A Survey on Image Style Transfer Approaches Using Deep Learning

Changshen Zhao*  
1Department of Science, University of Technology, Qingdao, Shandong, 266033, China  
*Corresponding author’s e-mail: zhaneneversettled@gmail.com

Abstract. Image style transfer has been an important branch of computer vision and image processing. Inspired by the development of deep learning, applications of Convolutional Neural Networks (CNNs) in recomposing content and style of two separated images were proven to be effective by the recent works of image style transfer. This paper provides a survey of image style transfer, which focuses on methods which using deep learning. In addition, this paper also provides information of methods without deep learning, and seeks their points of innovation. In this paper, all references are published ranging from 2001 to 2017, which presents an overview of progress in this realm over the past two decades.

1. Introduction
Deep Learning today has been evolved with widespread applications in computer vision and image processing. For instance, artistic creation or computer-synthesized painting, as one of the significant applications, is a popular combination of traditional approach and artificial intelligence technology. In the past, how to transfer one painting style to another is arduous and time-consuming. Now, image style transfer algorithms were developed to solve this problem, and the study of this field has attracted the attention of researchers.

Before deep learning rises, the style transfer methods were usually called texture transfer. These algorithms provide methods and inspirations for later algorithms with deep learning. This paper mentions two types of texture transfer methods, non-photorealistic rendering and photorealistic rendering. Non-photorealistic rendering often works on paintings or synthesized artistic image. In the work of Ashikhmin [2], a fast texture transfer algorithm is first proposed to use the MRF method instead of the exhaustive search when searching for candidate pixels, so that the texture synthesis speed could be distinctly improved. Based on Ashikhmin’s approach, Lee et al. [3] proposed an improved algorithm. Their method works faster and more effective on style transfer. Photorealistic rendering algorithms often proceed with real photographs or synthesized realistic images. The image quilting algorithm which Efros and Freeman [4] presented for texture synthesis and texture transfer is one of the most important Photorealistic Rendering methods. The key idea of their approach is to quilt texture pieces by calculating the best path which minimized the loss.

Nowadays, as GPUs’ significant evolution, studies on deep learning and neural networks became prevalent again. Researchers combined texture transfer technique and deep learning models, and presented Neural Style Transfer (NST), which now is a greatly influential research area. The first effective and complete neural method was studied by Gatys [1] et al. They defined two specific elements style and content, and proposed to combine the two elements to redraw a new image with a pre-trained convolutional neural network (CNN). Represented by Gatys et al.’s method, Deep Learning methods based on online models are the algorithms that based on online pre-trained structure such as VGG
networks [5]. In view of Gatys et al.’s approach is rather incompatible with Photorealistic transfer; Luan et al. [6] proposed an algorithm improved on Photorealism. They modify the original algorithm by altering the formulas.

Online image reconstruction methods are relatively complex and ponderous, so some researchers presented Deep Learning methods based on offline models in order to overcome the lack of flexibility. These algorithms use offline and pre-trained models that are trained by data-driven methods. Actually, transfer speed of these algorithms is rather faster than ones based on offline models. According to these model’s capacity of style types, offline models could be classified into three genres: single-style models, multi-style models and arbitrary-style models.

Methods using single-style models are the algorithms which each trained model could transfer only one kind of style without retrain. Models they used are the earliest offline models which are relatively small in size and provide faster speed. Johnson et al. [7] used single-style models to reach fast speed and reduce the computational cost. They introduced a new approach to measure perceptual loss, which could be applied both on style transfer and image super-resolution. Zhu et al. [8] exploited generative adversarial networks (GANs) to present a new type of loss, and introduced image translate technique without paired data. Methods using multi-style models have the capacity to transfer image to multiple style without retrain the model. Chen et al. [9] presented StyleBank, which is the key structure of their method to support multi-style capacity. These models are the most flexible methods which cost distinctly little time to train models for new style. Huang et al. [10]’s approach was one of the first methods which satisfied the need of arbitrary-style and real-time. Based on conditional instance normalization (CIN), they proposed adaptive instance normalization (AdaIN), which is the core structure to achieve arbitrary-style.

The remaining of the paper is organized as follows. Section 2 introduces several datasets in experiments. Section 3 reviews the deep learning models in some image style transfer methods. Section 4 describes the applications. Section 5 presents the conclusions and some future works.

2. Datasets

2.1. ImageNet Dataset
The ImageNet [11] is the world largest dataset for visual recognition application such as image classification and object detection. The dataset consists of 22 thousand classes and over 15 million images, which has been manually classified and annotated.

2.2. Set5, Set14
Set5 and Set14 datasets are based on Nonnegative Neighbor Embedding, their main purpose is image super-resolution. The datasets were introduced by Bevilacqua et al. [11] in their paper.

2.3. BSD100 Dataset
BSD100 is a dataset for image segmentation and boundary detection. It consists of 200 images in the training set and 100 images in the test set. Images in the dataset were annotated manually by 30 volunteers. The dataset was introduced by Martin et al. [13] in their paper.

