Application of in Information Visualization and Generative Adversarial Network in the Review of Style Transfer

Tongxin Shi*

University of Sydney, Australia

* Corresponding Author Email: stxsherry6@gmail.com

Abstract. Style transfer is the task that aims to adopt the appearance or visual style of an image to another image. For example, we can transfer the style of a normal landscape photo into the style of a Monet's painting, in which objects are generally drown in an impressionism way. Nowadays, style transfer has become one of the most important application in photo editing. It has attracted artists and designers from around the globe.

Keywords: Style Transfer; GAN; Neural Style Transfer.

1. Introduction

Style transfer is the task that aims to adopt the appearance or visual style of an image to another image. For example, we can transfer the style of a normal landscape photo into the style of a Monet's painting, in which objects are generally drown in an impressionism way. Nowadays, style transfer has become one of the most important application in photo editing. It has attracted artists and designers from around the globe.

Over recent years, researchers have achieved great progress in developing more effective style transfer models. In general, we found that current research can be categorized into two major groups. The first group is the feature-based methods which aim to achieve effective style transfer by investigating the style transfer in feature space. For feature-based methods, they improve the results based on changing or optimizing loss function. Depending on the aspects of improvement that they focus on, several distinguish loss functions have been applied to enhance the results.

In addition to feature-based methods, the second group is Generative Adversarial Network (GAN) based methods. A GAN is a type of deep neural network training strategy consists of a generative network and a discriminative network. In GAN, the generative network aims to generate fake data samples that are very similar to the true data, while the discriminative network aims to distinguish the generated fake data from true data. By applying GAN on style transfer, its trained generative network would be capable of generating images which look like the desired styles. Such GAN-based methods mainly study how to develop better technologies that can help generative network transfer styles more appropriately.

In the past year, there have been many new ways of playing in various photoshop applications, one of which is style transfer. Style transfer is currently a popular class of software algorithms that aims to transfer the style from one image onto another image. It can be divided into two groups, neural style transfer, and GAN-based style transfer. GAN-based methods [4, 8, 7, 13, 19, 20, 21, 22, 23, 24, 25, 29] and the key success of GAN-based style transfer is the adversarial loss force the generated images to be indistinguishable from the real photo. Neural style transfer methods [1, 2, 3, 5, 6, 9, 10, 11 12, 14, 15, 16, 17, 18, 26, 27, 28, 30] require a reference style image where the network transfers the content image similar to the reference image while containing the original context.

1.1. Feature-based Methods

Feature-based methods aim to achieve style transfer by translating the high-level feature of the input image according to the feature of the target image style. The most famous Neural style transfer proposed by Gatys [6] in 2016, proposed a seminal work that captures the style of autistics images and transfers it to other images using CNN. Gatys used a feature map that provide from the pre-trained VGG network to calculate the Gram matrix and then model and extract the style in the image.

Try to let the Gram matrix of the image close to the Gram matrix of the style image and the feature map from the image close to the feature map of the content image where Gram Matrix is the matrix of the inner product of each vector. Later on, plenty of Neural style transfer is presented. These methods can be further grouped into quality-optimized methods and efficiency-optimized methods. The quality-optimized methods are intended to improve the quality of resulting images by optimizing loss functions, such as using MMD, Laplacian loss, or improving the Gram matrix. The efficiency-optimized methods attend to accelerate the process on the basis of excellent style transfer research results. Through implementing multiple instance normalization, defining new optimization objectives, or applying AutoEncoder to accomplish acceleration.

