Studies Advanced in Image Style Transfer based on Deep Learning

Haoyin Deng [1, †, *], Haoyu Lin [2, †] and Yibing Guo [3, †]

[1] Xinjiang University, Urumqi, China

[2] Lanzhou University, Lanzhou, China

[3] Agricultural University of Hebei, Baoding, China

* Corresponding author email: harvey_deng@stu.xju.edu.cn

†These authors contributed equally

Abstract. A popular project in computer vision in recent years is called "image transfer", which attempts to create new images containing features from abstract properties of the content and style of a given image. Convolutional neural networks have generated the basic components of image style transfer, significantly improving transfer accuracy and visualization effect, thanks to deep learning's quick progress. The purpose of this study is to highlight the most recent developments in the area of deep learning-based picture style transfer. We first provide a detailed introduction to the traditional style transfer algorithms from the frameworks of generative adversarial networks and convolutional neural networks., including the design ideas research issues, key steps, advantages and disadvantages. On a few application tasks, we qualitatively compare the visualization outcomes of various approaches, and we talk about the subject's current issues and potential future developments.

Keywords: Image Style Transfer; Loop Always Adversarial Network; Generative Adversarial Network.

1. Introduction

Image style migration has long been a popular work in the computer vision community. For a given style feature image dataset and content feature image dataset, the image style transfer task aims to generate new images, which can retain the image content information while having style features. Image style transfer has gained significant interest from academia and industry as a result of the rapid growth of deep learning. It is now frequently utilized in the creation of animations, the restoration of historical photographs, the creation of game scenes, among many other things.

Since each image can be a different artistic style, everyone has a unique opinion to the style of randomly given images, especially some images are not clearly defined even in the art world. Vague definitions are a fatal problem for image researchers, so exactly how to turn something that cannot even be described into an executable program is an issue that has plagued many works in image style transfer. In the early days, the main way to evaluate a particular image-style was to build a mathematical or statistical model for that approach, and then to change the image which to be migrated so that it could better fit the model built. According to Portilla et al., a statistical model of the local properties of the image can adequately characterize texture. A straightforward texture algorithm was proposed by Efros et al. [1] that creates new textures by stitching and recombining sample textures. Semmo et al.[2] proposed that the qualities of oil paintings are described and transferred using seven separate phases. However, the drawback of this way is that the images need to be modeled manually before processing and a program can basically only do a specific style or scene. Traditional style transfer-based research has relatively few practical applications since it produces final extracted characteristics that are coarse and difficult to match the actual requirements when working with image that have more intricate color and texture schemes.

Changing this status is deep learning's rise. Utilizing convolutional neural networks for texture modeling, invented by Gatys et al. and they employed convolutional neural networks to partition

content abstract characteristics from style abstract characteristics of images, thus effectively achieving image style transfer and obtaining considerable artistic results. According to different network structures and design ideas, generative adversarial networks and convolutional neural networks are frequently employed in deep learning-based visual style transfer techniques.

The primary deep learning-based image style transfer techniques are systematically reviewed in this work, along with their most recent research findings and potential applications. Additionally, the paper highlights the present issues that need to be resolved and provides some helpful examples. In order to build the groundwork for more in-depth research, the paper analyzes existing issues, evaluates recent research findings and application potential, and provides some useful recommendations with citations. Finally, the problems and prospective development tendencies are then summed up.

2. Methods

Because it has become difficult to complete complex industrial requirements using only traditional non-parametric image style transfer since deep learning's advent, this paper focuses on deep learning-based image style transfer.Starting with deep learning, this paper summarizes existing image style transfer methods based on convolutional neural networks and generative adversarial networks.

2.1 Convolutional Neural Network-based Method for Transferring Image Style

Convolutional neural network is the cornerstone of deep learning. It is a feedforward neural network with convolutional computation and depth structure .Convolutional neural networks are generally composed of the following six parts: input layer, convolutional layer, activation function, pooling layer, full connection layer and loss function. Among convolutional neural networks, Alexnet, Vgg and Resnet are the three most widely used networks in academia and industry.

