Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/35/20230403

Research on image style transfer based on deep learning

Junlin Li[1, 4], Chen Lin[2] and Yingbo Zhao[3]

[1]School of software, South China Normal University, Guangzhou City, China
[2]Sussex Institute of Artificial Intelligence, Zhejiang Gongshang University, Hangzhou, China
[3]Department Of Computer Science, University of Liverpool, Liverpool, United Kingdom

[4] 20202005377@m.scnu.edu.cn

Abstract. In many fields, image style transfer has been a popular topic. It also has a long history; the earliest examples date back to the previous century. A number of style transfer techniques are currently flourishing, from manual modeling to the use of neural networks. Image style transfer methods is improving and taking less time. This paper is separated into two main categories to analyze based on picture iteration and model iteration - and summarizes the various techniques of image style transfer based on deep learning in accordance with the timeline. This paper introduces maximum mean difference, Markov random field, and deep image analogy in picture iteration-based image style transfer. In terms of model iteration-based image style migration, this paper introduces generative models and image reconstruction decoders. Finally, the paper presents some recommendations and outlook on the future.

Keywords: deep learning, image style transfer, computer vision.

1. Introduction

In today's society, artificial intelligence has attracted widespread attention in various fields, colliding and resonating with various technological fields. Among them, image processing has begun to emerge, and the image style transfer has attracted widespread attention. Image style transfer, which provides an image A with a unique style, mostly in the case of artistic paintings, and then transforms any other image B into that style while retaining most of the content of B [1].

Image style transfer can be used in art and in everyday life, for example, the principle behind the various filters in mobile phone selfies is an image style transfer algorithm, which is certainly very interesting. Style transfer based on Adversarial Generative Networks can even achieve semantic transfer, allowing a horse in a picture to become a zebra, the moon to become the sun, etc. Image style migration has also been combined with other computer vision tasks to make it more practical.

The earliest Image transfer techniques usually used simple manual deduction algorithms. With the development of computer vision, people began to use neural network models to design fast image style transfer algorithms [2]. Compared to early style transfer technology, it not only greatly improves efficiency, but also achieves a perfect combination of original content images and style images. Today, more and more image style transfer methods are being proposed, some of them takes even less than 10 seconds to migrate any style in real time.

So, since the analysis of the latest image style transfer method has not been sufficiently explored, this paper will list and illustrate the various deep learning based (image iteration and model iteration) image style transfer method in detail through a timeline.

2. Image style transfer method based on deep learning

One of the hottest topics in the world of image research is the application of deep learning techniques for style transfer. Image iteration-based image style transfer methods and model iteration-based image style transfer methods are the two primary categories of the current mainstream deep learning-based image style transfer techniques. One approach relies on optimization iterations on white noise images to achieve image style transfer, with white noise images serving as the target for image optimization in this kind of approach. The second type of later involves quick style transfer via network feedforward while optimizing the neural network model with iteration. The optimization goal of this type of method is the neural network model.

2.1. Image style transfer method based on image iteration

It is a method of transforming any style of the network without style. First, input the content image, style image and initialize the white noise image, then use shallow layer of VGG19 pre trained by ImageNet to extract style features, use the deep layer of extracted content features, and then use the gradient descent approach to minimize loss and optimize the white noise image after computing the content and the style Loss function. The method based on image iteration can just stylize one image at a time, requiring a reasonable number of iterations and setting different content and style weight coefficients for different content style images to achieve reasonable results.

The main advantages of style transfer technique based on picture iteration are high-quality synthetic images, simple parameter and control adjustment, and control. However, this method has a longer calculation time and a greater dependence on the pre trained model. By matching the white noise picture to the style feature on the style map and the content feature of the content image, image iteration aims to produce the styled composite image.

2.1.1. Based on Maximum Mean Difference (MMD). The most used Loss function in transfer learning is the maximum mean difference (MMD), which is primarily used to compare the distributions of two dissimilar but related random variables [1-2]. Finding a mapping function that maps variables to a higher dimensions space is the fundamental tenet of MMD. After mapping, finding the predicted difference between two dispersed random variables follows. Find the supremum of this difference, which is known as the mean discrepancy.

