A Review on Neural Style Transfer

Jiayue Li¹, Qing Wang¹a*, Hong Chen¹, Jiahui An¹ and Shiji Li¹

¹Department of Computer Science and Technology, China Agricultural University, Beijing 100083, China
²wangqingait@cau.edu.cn, *1061214320@qq.com

Abstract. Image style transfer is a method that can output styled images, which can both retain the original image content and add new artistic style. When using neural network, this method is referred as Neural Style Transfer (NST), which is a hot topic in the field of image processing and video processing. This article will provide a comprehensive overview of the current NST methods. Firstly, we introduce the current progress of NST from two aspects: the image-optimisation-based method and model-optimisation-based method. Then we compare and summarize different types of the NST algorithms. The review concludes with a discussion of applications of NST and some proposals for future research.

1. Introduction

Image style transfer method, so called image stylization, is a kind of computer image processing strategy that retains common photo content and adds certain artistic style. The effect of style-transferred image caters to people's curiosity psychology and pursuit of beauty, and it has been widely praised and independently socialized. This image style transfer method also has many practical applications in different fields, such as short video creation, live video, and film special effects.

Before the application of neural networks, the traditional nonparametric image style transfer often chose to analyse a certain style of image, and establish a corresponding mathematical statistical model to achieve the specific effect. This type of method can only extract the low-level features of the image, instead of the high-level abstract features. When dealing with images with more complex color and texture, the final image synthesis effect is relatively rough. The traditional style transfer research methods such as Non-Photorealistic Rendering (NPR) [1] can only do a certain style or scene, and its application scope is limited, which does not meet the actual industrial needs.

Deep learning uses multi-layer neural network, which can automatically extract different features from the target object. With the help of neural network, the extracted feature information is richer and the feature level is higher. Therefore, the neural network has excellent feature extraction ability and generalization ability correspondingly, which can effectively solve the traditional defects of artificial feature extraction algorithm.

The idea of utilizing deep learning to accomplish image style transfer originally comes from the research of Gatys et al. [2-4]. Gatys and others put forward that texture can be represented by statistical model of image local features, and Convolution Neural Network (CNN) is used to automatically extract and synthesize texture from a specific image, which can represent the style features of this image. Then, through weighted matching of image content, better artistic image output can be obtained, as shown in Figure 1.
Neural style transfer methods often work in two ways. One is based on image iteration, which starts from white noise image and optimizes the loss function iteratively, as shown in Figure 2; the other is a fast style transfer method accomplished by deep neural network model iteration, as shown in Figure 3.

Figure 1. Neural Style Transfer Output Effects [3].

Figure 2. Texture Analysis and Synthesis Method Using a White Noise Image [2].
Based on that, this paper mainly introduces neural style transfer, presents and analyzes the different research methods and experimental results, including the image-optimisation-based and model-optimisation-based neural style transfer methods. At the end of the paper, the challenges and opportunities of neural style transfer methods are summarized.

2. Image-Optimisation-Based Neural Style Transfer Method

Convolutional neural network (CNN) [6] is one of the most widely used deep neural networks. It has been successfully applied in natural language processing, computer vision and other fields, especially in many cases of computer vision. Based on CNN, Gatys et al. extracted the style feature from an oil painting, combined with the content feature of another image and output a unique art painting with the original photo content. About image content extraction [3], Gatys et al. proposed that the image can be processed with different convolution layers to obtain more clear local feature information, then calculated and processed by Gram matrix. Finally, through the fusion and reconstruction of content features and style features, the final result come as unique works with both image content and classic art style included [4].

Novak et al. [7] obtained more complex information by modifying the style representation, and imposed more strict constraints on the style transfer results. Risser et al. [8] used histogram loss function to synthesize texture, and provided a multi-scale synthesis method of convolutional neural network. The instability source of many previous methods was explained mathematically. It can converge in a small number of iterations, and is more stable in the optimization process. Yin and Rujie [9] proposed a style conversion algorithm based on content perception. Through the numerical experiments, they proved that the style features and content features were not completely separated by the neural network.