2.4. Microsoft COCO Dataset
COCO is a large-scale object detection, segmentation, and captioning dataset, which now becomes standard test platform for image captioning.

2.5. Cityscapes Dataset
The Cityscapes Dataset focuses on semantic understanding of urban street scenes. It consists of video sets from 50 different cities. The dataset was introduced by Cordts et al. [14] in their paper.
2.6. **CMP Facade Dataset**

The CMP Facade Dataset is a dataset of facade images, which includes 606 rectified images of facades from various sources.

2.7. **WikiArt Dataset**

WikiArt Dataset consists of over 150 thousand painting images by 2500 artists. The images in the dataset were collected and annotated by WikiArt. This dataset was introduced by Tan et al. [15] in their paper.

![Image](image.png)

**Fig. 1** The general examples from seven datasets.

3. **The Deep Learning Models**

3.1. **Style Transfer without Deep Learning**

3.1.1. *Non-Photorealistic Rendering*

Ashikhmin [2] presented a relatively faster method. He modified coherent synthesis algorithm he proposed earlier by two aspects. The original algorithm utilizes MRF method, and works through scan line order, one pixel at a time, search and pick the target pixels which are located in the same L-shaped neighborhood where already synthesized pixels in. His first modification is that perform a slight increase in search space, for example, to add a candidate pixel. Hence, the increase in space not only enhance the quality of synthesized image, but also solve the horizontal edges problem, which is raised by himself after checking his original coherent synthesis method. The second modification is to specify metrics according to different subjects. He proposed that to modify the neighborhood measure equation could effectively adjust the algorithm, so that it could solve specific problems more efficiently.

3.1.2. *Photorealistic Rendering*

Efros and Freeman [4] proposed image quilting algorithm for texture transfer. They introduced an efficient process to quilting texture: (1) Step1: randomly pick some texture blocks in specific size, and generate a sequence of blocks. (2) Step2: pick a random block $b_1$ and place it on the output image. (3) Step3: In each block adjoins $b_i$ which awaits synthesis, pick a random block $b_n$ and place it on the output image, make it have certain overlap area with $b_i$. (4) Step4: Calculate the error surface of the overlap area and find the best partition path, which minimized the loss. (5) Step5: Check the path Step4 offered, if the error does not exceed the upper limit, then keep $b_n$ on the output image, if the error exceeds the upper limit, reverse to Step3 and repeat. (6) Step6: Follow the raster scan order and repeat the process from Step3 to Step5, until the output image is completely synthesized.

3.2. **Style Transfer using Deep Learning methods**

3.2.1. *Deep Learning methods based on online models*
For the first one who introduced neural style transfer (NST), Gatys et al. exploited VGG-19 network which was pre-trained on ImageNet database. The key idea of the algorithm is to compute style loss with Gram matrix (i.e., use Gram matrix to represent style feature of an image). They first input a content image and a style image, and generate a random white noise image. Then, let $F^l$ be the style feature map at layer $l$ of the VGG net, afterwards compute the Gram matrix $G^l$, while $G^l$ is inner product of two sets of vectorized feature maps, and $l^\mathbb{C} \times C_t$ is the number of filter channels of layer $l$.

Luan et al. [6] studied follow-up research to improve the original algorithm by Gatys et al. Their approach focuses on photorealistic style transfer. They present two main modifications. First, they alter the optimization function by affix a regularization term for photorealistic transfer, so that the visual distortion of reconstructed image could be restrained. Second, they applied semantic segmentation method for guidance in order to prevent deviation during content transfer, therefore significantly improves the Photorealistic performances.

### 3.2.2. Deep Learning methods based on offline models

**a) Methods using single-style models**

Johnson et al.’s approach [7] is one of the representative algorithms which based on single-style models. They use stochastic gradient descent method to train the residual convolutional network for style transfer, and use the pre-trained network to compute the perceptual loss. Similarly, Zhu et al. modify the Johnson et al.’s network architecture (i.e., residual convolutional network) for the style transfer by using GANs and cycle consistency, in order to measure the adversarial loss and the cycle consistency loss and then sum them up obtaining a full loss function.

**b) Methods using multi-style models**

Chen et al. proposed StyleBank to make their method support multi-style transfer [9]. The StyleBank model contains banks of convolutional filters, these filter banks use many filter channels to store stylish information. So one bank could represent one kind of style, and the whole network composed by multiple banks could represent multiple styles. The style transfer process relies on a set of an encoder network and a decoder network, which both contain convolutional layers and activation functions.

**c) Methods using arbitrary-style models**

Huang et al.’s approach was one of the first methods which meet the request of arbitrary-style and real-time [10]. Based on conditional instance normalization (CIN), they proposed adaptive instance normalization (AdaIN), which is the core structure to achieve arbitrary-style.

### 4. Applications

#### 4.1. Style Transfer without Deep Learning

**4.1.1. Non-Photorealistic Rendering**

Ashikhmin’s texture transfer method has two main application. First, it is capable with simple artistic style transfer, with MRF search method, his method is rather faster than previous approach. Second, the model could change images’ visual performance by alter their basic property, such as brightness, contrast and color vibrancy.

