Comparison of Neural Style Transfer and Cycle Generative Adversarial Network approaches for image style transfer

Nemanja Vujadinović
Software engineering and information technologies
Faculty of Technical Science, University of Novi Sad
Novi Sad, Srbija
vujadinovic.sv28.2020@uns.ac.rs

I. Intro

Visual records serve as one of the earliest methods for conveying both conceptual and visual information. Throughout history, painting has evolved, reaching its peak in the modern era. The need for artistic expression persists to this day, bringing for both personal satisfaction and aesthetic fulfillment, as well as for meeting the demands of film industry. Nevertheless, creation of art presents challenges in terms of skill and inspiration and the negotiation of temporal and financial constraints.

The advancement of software has enabled to overcome these limitations. Early algorithms for style transfer [1] [2] often failed to capture the full content of the image to which the style was applied [3]. Neural style transfer (NST) algorithm [3] made it possible to generate higher-quality stylized images. However, the NST algorithm resulted in low-resolution images and required longer generation time [3].

In need of better results, neural networks specialized for image synthesis problems [4] have been developed. Cycle Generative Adversarial Network (CycleGAN) [5] accelerated the process of image stylization, but its inconsistent style transfer posed a problem.

It became evident that each software for style transfer had its drawbacks. If users demand higher image quality, they must accept longer generation time. Conversely, shorter generation time often results in an unsatisfactory visual experience. By comparing different approaches, users can find a balance according to their personal needs and choose a more suitable option for use.

This paper explores and compares models for image stylization based on NST and CycleGAN approaches. The models expect input images of portraits or landscapes. For portrait images, the models are trained to generate images in the Cubism style. In the case of landscape images, the generated image will be in the Ukiyo-e style [6]. The results provide insight into the generated images, highlighting differences in style application. Additionally, the results indicate differences in the models' time and hardware requirements.

II. Method

Our implementation and demo are avaliable at link.

A. Data collection and analysis

The data was collected from publicly available datasets. Cubism and Ukiyo-e style images were obtained from the *WikiArt* dataset. The *CelebA* dataset was used for portrait images, and the *Flickr* dataset was used for landscape images.

The number of images used for training the models was significantly smaller than the total number of images in the public datasets. The main reason for reducing the number of images were hardware limitations. Subsets of the public datasets were randomly sampled.

In new, reduced dataset, there were style images that lacked sufficient distinctiveness. We have manually removed images that were assessed as ones that would not achieve satisfactory results. Portrait images were also manually filtered to ensure an equal gender and age ratio. Detailed data on the number of images in the public and final datasets Table 1.

TABLE I
Number of images in datasets

	Portrait	**Landscape**	**Cubism**	**Ukiyo-e**
Public dataset	202599	1273	383	4036
Our dataset	500	1000	100	1100

B. Data preprocessing

For input to the NST model, images were scaled to 224×224 pixels. The input images for the CycleGAN model were scaled to 256×256 pixels. Due to the smaller size of the training dataset for the CycleGAN model, we applied random cropping and horizontal flipping augmentations.

C. CycleGAN model training

We used 80% of dataset for training set and 20% of dataset for testing set.

Adversarial loss was used as the loss function for the discriminator. For the generator, the loss function consisted of the sum of three independent functions:

- Adversarial loss, with a weight of 1
- Identity loss, with a weight of 0.1
- Cycle consistency loss, with a weight of 10

We used Adam optimizer [7]. Initially, models were trained with a learning rate of 0.001; however, around the 30th epoch, the loss began to stagnate. Generated images were blurry and

style wasn't transfered successfully. By adjusting the learning rate to 0.0002, we achieved significantly improved results.

The model for stylizing landscapes was trained for 100 epochs, while the model for stylizing portraits was trained for 50 epochs. The number of epochs was determined experimentally. We used batch size of 1 in both cases. The models were trained from scratch.

D. Transfer learning for NST model

Transfer learning technique [8] was performed on a pre-trained VGG-19 network [9]. One content layer and five style layers were extracted from this network. Each of the style layers had an equal weight of 0.2.

The total loss function comprised the sum of content and style loss functions. Content loss had a weight of 5, and style loss had a weight of 80. These weights were empirically determined to maintain content integrity while effectively applying the style. We used Adam optimizer with a learning rate of 0.001.

Both models for stylizing portraits and landscapes were trained for 10000 epochs. The number of epochs was determined experimentally. Higher number of epochs caused distortion of the original content, while fewer epochs resulted in weaker style transfer.