In 2015, Gates et al.[3] proposed a graphic style transmission scheme based on Vgg 19 network. Convolutional neural networks have since been used to transmit visual styles. As can be seen from Figure 1, Vgg19 network can be divided into five blocks, each of which is made up of a pooling layer and a convolution layer.There is the same pooling layer in each block, which is the Max pooling layer, but there are various convolutional layers in each block: the first and second blocks have 2 convolutional layers, and the remaining blocks have 4 convolutional layers. Finally, there are three fully connected layers, namely FC1, FC2, and Softmax, but they are not used for style conversion work.

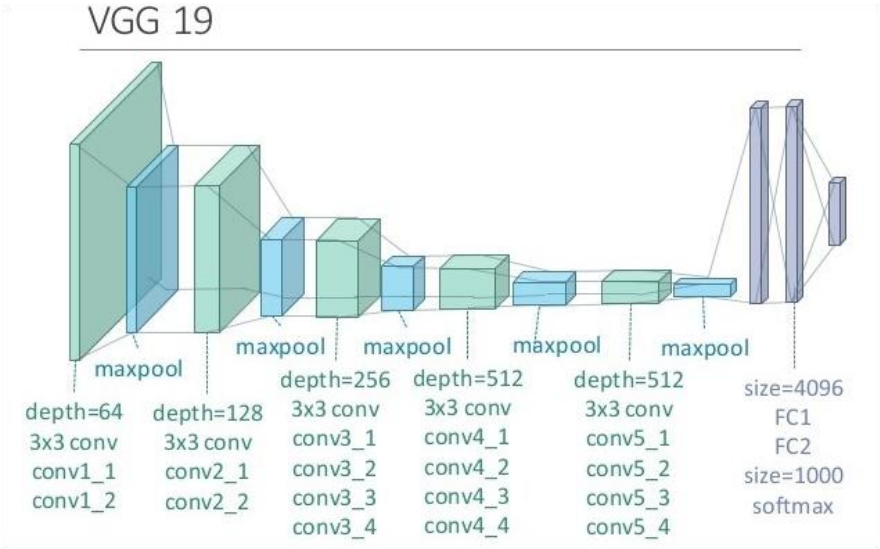


Fig 1. The network structure of VGG19

In Figure 2, the left picture represents the style input, while the right image represents the content input. The original picture information of the style image was mostly lost after being processed in each of the five blocks of Vgg 19, but much of it was preserved in the content data information. The style data is marked as α, the content image as p, and the randomly generated noise image as x. The content picture is marked as pl and the noise image is flagged as Fl after the lth block has been processed. Gates used block 4 to calculate content loss, and the following equation was used to determine the Euclidean distance between image features and noisy image features:

$$\mathcal{L}_{content}(\vec{p},\vec{x},l) = \frac{1}{2}\sum_{i,j}(F_{ij}^{l} - P_{ij}^{l})^2 \tag{1}$$

When calculating the style loss, the features obtained from the style image after five blocks need to be calculated by the Gram matrix first, and the calculated value is noted as Al, The feature Fl obtained after the lth block of the noisy image is calculated by the Gram matrix and marked as Gl. The Gram matrix formula is.

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l} \tag{2}$$

To calculate the Euclidean distance between Al and Gl, the formula is:

$$E_l = \frac{1}{4N_l^2 M_l^2}\sum_{i,j}(G_{ij}^{l} - A_{ij}^{l})^2 \tag{3}$$

The total loss can be obtained by linear summation of content feature loss and style feature loss, and the percentage of content loss and style loss can be changed by changing the values of α and β. The total loss formula is:

$$\mathcal{L}_{total}(\vec{p},\vec{a},\vec{x}) = \alpha\mathcal{L}_{content}(\vec{p},\vec{x}) + \beta\mathcal{L}_{style}(\vec{a},\vec{x}) \tag{4}$$