**4.1.2. Photorealistic Rendering**

Efros and Freeman’s approach [4] is regarded as one of the most important achievements in the field of computer vision. The main application of their method is texture synthesis that uses one small-sized image to generate large-sized image with basic features from original images. In addition, due to its flexibility and compatibility, image quilting method has been cited and exploited by many studies, and also developed plenty of improved version, such as improved image quilting [16]. The following studies has continued along recent years until 2017 [17].
4.2. Style Transfer using Deep Learning methods

4.2.1. Deep Learning methods based on online models

Gatys et al.’s seminal work profoundly affected the field of image style transfer [1]. Their introduced Gram loss, which is adapted by many studies. Transfer result using their method is generally accepted to be the standard of image transfer quality. In view of Gatys et al.’s algorithm is not relatively effective on photorealistic transfer, Luan et al. proposed their approach which focused on the realm of photorealistic. Photorealistic transfer can be applied on various areas e.g. virtual reality.

4.2.2. Deep Learning methods based on offline models

a) Methods using single-style models

Johnson et al.’s approach [7] uses single-style model which possesses the fastest speed and occupies least space in the class of offline models, so that it could be applied on lightweight web applications and mobile applications. However, in view of lacking flexibility while retraining, this method can barely be used on applications which provide large amount of styles. The method is also competent to execute image super-resolution task, which is to generate a high-resolution image from a low-resolution input image. By means of perceptual loss, their super-resolution approach can reach different result from the traditional methods which use per-pixel loss.

Zhu et al.’s approach focused on transferring image without paired training data. According to the paper, their algorithm could be used on multiple applications, both photorealistic and non-photorealistic: in the realm of photorealistic, the model could transfer one object to another one that belongs to a category bears visual resemblance. It could also switch the season in a scenery photo, or improve the photo’s visual property. For non-photorealistic, the model could change one photo to a specific artistic style created by an artist (e.g. Van Gogh) instead of transferring the photo to a style of single painting (e.g. The Starry Night). Meanwhile, the algorithm could transfer a painting to a visually similar image which is more like a photograph.

b) Methods using multi-style models

Chen et al. proposed StyleBank model in order to reduce the cost while adding new styles [9]. Their method improved the model’s style capacity so that substantially overcame problem of flexibility which single-style models have. Thus, it can be well exploited on some industrial application.

c) Methods using arbitrary-style models

Huang et al.’s algorithm [10] reached arbitrary-style and real-time, so it can be applied on many areas that request flexibility, e.g. scientific research. In fact, the model’s performance on style transfer depends highly on amount of the training data because of its data-driven training. It is noted that larger amount of training input would improve the quality of the transfer result otherwise the result would be debased.

5. Conclusion

Image Style Transfer has considerable growth in the latest decades. Since some researchers introduced seminal algorithm of neural style transfer, this field has attracted more and more researchers. In this paper, we survey several image style transfer algorithms based on non-deep learning and deep learning techniques. Though the lack of consideration for high-level semantic information, non-deep learning algorithms provide the methods of image reconstruction and extracting the style feature. With the inspiration of these algorithms, deep learning methods have grown rapidly in recent years. More and more research concentrates on improving the transfer speed and quality. However, most studies at present evaluate the experiment result by subjective judgment, so a universal evaluation of image transfer quality is needed in the future.

References

[1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).
[2] Ashikhmin, N. (2003). Fast texture transfer. IEEE Computer Graphics and Applications, 23(4), 38-43.
[3] Lee, H., Seo, S., Ryoo, S., & Yoon, K. (2010, June). Directional texture transfer. In Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering (pp. 43-48). ACM.
[4] Efros, A. A., & Freeman, W. T. (2001, August). Image quilting for texture synthesis and transfer. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques (pp. 341-346). ACM.
[5] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[6] Luan, F., Paris, S., Shechtman, E., & Bala, K. (2017). Deep photo style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4990-4998).
[7] Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision (pp. 694-711). Springer, Cham.
[8] 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 (pp. 2223-2232).
[9] Chen, D., Yuan, L., Liao, J., Yu, N., & Hua, G. (2017). Stylebank: An explicit representation for neural image style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1897-1906).
[10] Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1501-1510).
[11] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.
[12] Bevilacqua, M., Roumy, A., Guillemet, C., & Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding.
[13] Martín, D., Fowlkes, C., Tal, D., & Malik, J. (2001, July). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Vancouver:: Iccv.
[14] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3213-3223).
[15] Tan, W. R., Chan, C. S., Aguirre, H. E., & Tanaka, K. (2016, September). Ceci n’est pas une pipe: A deep convolutional network for fine-art paintings classification. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 3703-3707). IEEE.
[16] Long, J., & Mould, D. (2007, May). Improved image quilting. In Proceedings of Graphics Interface 2007 (pp. 257-264). ACM.
[17] Raad, L., & Galerne, B. (2017). Efros and Freeman image quilting algorithm for texture synthesis. Image Processing On Line, 7, 1-22.