Li Chuan [10] enhanced Gatys’ framework with a combination of general Markov random field (MRF) models and discriminatively trained deep convolutional neural networks (dCNNs) for synthesizing 2D images. Markov describes the set of similar feature information, so CNN image feature mapping is divided into many blocks and matched to improve the visual rationality of the composite image. Champandard [11] introduced a new concept, and augmented the general architectures with semantic annotations, either by manually authoring pixel labels or using existing solutions for semantic segmentation. The result is a content aware generation algorithm, which provides meaningful control of the results, improves the quality of generated images by avoiding common faults, makes the results look more credible, and expands the functional scope of these algorithms. Refer to Figure 4. [12] proposed an approach consisting of a dual-stream deep convolution network as the loss network and edge-preserving filters as the style fusion model, with an additional similarity loss function added.
Liao Jing [13] proposed the concept of Deep Image Analogy, a coat-to-fine strategy used to compute the nearest-neighbor field for generating features, which are extracted from a deep Volatile Neutral Network for matching. They also verified its effectiveness in various situations, including style or texture conversion, color or style exchange, sketch or painting into photo and time interval.

3. Model-Optimisation-Based Neural Style Transfer Method

The model-optimisation-based method for neural style transfer is still for the fixed style training, but its speed has been greatly improved compared with the method based on image-optimisation. Specifically, the per-style-per-model neural methods was brought out firstly, then followed by the multiple-style-per-model neural method and the arbitrary-style-per-model neural method with more choices provided and faster processing speed.

3.1. Non-Image-Reconstruction-Decoder Methods

Johnson Justin [5] propose the use of perceptual loss functions for training feed-forward networks for image transformation tasks. network gives similar qualitative results but is three orders of magnitude faster using the feed-forward network. [14] proposed a multimodal convolutional neural network that takes into consideration subtle and exquisite representations of both color and luminance channels, and performs stylization hierarchically when losses multiply and scales increase. Ulyanov et al. [15] showed that swapping batch normalization with instance normalization can results in a significant qualitative improvement in the generated style-transferred images.

Huang Haozhi et al. [16] used the feedforward network to force the output of continuous frames to train, so as to maintain consistency in style and time. Zhang hang et al. [17] introduced CoMatch Layer that learns to match the second order feature statistics with the target styles. They also build a Multi-style Generative Network (MSG net), acquiring real-time performance and superior image quality. Carlos Castillo [18] proposed a method of target style transformation, which can simultaneously segment and stylize a single object selected by user. In this method, the outliers near the target boundary are smoothed and anti-aliased by using Markov random field model, so that the stylized objects are naturally integrated into the surrounding environment.
3.2. Image-Reconstruction-Decoder Methods

Although the arbitrary style transfer can alleviate the problem of low efficiency, it can only train models for specific styles, and still can not avoid the problem of parameter adjustment. In order to overcome these problems, scholars propose an image style migration algorithm based on image reconstruction decoder. This algorithm does not need model training for specific styles, avoids the problem of parameter adjustment, and can realize fast image style transfer of arbitrary style.

Li Yijun et al. [19] use a pair of feature transforms, whitening and coloring, that are embedded to an image reconstruction network without training on any pre-defined styles, so that it overcome the limitation of inability of generalizing to unseen styles or compromised visual quality. [20] proposed a method using a stylization step and a smoothing step to ensure spatially consistent stylizations and remain the stylized photo’s photorealistic. [21] is more effective in preserving content integrity, but it is still limited in maintaining the consistency of depth information, as shown in Figure 5. [22] introduce an interface to manually and spatially control the stylization levels, and combine multiple styles in the generator. They also use a conditional discriminator based on the coarse category of styles to tackle the challenging adversarial training for arbitrary style transfer, as both the input and output of our generator are diverse multi-domain images.