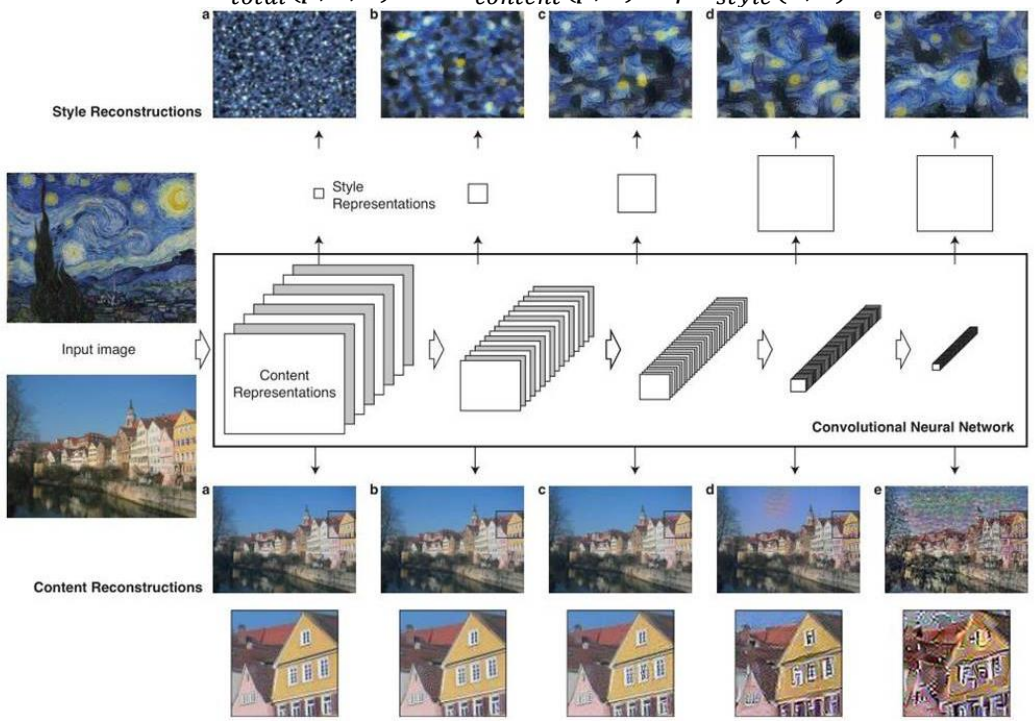


Fig 2. The framework of image style transfer based on VGG19

Johnson et al. [4] train a forward propagation network for the graphics transformation task using the perceptual loss function.Image Transform Net is the network with untrained parameters and Vgg 16 is the network with pre-trained ones,both of which constitute the network together.

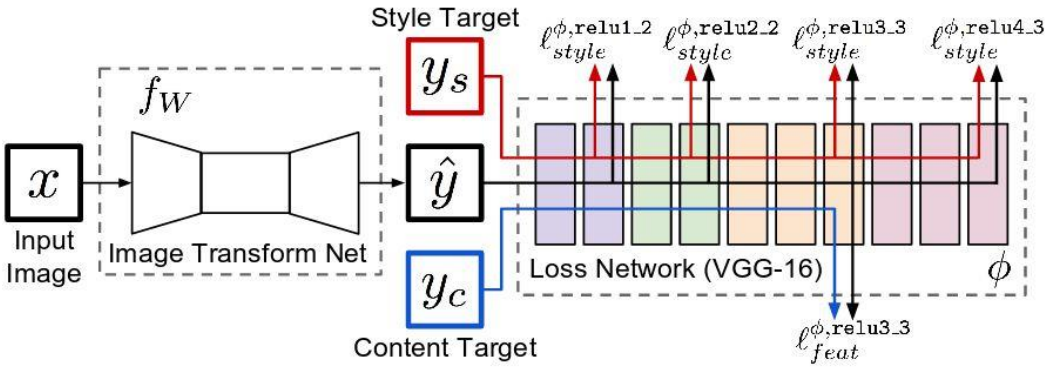


Fig 3. The framework of image style transfer based on VGG16

Image Transform Net is a residual network. As can be seen from Figure 4, the input image x passes through Image Transform Net to obtain a $\hat{y}$, and the mapping relationship between the two is $\hat{y} = fw(x)$. Input y into the Vgg16 network, and calculate the content loss by using the Euclidean distance between y and the target graph yc generated after the input content passes through Vgg16. The smaller the Euclidean distance is, the closer the two input images are. The formula is:

$$\ell^{\emptyset,j}_{feat}(\hat{y},y) = \frac{1}{C_j H_j W_j}||\emptyset_j(\hat{y}) - \emptyset_j(y)||_2 \quad (5)$$

