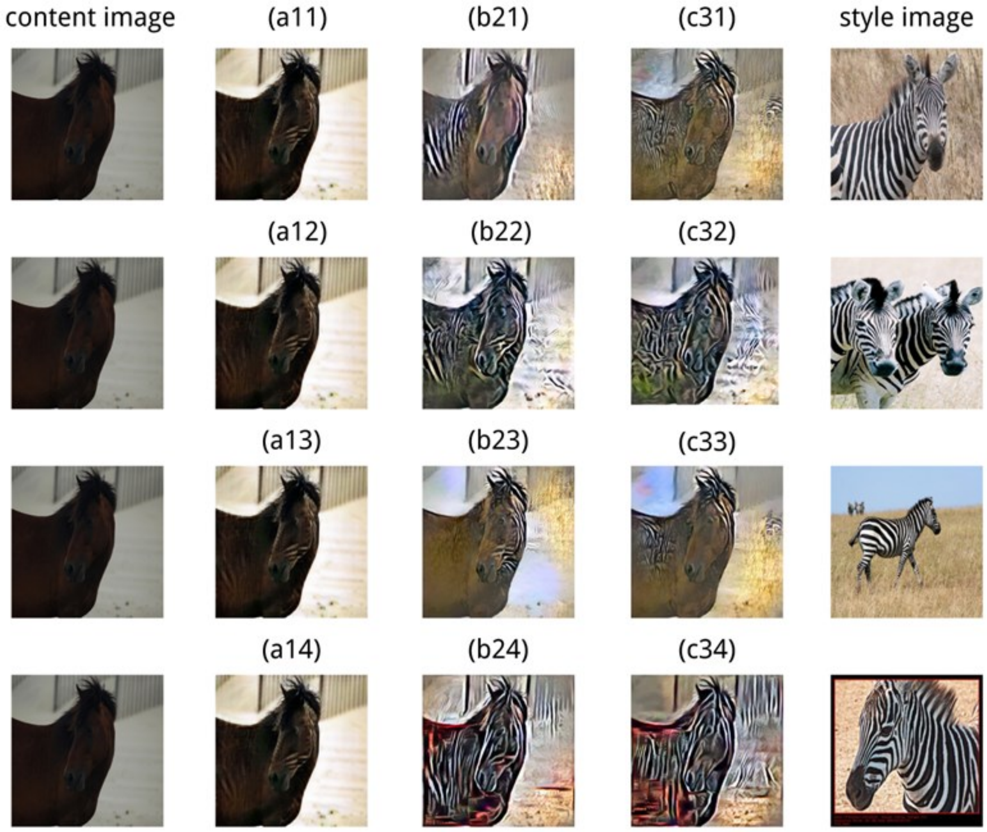

**Figure 4.** Results of double training process of poor transfer effect (Figure Credits: Original)

Analyzing the results in Figure 4, it can be observed that for input images with poor performance in CycleGAN, all three methods, (a), (b), and (c), exhibit unsatisfactory style transfer effects. The (b22) that demonstrated good performance in the Figure 4 and Figure 5 only achieves a larger area of zebra pattern migration in Figure 6, but the stroke, color, and spatial fidelity of the pattern are all poor. On the other hand, (b21) achieves good color and texture authenticity in a small localized area. Overall, the (b) and (c) sets outperform the (a) set in terms of zebra pattern migration clarity and area, but they also achieve good results in other aspects. Possible reasons for this discrepancy could be attributed to the low contrast between the main subject (horse) and the background in the content image, which increases the difficulty of migration. Additionally, the perceptual loss exhibits a characteristic of emphasizing color feature restoration from the style image, as compared to CycleGAN. The rank of outputs visual quality in Figure 4 could be: (b21)>(c31)=(a11); (c32)>(c32) ⩾ (a12); (a13)=(b23)=(c33); (a14)=(b24)=(c34).

Further analysis of the results reveals several aspects in which variables can be controlled for summaries. From the perspective of the variable being the input generated by CycleGAN, i.e.,the fake Y in (a) set, it can be observed that good or moderate quality inputs in the (a) set contribute to the improvement of the (b) set, which combines CycleGAN and perceptual loss through double training. This improvement is primarily reflected in color correction. This suggests that high-quality input images can better leverage the advantages of CycleGAN and perceptual loss. From the perspective of the variable being the style image, it is observed that style images with a white background yield better results for double training method in this study, as seen in the (b) set. This indicates that the perceptual loss method focuses more on extracting overall color and pattern information from the style image, and a white background is advantageous for extracting the zebra pattern, which is the main subject of the image. This implies that for image translation tasks like horse-to-zebra that only require subject transformation, identifying and segmenting the main subject and background of the input image, and whitening the background of the style image to allow the network parameters to learn the subject's pattern features may enhance the output image quality. Although CycleGAN performs well in background and subject recognition, it still struggles with accurately capturing light and shadow edges, resulting in many failures. This background segmentation and whitening approach can aid CycleGAN in improving its performance in judging light and shadow edges.