2.1.2. Based on the Markov Random Field (MRF). Most of the early data-driven synthesis methods are based on MRF, which consider that the most relevant part of the image depends on the local feature distribution, and learn the possible features of the patch by the pixel points in the local k*k range, which is usually obtained by the nearest neighbour algorithm, but a big problem of Markov Random Fields - based texture synthesis is that it will lead to the deformation of the image within the patch, and even if the search space is in a higher dimensions still cannot suppress the deformation.

Considering the model generality and non-deformation, we consider deep convolutional neural networks with discriminative ability to identify complex features and nonlinear transformations of the image as an alternative to the nearest-neighbour search in Markov Random Fields. However, the continuous convolution-pooling operation causes the CNN to continuously lose detail information during the feature extraction process, which makes the final feature representation represent high-level semantic features but is very rough, such as the visualisation results in Figure 1.

DOI: 10.54254/2755-2721/35/20230403

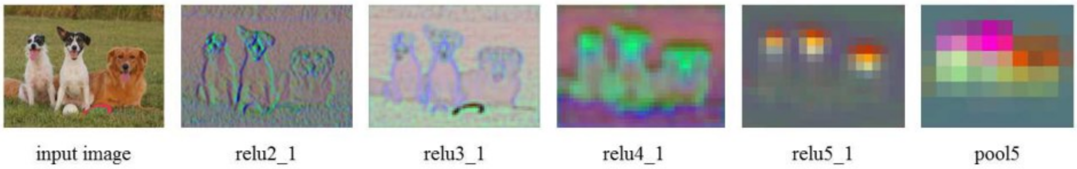

Figure 1. Visualisation Results [2].

2.1.3. Based on depth image analogy. Based on the depth image analogy, two images A1 and A2(unprocessed source image and processed image, respectively) and an unprocessed target image B1 are given to synthesize a new target filtered image B2, so that Including texture transfer, image blur, super-resolution, texture synthesis, artistic style transfer, it also effects of hand drawn images transforming into photos [3]. However, this method has a longer calculation time and a greater dependence on the pre trained model.

2.2. Image style transfer method based on model iteration

Low computational efficiency is caused by the picture iteration-based image style transfer's excessive iterations and lengthy iteration times. The model-based iteration approach, which can be divided into based on reconstruction decoder and based on image generative model, can significantly cut down on time requirements and increase computational efficiency.

The primary benefit of the picture style transfer approach based on model iteration, which can be utilized for quick video stylization and is currently the standard technology in industrial application software, is fast calculation speed. The primary drawbacks include the need for more training data and the necessity to increase the quality of image production.

2.2.1. Based on generative model. The image style transfer method based on iterative optimization generation model was firstly proposed by Johnson et al. [4], also known as fast style transfer. Different from the previous loss function that compares each pixel during training, the perceptual Loss function performs square difference on the high-level abstract feature retrieved from the pre-training VGG model. The algorithm used by Gatys et al. is consistent with the problem-solving in this section. The residual network was employed by Johnson et al. as the fundamental building block of the generative model, and the COCO data set as training data. The following function is used as perceptual loss function [4]:

$$y^* = \lambda_s L_c(f(x), y_s) + \lambda_{TV}\big(f(x)\big) + \text{argmin}\lambda_c L_c(f(x), y_c) \tag{1}$$

Among them, x indicates the input of network, and f(x) indicates the Generative model function. λs, λc, and λTV respectively indicate style loss function, the weight coefficients of the content Loss function and image smoothing function. Yc stands for the content feature representation that the pre-trained VGG model extracted from the training image, and ys stands for the style feature representation that the pre-trained VGG model extracted from the style image (Figure 2).

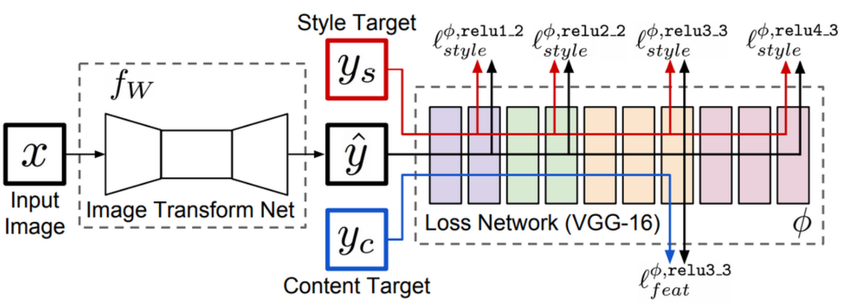

Figure 2. VGG16 convolutional neural network system [4].