Where j represents the middle layer of VGG-16; y represents the feature target image; $\emptyset j(y)$ represents the output of y on layer Vgg16; C_j, H_j and W_j respectively represent the number, height and width of channels in layer VGG-16. The Euclidean distance weighted sum of the Gram matrix generated by $\hat{y}$ and the input content target map yc in Vgg multiple intermediate layers obtained from the feature map is used as the style abstract feature loss, with the Gram matrix equation :

$$G^{\emptyset}_j(x)_{c,c'} = \frac{1}{C_j H_j W_j}\sum_{h=1}^{H_j}\sum_{w=1}^{H_j}\emptyset_j(x)_{h,w,c}\,\emptyset_j(x)_{h,w,c'} \quad (6)$$

The parameters C_j, H_j, and W_j denote the total number of channels in layer VGG-16j, their height, and their width, respectively. The Style loss formula is as:

$$\ell_{pixel}(\hat{y},y) = \frac{1}{CHW}||\hat{y} - y||_2^2 \quad (7)$$

After determining the overall loss, the network's parameters are continually updated using the gradient descent optimization approach, with a higher weight of content loss indicating more prominent content.

2.2 Generative Adversarial Networks-based Image Style Transfer

In 2014, Goodfellow et al.[5] invented the Generative Adversarial Network GAN, and It is made up of a discriminator and a generator. The generator generates certain noise vectors at random, and the discriminator decides whether the image is real. Equation[6] illustrates the adversarial interaction in the generative adversarial network Gan developed by Goodfellow et al.

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log(1 - D(G(\cdot)))] \quad (8)$$

Where x represents some numbers in the image, z is the noise sent to G network, $G(z)$ the final graph, $P_{data}(x)$ is some data output of the actual image, $P_z(z)$ represents the distribution of some noise sent to network G.

Without supervision, CycleGAN can do out image and image style change, so that images do not need to be paired, just two images can be passed in, and the structure itself does not make major changes.When doing image-to-image conversion,CycleGAN primarily learns the mapping relationship between the input image and the output image[7]. In the improved model of Du Zhenlong et al. [8], in the process of image transfer, most of the samples that need to be trained appear as a pair of original image and style image, and also need some labels, but it is difficult to realize in most cases. For this, Zhu et al.[9] proposed the CycleGAN structure, which can realize style transfer between any

two images without some labels. This structure has two G and F and Dx and Dy structures, namely production and discriminant structures, as shown in Figure 4.

In order to more effectively blend the style image with the original image without labels, CycleGAN made some changes from the feature combination generator model of the style image according to the transfer mechanism. The structure of the CycleGAN network is shown. It is made up of a discriminator, a converter, a decoder, and an encoder.

(1) Encoder: input the images into the neural network successively and extract the different styles. The first convolution layer has a number of 64 filters, a size of 256x256 when it is supplied to the encoder, and after it is finished, it has a 25664x64 feature vector.

(2) Converters: transfer a picture between domains.

(3) Decoder: The opposite of an encoder is a decoder.The work of restoring low-level features from feature vectors can wait for image generation.

(4) Discriminator: The discriminator will determine whether each image is a created one or the original.

The generator structure of CycleGAN follows the residual network used in fast neural style transfer[6], with some residual modules outside the lower sampling layer and upper sampling layer. In the builder structure, two downsampled convolution layers with a step size of 2, nine residual modules, and the BN (BatchNormalization) layer, the structure of the convolution layer and ReLU layer[11] are used. In order to integrate the depth feature transfer method into CycleGAN, the first three of the downsampled convolution layer and the residuals modules of the generator were used as encoders, the last three residuals modules and the upsampled convolution layer were used as decoders, and the middle three residuals modules were used as feature transfer modules.