1.1.1 Quality-optimized methods

Although Gatys has created an impressive result, the reason that the Gram matrixes could represent style remains unclear. Therefore [18] introduce Maximum Mean Discrepancy (MMD) which demonstrates the Gram matrices of feature maps are equivalent to minimizing the Maximum Mean Discrepancy (MMD) with the second-order polynomial kernel. Further, they match the Gram matrices of the neural activation which can be reformulated as minimizing the MMD with the second-order polynomial kernel to achieve domain adoption in style transfer. For the problem that the low-level features of the content image in traditional optimization objective are absent, [16] apply Laplacian loss which measures the difference of the Laplacians, and correspondingly the difference of the detail structures, between the content image and new image. It is flexible and compatible with the traditional style transfer constraints. Thus a new optimization objective has been obtained, when minimizing this objective will produce a stylized image that better preserves the detail structures of the content image and eliminates the artifacts. Although the pre-trained networks demonstrated amazing results, the high-level features often focus on the primary target and neglect other details. Therefore, [20] introduced a novel approach for neural style transfer that integrates depth preservation as additional loss which preserves overall image layout while performing style transfer. Follow the progress of style transfer. Since the difficulty of achieving comprehensive style modeling using 1-dimensional style embedding, thus [29] introduce CoMatch Layer that learns to match the second-order feature statistics with the target styles and build a Multi-style Generative Network (MSG- Net) that is a feed-forward network that runs in real-time after training. The remaining problem of feed-forward-based methods trade-off generality for efficiency, which suffers from many issues, such as shortage of generality, lack of diversity, and suboptimality. [17] first improve the Gram matrix loss by subtracting the feature mean, so that the newly designed loss is more stable in scale and adds stability to the learning. Also, they add texture loss that computes between the synthesized result and the given texture to ensure that these two images share similar statistics. By empowering the ability to generate diverse samples and prevent overfitting, they correlate the output samples with the input noise vector. [19] also, focus on the inability of generalizing to unseen styles or compromised visual quality. They propose to use feature transformation whitening and coloring transforms (WCTs) to reflect a direct matching of feature covariance of the content image to a given style image, and couple feature transformation with a pre-trained general encoder-decoder network. After Arbitrary-Style-Per- Model Fast Style Transfer Methods (ASPM) has been proposed, improving ASPM's quality becomes a challenge. One of the aspects of optimization is focusing on brush strokes of images. Since different brush strokes will render a diverse level of detail or coordinate spatial distribution of visual attention between the content image and stylized image. Thus [27] develop an attention-aware multi-stroke (AAMS) style model for arbitrary styles transfer to encourage attention consistency for corresponding regions between content image and stylized image. It achieves both scalable multi-stroke fusion control and automatic spatial stroke size control in one shot. Another aspect is related to the work of [20], it has demonstrated the issue of the traditional style transfer methods that have failed to reproduce the depth map of the content image. Therefore, [13] proposes an extension method of AdaIN [8] which allows depth map preservation by applying variable stylization strength. For works on Arbitrary Style Transfer, [4] explore that more than 50% of the time, the Arbitrary Style Transfer stylized images are not acceptable to human users due to the under-or over-stylization.

Hence, they propose a new loss that largely overcomes imbalanced style transferability which happened on Arbitrary Style Transfer.

1.1.2 Efficiency-optimized methods

To optimize the efficiency of feature-based methods, [10] proposes to use perceptual loss functions (works by summing all the squared errors between all the pixels and taking the mean) for training feed-forward networks and show that they achieve Gatys [6] in real-time. Concurrently, [26] find out Gatys et al. require a slow and memory-consuming optimization process. Thus they focus on the problem of achieving the synthesis and stylization capability of descriptive networks using feed-forward generation networks. They train the compact feed-forward convolutional networks to generate multiple samples of the same texture of arbitrary size. Then transfer artistic style from a given image to any other image where the result of their network is not only lightweight and can generate textures which the quality can compare to Gatys, but also hundreds of times faster. The multi-style transfer is a challenge of improving efficiency since if the style transfer network is tied to a single style, when needs multiple styles it will train for every style end-to-end. Thus, [5] introduces conditional instance normalization that allows it to learn multiple styles due to only acting on the scaling and shifting parameters. They train a style transfer network on multiple styles requires fewer parameters than train same amount of separate networks. They demonstrated that the role of normalization in style transfer networks is all convolutional weights of a style transfer network can be shared across many styles, and it is sufficient to tune parameters for an affine transformation after normalization for each style. The impressive result of this model is it reduces each style image into a point in an embedding space that permits one to arbitrarily combine artistic styles in novel ways not previously observed. Furthermore, for multi-style transfer, StyleBank has been proposed by [1] .

StyleBank is composed of multiple convolution filter banks and each filter bank represents one style. The intermediate feature embedding convolved with the corresponding filter bank of a specific style, and produced by a single suto-encoder which decomposes the original image into multiple feature response maps. [25] explored that to improve the efficiency, not only constrained with the limited number of trained styles, but also the expensive cost of an optimization procedure that works for any style image. Therefore, they proposed a patch-based operation for constructing the target activation in a single layer, given the style and content images as the content image is replaced patch-by-patch by the style image. Hereafter, some works propose the network is usually tied to a fixed set of styles and cannot adapt to arbitrary new styles. [8] propose a novel adaptive instance normalization (AdaIN) layer that aligns the mean and variance of the content features with those of the style features for the first time enables arbitrary style transfer in real-time. Also, [7] build upon recent work, leveraging conditional instance normalization for multi-style transfer networks by learning to predict the conditional instance normalization parameters directly from a style image. Further, [21] describes the Variational AutoEncoder (widely used likelihood-based generative models) for latent space-based Style Transfer (ST-VAE) that performs multiple style transfer by projecting nonlinear styles to linear latent space, and it fuses them by linear interpolation for improving the efficiency of multi-style transfer.