Overall, double training method in this study demonstrates good performance when the input consists of high-quality CycleGAN outputs and the style image has a white background. Its output quality surpasses that of the individual CycleGAN method or the perceptual loss method alone. Additionally, the perceptual loss method exhibits the characteristic of extracting style features based on the overall style image's color and texture, thereby achieving more pure style extraction effects, particularly with white background style images. On the other hand, CycleGAN excels in learning mapping features from many similar images, and its ability to distinguish between the subject and background is stronger than that of the perceptual loss method. However, it is relatively weaker in color restoration fidelity as demonstrated in Figure 5. Based on the analysis, The following improvement hypothesis are proposed: utilizing background-subject segmentation and background whitening to enhance the pure color and texture restoration capability of the perceptual loss method and further addressing the light and shadow edge misjudgement issues caused by CycleGAN.

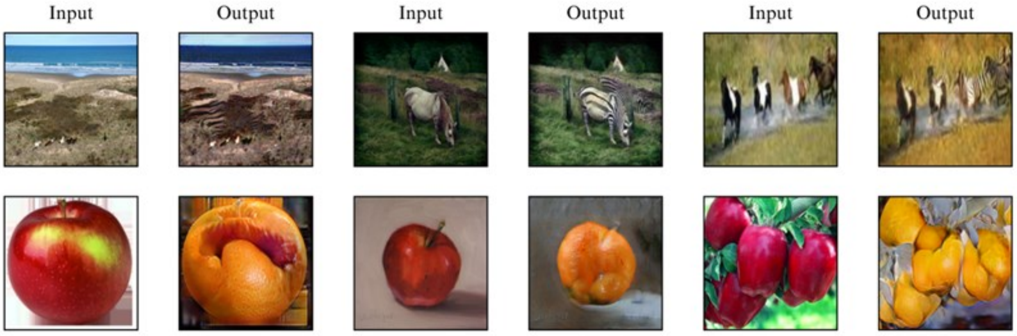

**Figure 5.** Failure examples (Figure Credits: Original).

## 4. Conclusion

This paper presents a conceptual method called double training, which combines CycleGAN and perceptual loss, aiming to explore further improvements in the quality of style-transferred generated images. By comparing and analyzing the results, a qualitative study is conducted on the characteristics and properties of CycleGAN and perceptual loss. Through experiments with double training using different content-style image combinations as input, it is observed that double training surpasses the individual training of CycleGAN and perceptual loss when the style image has a white background. Furthermore, the quality of the generated fake Y images by CycleGAN, which serves as the optimization target, plays a crucial role in enhancing color restoration using the second method. Additionally, the quality of content and style images, as well as background color and style features, also impact the effectiveness of the second method. Comparative analysis reveals that CycleGAN tends to recognize and transfer overall features of the input image, while style transfer methods based on perceptual loss tend to focus on overall recognition and transfer of color and texture in the style image. Based on these findings, this study proposes a style transfer method that incorporates background-object segmentation and background whitening, which effectively leverages the advantages of CycleGAN and perceptual loss. Further research can explore network architectures and parameter settings suitable for this proposed method to achieve even better results.

## References

[1] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2019). Neural style transfer: A review. IEEE transactions on visualization and computer graphics, 26(11), 3365-3385.

[2] Singh, A., Jaiswal, V., Joshi, G., Sanjeeve, A., Gite, S., & Kotecha, K. (2021). Neural style transfer: A critical review. IEEE Access, 9, 131583-131613.

[3] Cai, Q., Ma, M., Wang, C., & Li, H. (2023). Image neural style transfer: A review. Computers and Electrical Engineering, 108, 108723.

[4] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576.

[5] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1125-1134.

[6] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, 2223-2232.

[7] Oh, G., Sim, B., Chung, H., Sunwoo, L., & Ye, J. C. (2020). Unpaired deep learning for accelerated MRI using optimal transport driven CycleGAN. IEEE Transactions on Computational Imaging, 6, 1285-1296.

[8] Song, J., Yi, H., Xu, W., Li, X., Li, B., & Liu, Y. (2023). ESRGAN-DP: Enhanced super-resolution generative adversarial network with adaptive dual perceptual loss. Heliyon, 9(4), e15134.

[9] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[10] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In European Conference Computer Vision, 694-711.