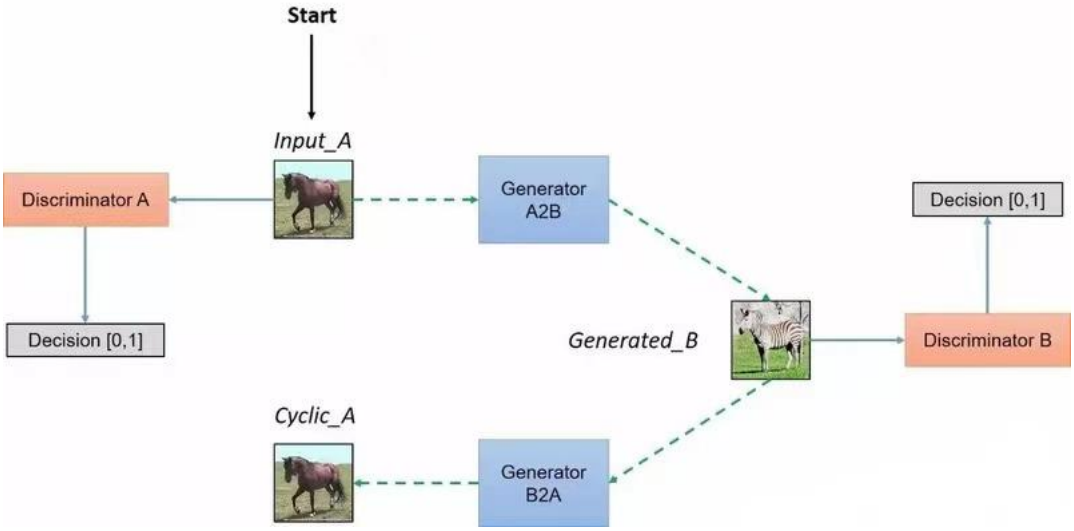


Fig 4. The framework of CycleGAN

CycleGAN's Loss function is divided into two sections: adversarial loss L_{GAN} and cycle consistency loss L_{cycle}. Among them, L_{GAN} is ensured that the generator and the discriminator evolve each other to generate better quality pictures, while making sure that the output image' content resembles that of the original image as much as possible. The totle loss of CycleGAN is as:

$$\begin{aligned} \mathrm{Loss}_{\mathrm{GAN}} &= \mathcal{L}_{\mathrm{GAN}}(\mathrm{G}, \mathrm{D}_{\mathrm{Y}}, \mathrm{X}, \mathrm{Y}) + \mathcal{L}_{\mathrm{GAN}}(\mathrm{F}, \mathrm{D}_{\mathrm{X}}, \mathrm{X}, \mathrm{Y}) \\ &= \mathbb{E}_{y\sim data(Y)} \log\big(D_Y(y)\big) + \mathbb{E}_{x\sim data(X)} \Big(\log\Big(1 - D_Y\big(G(x)\big)\Big)\Big) + \mathbb{E}_{x\sim data(x)} \log\big(D_X(x)\big) + \mathbb{E}_{y\sim data(Y)} (\log\Big(1 - D_x\big(F(y)\big)\Big)) \end{aligned} \tag{9}$$

3. Experiment

3.1 Evaluation Metrics

The essence of generating adversarial network-generated images is still an image processing problem. The picture quality evaluation standards are used to evaluate the quality of images generation in this experiment. The evaluation method we used in this experiment is structural similarity (SSIM), included in objective image quality evaluation indicators [12]. In this experiment, the closer the SSIM value of an image is to 1.0, the more accurate the target image changed from the original image.

3.2 Experiment Analysis

To evaluate the relative transfer effectiveness of various algorithms, we undertake a series of tests. During network training, the improved CycleGAN-based image style transfer algorithm picks 1500 sketch images as image style data, 1000 van Gogh paintings to transfer facial hair features and van Gogh oil painting features respectively. The training set contains 1,500 images of architectural styles, as well as drawings. To find the ideal parameter values, the model use stochastic gradient optimization with a batch size of 1 and an initial learning rate of 1e-5.