1.2. GAN

GAN is the most common learning model in both semi-supervised and unsupervised learning which learn by generating fake or synthetic looking data, while simultaneously training a generative model to minimize this loss. Practically, GAN has introduced many applications in style transfer. These methods can be grouped into Typical GAN-based methods and CycleGAN-based methods. The Typical GAN-based methods are upgrades based on GAN, such as adding on conditions, using Markovian patches to captures the feature statistics, or creating a new learning

framework based on GAN knowledge. CycleGAN mainly focuses on the problem that translating an image from one domain to another domain. The CycleGAN-based methods optimize CycleGAN in various aspects, such as several enhanced CycleGAN methods clarify the problem of bi-directional translations, the difficulty of using labeled data, limitations of adding generated images, and the instability of CycleGAN. They committed to optimize CycleGAN for style transfer by solving above issus.

1.2.1 Typical GAN-based methods

Since several papers have demonstrated that use GAN for style transfer but only applied the GAN unconditionally. Therefore they rely on other terms, such as L2 regression to force the output to be conditioned on the input. Therefore [9] introduce Conditional GAN aims to simple the setup than most others. Conditional GANs (cGANs) learn a conditional generative model which not only learns the mapping from the input image to the output image, but also learns a loss function to train this mapping. By solving the classic problem that synthesizing content by example, [14] proposed a combination of generative Markov random field (MRF) models and discriminatively trained deep convolutional neural networks (dCNNs) for synthesizing 2D images. At the same time, [15] figure out deep neural network approaches still come at considerable computational costs. Thus they focus on the efficiency issue by precomputing a feed-forward, they proposed stridden convolutional network that captures the feature statistics of Markovian patches and can directly generate outputs of arbitrary dimensions. Due to the style transfer is not able to conduct when a single target-style domain contains semantic objects. In the other words, when they have their own distinct and unique styles, such as anime-style domain, the current style transfer is not working. Therefore [11] propose pseudo-supervised learning framework, it provides the semantic connections between the objects of the two style domain in the image space. They introduce a generator network SMSTnet and a new training method for the discriminator. By cause of the domain adoption problem is still unclear, [12] introduce discover cross-domain relations with GANs (DiscoGAN) that aims to achieve cross-domain transformation which accomplish to constraint all images in one domain to be representable by images in the other domain. Later on, [3] has considered the case that transfer multiple artists styles within a single model, and is usually difficult to be preserved in a fine-grained locally consistent manner. Thus, they introduce a Conditional Generator, a novel module called Anisotropic Stroke Module, and a Multi-Scale Projection Discriminator, which constitute the Stroke Control Multi-Artist Style Transfer framework

1.2.2 Based on CycleGAN

Typical GAN-based methods has produce incredible results in style transfer. Nonetheless, CGANs and some other image-to-image transformation require the paired images to learn the mapping between the input images and the output images which is difficult to acquire in many tasks. Thus, [30] proposes to translate an image from a source domain X to a target domain Y in the absence of paired examples which known as CycleGAN. The invention of CycleGAN solve one of the challeging issue of style transfer that train with unpaired images, then came a lot of optimization methods based on CycleGAN. Concurrently, [28] realized human labeling is expensive and millions of labeled image pairs are needed to train a CGAN, develops DualGAN to train the image translators from two sets of unlabeled images from two domains. Based on CycleGAN and DualGAN, [22] discover that the explicit regularization term penalizes the perceptual quality between the images from the two domains is absent. Therefore, they introduce ARGAN which optimizes CycleGAN on an activation reconstruction loss. It can help enhance the sense of perception between the original image and the generated image as well as improve the stability of the generated image. Later, [24] figure out that CycleGAN lacks effective constraints since the presence of artifacts and degenerated transformations, thus they proposes HarmonicGAN to learn bi-directional translations between the source and the target domains, and enforce consistent mappings during the translation. LeafGAN [23] aim to improve the limitation of adding generated images to the training set of CycleGAN (since the generated images are close to the distribution of the original training data) and CycleGAN has no explicit attention mechanism (which cause it tends transform the entire image from the source to the target domain rather than transforming the specific objects). It determines the area of the image that is important and translates only that area from the source to the target domain which focuses to develop a segmentation module that segments the area of interest from the background and pays attention to the ROIs(Role of Interest). Also, [2] discover that CycleGAN will easily generate inappropriate image results when the number and shapes of objects in the style offering image and the source image are greatly different. They present an arCycleGAN which can transfer the freshness styles from the style offering images to the unpaired input source images and the result will have the freshness attributes of the style image and maintain the shapes and key features of the input source images.