![Figure 5. Model-Optimisation-Based Method Using VGG Encoder and Decoder [21].](image)

4. Conclusion

In this paper, the image style transfer method based on deep learning, neural style transfer for short, is introduced in detail, and its research ideas and research methods are discussed and analyzed in depth. In this field, the trained image style has been gradually diversified, the generated effect has gradually met people's aesthetic requirements, and the generation speed has been gradually increased. Generally, the advantage of the image-optimisation-based method is that the synthesized image has high quality, good controllability, easy parameter adjustment and no need of training data. However, the calculation time is relatively long and it largely depends on the pre-training model. The model-optimisation-based method has the advantage of fast computing speed, which can be used for real-time video style transfer, and is also the mainstream technology of commercial application software, despite of the relatively poor quality of image generation and enormous training data needed.

On the whole, with the popularity of deep learning, researchers have more and more works and ideas on image style transfer. New breakthroughs have been made, therefore this emerging field has attracted great attention of practitioners from academic and business fields. Prisma is the first mobile application to provide free image neural style transfer service. Similarly, Tiktok has recently launched a variety of short video experience services using neural style transfer. Meanwhile, there are still some unsolved problems and challenges in this field, such as the limitation of per-style-per-model, the choice of pre-processing and post-processing, the qualitative and quantitative evaluation criteria, the perfection of
style transfer principle and the 3D objects style transfer. These listed challenges are worthy of further exploration in the future research.

Acknowledgments
Thanks to my mentor Qing Wang and Hong Chen. Thanks for their valuable suggestions for my research and my article.

References
[1] Bruce, G., Amy, G. (2001) Non-Photorealistic rendering. AK Peters, Ltd., USA.
[2] Gatys, L. A., Ecker, A. S., Bethge, M. (2015) Texture synthesis using convolutional neural networks.
[3] Gatys, L. A., Ecker, A. S., Bethge, M. (2015) A neural algorithm of artistic style. Journal of Vision.
[4] Gatys, L. A., Ecker, A. S., Bethge, M. (2016) Image style transfer using convolutional neural networks. In: Computer Vision & Pattern Recognition.
[5] Johnson, J., Alahi, A., Li F. (2016) Perceptual losses for Real-Time style transfer and Super-Resolution.
[6] Lecun, Y., Bottou, and L., et al. (1998) Gradient-based learning applied to document recognition. Proceedings of the IEEE.
[7] Novak, R., Nikulin, Y. (2016) Improving the neural algorithm of artistic style.
[8] Risser, E., Wilmot, P., Barnes, C. (2017) Stable and controllable neural texture synthesis and style transfer using histogram losses.
[9] Yin, Rujie. (2016) Content aware neural style transfer.
[10] Li, C., Wand, M. (2016) Combining markov random fields and convolutional neural networks for image synthesis. In: 29th IEEE Conference on Computer Vision and Pattern Recognition.
[11] Champandard, A. J. (2016) Semantic style transfer and turning Two-Bit doodles into fine artworks.
[12] Li, W., Zhao, W., Yang, X., et al. (2018) Photographic style transfer. Visual Computer.
[13] Liao, J., Yao, Y., Yuan, L., et al. (2017) Visual attribute transfer through deep image analogy. Acm Transactions on Graphics, 36(4): p. 120.
[14] Wang, X., Oxholm, G., Zhang, D., Wang, Y. (2016) Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer.
[15] Ulyanov, D., Vedaldi, A., Lempitsky, V. (2016) Instance normalization: The missing ingredient for fast stylization.
[16] Huang, H., Wang H., Luo, W. et al. (2017) Real-Time neural style transfer for videos. In: 30th IEEE Conference on Computer Vision and Pattern Recognition.
[17] Zhang H., Dana K. (2018) Multi-style generative network for real-time transfer. In: 15th European Conference on Computer Vision.
[18] Astillo, C., De, S., Han, X. et al. (2017) Son of Zorn's lemma: Targeted style transfer using instance-aware semantic segmentation. In: 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing.
[19] Li, Y., Fang, C., Yang, J. et al. (2017) Universal Style Transfer via Feature Transforms.
[20] Li, Y., Liu, M., Li, X. et al. (2018) A closed-form solution to photorealistic image stylization.
[21] Li, X., Liu, S., Kautz, J. (2018) Learning linear transformations for fast arbitrary style transfer.
[22] Xu, Z., Wilber, M., Fang, C. et al. (2020) Adversarial training for fast arbitrary style transfer. Computers & Graphics, 87.