Figure 5 illustrates the effect comparison. Image style transfer simply modifies certain aspects of the images appearance, such as color and contour, as well as the underlying color from an overall perspective.The Vangogh2photo dataset has two kinds of graphics, one of Van Gogh's works and one of nature's landscapes, but there is no one-to-one correspondence between the two. Therefore, this CycleGAN-based technique not only modifies the image's underlying hue, but also imparts certain Van Gogh-like traits to its outline. Currently, this strategy can be used to different contexts.



Fig 5. Comparison of transfer effect between different methods

4. Discussion

Deep learning-based visual style transfer algorithms have made tremendous progress, but the issues in this field are still disturbing scholars. In this paper, we discuss the major problems of the present and offer some potential answers.

(1) Parameter alterations. The image-based iterative methods and the model-based iterative methods require human parameter tuning to achieve satisfactory results, especially the iterative model-based approach. The model must be retrained for the model-based iterative technique after every parameter change. The parameter tuning issue can be solved using the image reconstruction encoder approach, but the image reconstruction decoder training procedure is more time-consuming and the image production outcomes are less than ideal. Finding a simple, controllable procedure that can ensure the quality of the image is thus the next crucial study area. Due to the method's ability to

successfully overcome the parameter tuning problem, further enhancing the quality of image generation based on the image reconstruction encoder method without taking the size of the model storage capacity into consideration is a very interesting research direction to take into consideration.

(2) Limitations of pre-training models. Although pre-trained VGG is an excellent convolutional neural network model with excellent performance in feature extraction, VGG model is extremely bulky and with a large size alone with computational effort. Moreover, its design was not particularly for image style transfer. In order to further deep learning-based image style transfer and its continued development, it is crucial to move away from dependency on compact and more effective feature extractors. The limitation problem of training models could be resolved by generating adversarial networks. Moreover, in the context of gaining new features, the improvement strategy based on scatter distribution have the same result of the image iteration based correlation strategy, and the adversarial training strategy produces excellent results.

5. Conclusion

The area of this study on which we are concentrating is visual style transfer using deep learning. This paper lacks the most recent research findings, particularly the current representative image style transfer algorithms from the perspectives of convolutional neural networks and generative adversarial networks, including their design ideas, research problems, key steps, loss functions, etc. Second, we visualize the transfer effects of different algorithms and qualitatively analyze the potential advantages and disadvantages of different methods, hoping to provide some insights for choosing the optimal algorithm in different application scenarios. Finally, we discuss existing research issues and future directions for style transfer.

References

[1] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]//Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 2001: 341-346.

[2] Proceedings of the workshop on Computational AestheticsJune 2015 Pages 149–158.

[3] Gatys L A, Ecker A S, Bethge M. Image Style Transfer Using Convolutional Neural Networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[4] Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[C]// European Conference on Computer Vision. Springer, Cham, 2016.

[5] Goodfellow J, Alahi A, Feifei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. [J]. 2016: 38-42.

[6] Isola P, Zhu J Y, Park T, et al. Image-to-Image Translation with Conditional Adversarial Networks [J]. 2016.

[7] Mo Ran, GAO Shuzhi. Multi-target tracking based on wavelet filter based JPDA algorithm [J]. Journal of Shenyang University of Chemical Technology,2018,32(01):91-96.

[8] Du Zhenlong, SHEN Haiyang, SONG Gome, LI Xiaoli. Image Style Transfer Based on Improved CycleGAN [J]. Optics and Precision Engineering,2019,27(08):1836-1844.

[9] Weng Zixin, ZHANG Kai, LUO Peicheng, WU Minghui. Image Style Conversion Method and Its Application Based on CycleGAN [J]. Computer Age,2019(08):24-28.

[10] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [J]. JMLR.org, 2015.

[11] Kong L R. Research on Image Style Transfer Based on Generative Adversarial Network [D]. Guangdong University of Technology, 2019:36-39.

[12] Weng Zixin, ZHANG Kai, LUO Peicheng, WU Minghui. Image Style Conversion Method and Its Application Based on CycleGAN [J]. Computer Age,2019(08):24-28.