References

[1] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua. Stylebank: An explicit representation for neural image style transfer. IEEE Computer Society, 2017.

[2] H. Chen, M. Guan, and H. Li. Arcyclegan: Improved cyclegan for style transferring of fruit images. IEEE Access, PP(99):1–1, 2021.

[3] Xuanhong Chen, Xirui Yan, Naiyuan Liu, Ting Qiu, and Bingbing Ni. Anisotropic stroke control for multiple artists style transfer. CoRR, abs/2010.08175, 2020.

[4] J Cheng, A. Jaiswal, Y. Wu, P. Natarajan, and P. Natarajan. Style-aware normalized loss for improving arbitrary style transfer. 2021.

[5] V. Dumoulin, J. Shlens, and M. Kudlur. A learned representation for artistic style. 2016.

[6] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[7] G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, and J. Shlens. Exploring the structure of a real-time, arbitrary neural artistic stylization network. 2017.

[8] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. IEEE, 2017.

[9] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. 2017.

[10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, ECCV (2), volume 9906 of Lecture Notes in Computer Science, pages 694–711. Springer, 2016.

[11] S. Kim, J. Do, and M. Kima. Pseudo-supervised learning for semantic multi-style transfer. IEEE Access, PP(99):1–1, 2021.

[12] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. CoRR, abs/1703.05192, 2017.

[13] V. Kitov, K. Kozlovtsev, and M. Mishustina. Depth-aware arbitrary style transfer using instance normalization. 2019.

[14] C. Li and M. Wand. Combining markov random fields and convolutional neural networks for image synthesis. IEEE, 2016.

[15] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In European Conference on Computer Vision, 2016.

[16] Shaohua Li, Xinxing Xu, Liqiang Nie, and Tat-Seng Chua. Laplacian-steered neural style transfer. In Qiong Liu, Rainer Lienhart, Haohong Wang, Sheng-Wei ”Kuan-Ta” Chen, Susanne Boll, Yi-Ping Phoebe Chen, Gerald Friedland, Jia Li, and Shuicheng Yan, editors, ACM Multimedia, pages 1716–1724. ACM, 2017.

[17] Y. Li, F. Chen, J. Yang, Z. Wang, and M. H. Yang. Diversified texture synthesis with feed-forward networks. IEEE, 2017.

[18] Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. Demystifying neural style transfer. CoRR, abs/1701.01036, 2017.

[19] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, NIPS, pages 386–396, 2017.

[20] Xiao-Chang Liu, Ming-Ming Cheng, Yu-Kun Lai, and Paul L. Rosin. Depth-aware neural style transfer. In Holger Winnemoeller, Lyn Bartram, Stephen N. Spencer, Bruce Gooch, and Yotam I. Gingold, editors, NPAR@Expressive, pages 4:1–4:10. ACM, 2017.

[21] Z. S. Liu, V. Kalogeiton, and M. P. Cani. Multiple style transfer via variational autoencoder. 2021.

[22] H. Nazki, S. Yoon, A. Fuentes, and S. P. Dong. Unsupervised image translation using adversarial networks for improved plant disease recognition. Computers and Electronics in Agriculture, 168, 2019.

[23] H. C. Quan, H. Uga, S. Kagiwada, and H. Iyatomi. Leafgan: An effective data augmentation method for practical plant disease diagnosis. 2020.

[24] Z. Rui, T. Pfister, and L. Jia. Harmonic unpaired image-to-image translation. 2019.

[25] Q. C. Tian and M. Schmidt. Fast patch-based style transfer of arbitrary style. 2016.

[26] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. 2016.

[27] Y. Yao, J. Ren, X. Xie, W. Liu, Y. J. Liu, and J. Wang. Attention-aware multi-stroke style transfer. 2019.

[28] Z. Yi, H. Zhang, P. Tan, and M. Gong. Dualgan: Unsupervised dual learning for image-to-image translation. IEEE Computer Society, 2017.

[29] H. Zhang and K. Dana. Multi-style generative network for real-time transfer. 2017.

[30] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. IEEE, 2017.