III. Results and Discussion

In Figure 1, we show results of transfering Ukiyo-e style to landscapes. The CycleGAN model demonstrated a superior understanding of Ukiyo-e's characteristic traits. In contrast, since NST relies on a single reference image, it struggled to capture the detailed and subtle features typical in Ukiyo-e style. This resulted in less satisfactory outcomes when using NST. Results highlight the advantage of CycleGAN. By training on a diverse set of images, CycleGAN excels in creating a cohesive style rather than relying solely on individual characteristics.

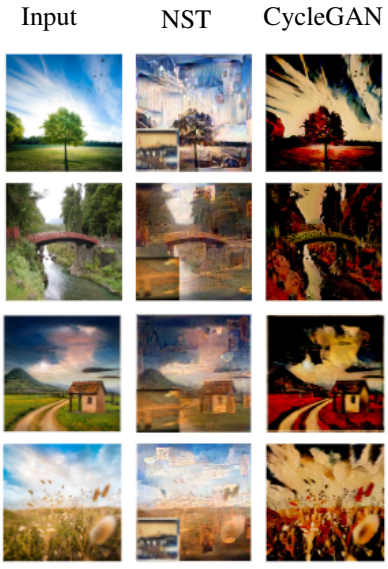

Fig. 1. Style transfer of landscapes into Ukiyo-e style with both approaches

In Figure 2, we show results of transfering Cubism style to portraits. The CycleGAN model maintains the integrity of key features such as eyes, nose, and body. In contrast, NST distorts portraits more extensively in its attempt to match the Cubist style. In this case, we could say that NST provides more visually pleasing results.

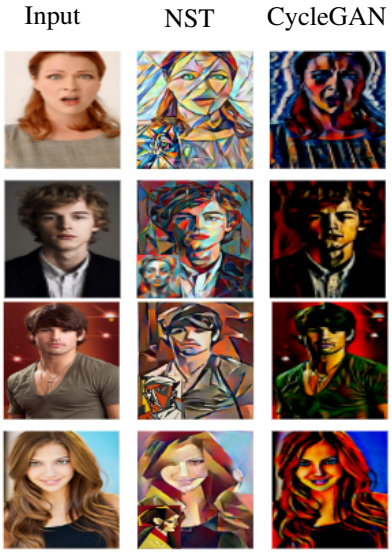

Fig. 2. Style transfer of portraits into Cubism style with both approaches

Both approaches deliver impressive visual results. CycleGAN model excels in realistic results, while NST one brings an artistic touch to generated images. NST model performs less optimally when the distinctive style features in the reference image are not prominent, like in Ukiyo-e style. However, if one style image is enough to define the preferred style transfer, which is the case with Cubism, NST model may produce better results.

Models have been trained on Nvidia RTX 3070 GPU. Unlike NST, CycleGAN requires a training process which demands more time for building a model. However, inference time is in favor of CycleGAN model - 10s compared to 240s needed for NST model. This scenario applies to the image resolutions mentioned in II-B. It is worth noting that in cases where higher resolution images are required, the difference in inference time would be even more significant and pronounced.

In general, if time is not the determining factor for user and if he or she wants to manipulate with output image while getting artistic results, NST approach would be the recommended choice. On the other hand, if time is of the essence and the user seeks to maintain realism in images or wishes to transfer style across video sequences, CycleGAN approach should be the preferred option. At the end, it is all about these balances and user's visual preferences.

References

[1] Efros, Alexei A., and Thomas K. Leung. "Texture synthesis by non-parametric sampling." In Proceedings of the seventh IEEE international conference on computer vision, vol. 2, pp. 1033-1038. IEEE, 1999.

[2] Efros, Alexei A., and William T. Freeman. ”Image quilting for texture synthesis and transfer.” In Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 571-576. 2023.
[3] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. ”Image style transfer using convolutional neural networks.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2414-2423. 2016.
[4] Huang, He, Philip S. Yu, and Changhu Wang. ”An introduction to image synthesis with generative adversarial nets.” arXiv preprint arXiv:1803.04469 (2018).
[5] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. ”Unpaired image-to-image translation using cycle-consistent adversarial networks.” In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.
[6] Harris, Frederick. Ukiyo-e: the art of the Japanese print. Tuttle Publishing, 2012.
[7] Kingma, Diederik P., and Jimmy Ba. ”Adam: A method for stochastic optimization.” arXiv preprint arXiv:1412.6980 (2014).
[8] Zhuang, Fuzhen, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. ”A comprehensive survey on transfer learning.” Proceedings of the IEEE 109, no. 1 (2020): 43-76.
[9] Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).