DOI: 10.54254/2755-2721/35/20230403

The fast style transfer algorithm has opened up new ideas for style transfer. In addition, Ulyanov et al. also used a similar architecture of network, and experiments show that in the training process of the Generative model, using case normalization instead of batch normalization can significantly improve the quality of generated images [5]. For the purpose of fast style migration of multiple styles, Zhang et al. developed a Generative model that can train numerous styles [6].

In addition, unsupervised Generative adversarial network (such as CycleGAN, DiscoGAN, etc.) also have excellent effect in image style migration. However, the current Generative adversarial network is quite unstable in model training, and the setting of Discriminative model makes it difficult for the network to point to a clear image style. In addition, the Generative adversarial network is based on the iterative optimization of image divergence distribution to conduct confrontation training, rather than on the image's color, texture and content, so it is hard to get command of the process of image style transfer using the Generative adversarial network.

2.2.2. Based on image reconstruction decoder. Image based iteration has two drawbacks: adjustment of parameters and low efficiency. Fast style transfer can only train models for specific styles and still cannot solve the parameter adjustment problem, even though it solves the problem of poor efficiency. To overcome these problems, a style transfer algorithm based on image reconstruction decoder was proposed by Li et al., which allows the network to transfer any style without repeated training [7]. The system's whole pipeline is depicted in figure 3. It was trained end-to-end using stochastic gradient descent in one step. The encoder is based on the VGG-16 network's 13 convolutional layers, while the decoder arranges them in reverse. During the test phase, Monte Carlo samples of the model were used to generate probabilistic outputs, and the variance of these softmax samples was used to represent the model uncertainty for each class. This algorithm has two main innovation points. The first one is that the content feature data is directly matched to the deep feature space of the style image by feature transfer (such as whitening and coloring). Another one is that combining feature transformation with pre trained general codec networks enables the migration process to be achieved through simple feedforward operations.

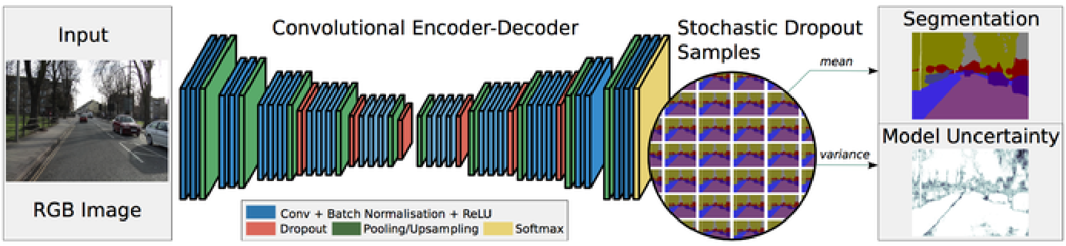

Figure 3. A schematic of the Bayesian SegNet architecture [7].

3. Existing problems

For now, has been able to produce some high-quality and high-definition style migration outcomes using the deep learning-based picture style migration technique, but there are still some shortcomings and some places that can be improved.

- Parameter Modification. Both iterative methods which based on different and divided images and iterative methods which based on different and divided models require manual parameter tuning in order to achieve adequate results, particularly model-based iterative methods where the models must be trained a lot after each parameter adjustment. Although the parameter tuning issue can be resolved by the image reconstruction-based encoder approach and it does not need to use different and divided model training for different and various styles, the image reconstruction decoder training procedure takes more time, and the image creation outcomes are not very satisfactory. Local smoothing enhances the decoder-based image reconstruction approach but removes the texture of the stylized image, producing a result that is almost identical to image color migration.

- Developing the theory of transfer learning. Transfer learning frequently uses the use case of image style transfer. Deep learning-based transfer learning techniques are now in their infancy and require the guidance of more sophisticated mathematical theories and procedures. For the further advancement of picture style transfer based on deep learning, the improvement of transfer learning theory is crucial. The related research on generic models suggests the creation of generic neural network models to enhance the models' ability to learn from migration, which is a crucial direction for the advancement of picture style migration.
- Some of the current results may not quite match expectations. To make the final findings more applicable to practical applications, both pre- and post-processing techniques, such as image fusion, image color migration, image smoothing, and there are many, many more can be used. These pre- and post-processing methods sometimes can improve the influence on picture style transfer. For instance, Castillo et al. combined the technique of image semantic segmentation to migrate the style of specific objects in the image [8]. Li et al. combined the technique of image fusion technology to provide a friendly interaction for the user [7]. And Gatys et al. used the technique of image colour migration to achieve the stylized image color control [9]. Li et al. on stylization the image is then partially smoothed to obtain a photo-like effect [10].

4. Outlook

At present, by analysing the results of the deep learning-based image style migration and calculating the data, it is possible to make some envisaged treatment of some problems and for the future outlook.

- The next crucial area for research is to discover a straightforward controlled procedure that can also ensure image quality. Because the approach can successfully avoid the problem of parameter adjustment, further improving the quality of picture production depend on the image reconstruction encoder method is an awfully worthwhile research and study direction. This is especially true if image reconstruction encoder-based image generation is improved without taking the size of the model's storage capacity into account.
- Connecting various stylized images in order to acquire the required look. For instance, limiting the consistency of stylized images produced from the same source domain image in various iteration cycles during the middle and late stages of model training reduces variability so that the model essentially learns the same style distribution or learns the desired style and does not learn different style distributions based on the iteration cycles.

5. Conclusion

The deep learning-based image style transfer algorithms are described in detail in this paper. At the end of the article, the author discusses the various benefits and drawbacks of these style transfer algorithms, as well as changes that could be made in the future and potential solutions. Through the paper, it is found that based on the existing algorithms, it is true that they can go to success in some practical use cases, and high-quality style migration images can be obtained, and the traditional methods of processing images can also play a good role, but there is still a certain distance from mass commercialisation, and the existing algorithms still need to be improved, and there is still a lot of room for progress. All in all, it is a creative and challenging topic for the field of image style migration, and there are many opportunities along the way, which have also attracted extensive attention from the academic community. In conclusion, image style migration also has a lot of research value and commercial value to explore in business, and there are many application prospects waiting for researchers to explore.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style. *arXiv preprint arXiv: 1508. 06576*, 2015

[2] Li Chuan, Wand M. Combining Markov random fields and convolutional neural networks for image synthesis *Proc of IEEE Conference on Computer Vision and Pattern Recognition,*2016: 2479-2486
[3] Liao Jing, Yao Yuan, Yuan Lu, et al. Visual attribute transfer through deep image analogy. *arXiv preprint arXiv: 1705. 01088*, 2017
[4] Johnson J, Alahi A, Li Feifei. Perceptual losses for real-time style transfer and super-resolution *Proc of European Conference on Computer Vision,* 2016: 694-711
[5] Ulyanov D, Lebedev V, Vedaldi A, et al. Texture networks: Feed-forward synthesis of textures and stylized images. *arXiv preprint arXiv:1603.03417*, 2016.
[6] Zhang H, Dana K. Multi-style generative network for real-time transfer. *Proceedings of the European Conference on Computer Vision Workshops*. 2018: 10-21.
[7] Li Y, Fang C, Yang J, et al. Universal style transfer via feature transforms. *Advances in neural information processing systems*, 2017, 30.
[8] Castillo C, De S, Han X, et al. Son of zorn's lemma: Targeted style transfer using instance-aware semantic segmentation, 2017 *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017: 1348-1352.
[9] Gatys L A, Bethge M, Hertzmann A, et al. Preserving color in neural artistic style transfer. *arXiv preprint arXiv:1606.05897*, 2016.
[10] Li Y, Liu M Y, Li X, et al. A closed-form solution to photorealistic image stylization, *Proceedings of the European conference on computer vision.* 2018: 453-468.