ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0023

Art Style Migration Based on Generative Adversarial Network

Song You[1, a], Guojun Lin[1, b, *]

[1]School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong 643000, China

[a]398215487@qq.com, [b]386988463@qq.com

Abstract

Using generative adversarial network to realize image style conversion for data enhancement has been favored by more and more researchers. At present, the research of style transfer based on generative adversarial network is a hot research direction, and the related algorithms emerge in an endless stream. This paper summarizes and discusses the mainstream methods and representative work of style transfer based on generative adversarial network, analyzes the main neural network models used in this field, and discusses their advantages and disadvantages and their improved models. Finally, the existing problems and future research directions of style transfer based on generative adversarial network are summarized.

Keywords

Data enhancement; Generative adversarial network; Style transfer.

1. INTRODUCTION

Image style migration is the process of synthesizing a given image into an image with a specified style. It means the conversion of two or even more images in different domains. Specifically, it is to provide a style image, convert any image into this style, and retain the contents of the original image as far as possible. Image style transfer is only the goal, there are many ways to achieve it. According to the time line, it can be divided into three categories: traditional image style transfer; Image style migration based on neural network and image migration based on generative adversarial network. The traditional style transfer method is to analyze the image of a certain style, then build a mathematical statistical model for the style, and then change the image to be transferred, so that it can better conform to the established model. Before the introduction of deep neural network to achieve style transfer, there are many non-parametric methods. By "transferring" the style of an image to another image, deep learning was initially realized by using CNN framework, but such a model has the problems of slow training speed and high requirements on training samples. However, the advantages of independent learning and random sample generation of GAN, as well as the reduction of the requirements for training samples, make GAN achieve fruitful research results in the field of image style transfer. The image migration of generating adhoc network has been continuously improved in recent years. At present, the technology has been quite stable and the generated effect is excellent. Style transfer based on generative adversarial network is an important research direction in the field of image style transfer and has a broad research prospect [1].

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0023

2. PROPERTIES

2.1. Generate an adversarial network

First generation against network 2014 Goodfellow (Generative Adversarial Nets, GAN), It has become a successful example of deep learning combining generative model and discriminant model. The generator generates data and the discriminator judges the authenticity of the data. The two compete with each other and promote each other. The ability of discriminator is also getting stronger, and the discrimination of true and false data is also getting stronger [2].The optimization process of GAN can be described as a "binary minimax" problem. The discriminator and generator constitute two players of the game. In order to achieve the final victory, the discriminator and generator will train and improve their discriminant and generative abilities, and finally maintain a relatively stable state -- Nash equilibrium state. At this time, the generator learns the probability distribution similar to the real sample, and the discriminator cannot correctly judge whether the input data is from the real sample or the generated fake sample, that is, the output probability value given by the discriminator is 1/2.The objective function of the generated adversarial network is:

$$\min_G \max_D L(D,G) = E_{x\sim p_r(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z))] \tag{1}$$

Where: $p(z)$ is the distribution of random noise; $p(data)$ Is the distribution of real samples; $E(\cdot)$ Is the expected value of the calculation. Its network model is shown in Figure1.

.

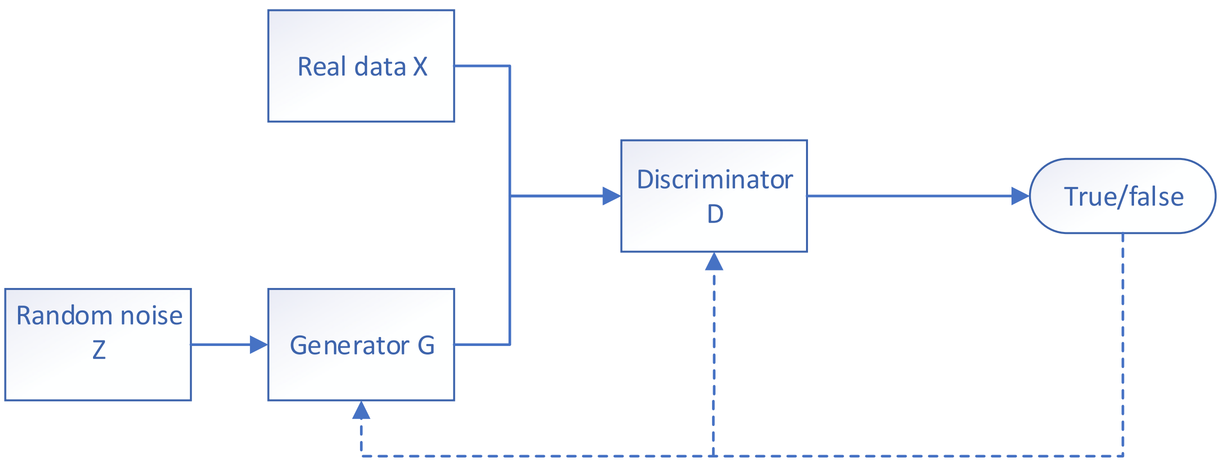

Figure 1. GAN frame

Next, this article will give the following introduction of style migration based on generative adversarial network.

2.2. Pix2pix

Pix2Pix is one of the must-read articles for image style transfer. Pix2pix as an image and model is different from the traditional GAN which uses MLP as the model structure. Pix2Pix uses the model structure of convolution +BN+ ReLU commonly used in CNN. What this structure actually learns is a one-to-one mapping between x and y! In other words, pix2pix is the reconstruction of ground truth: the input contour is decoded into the corresponding vector through Unet encoding and then decoded into the real map. There are three core technologies:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0023

loss function based on conditional GAN, generator based on Unet and discriminator based on PatchGAN.Pix2Pix can achieve amazing results in many image translation tasks, but because its input is image pairs, its resulting model is still limited [3].The limitation here refers to the fact that the model can obtain satisfactory generated content with x approximate to the data set. However, if the deviation between input x and the training set is too large, the results obtained by Pix2Pix will not be so ideal. This kind of data set to find one-to-one mapping, the application appears to be limited, when the data we input and the data in the training set are not in the same mapping range, the generated results appear to be unsatisfactory, which increases the difficulty of our training, to find the data set to cover as many types as possible.

2.3. Cyclegan

The core structure of Cyclegan network is two groups of generative adversarial networks, which are cooperative relations. The first group of generative adversarial network is a generator G (generation from X to Y) and a discriminator D_Y, which is used to judge whether the generated image belongs to domain Y.The second set of generative adversarial networks is the generator (generation from Y to X) and discriminator D_X, used to determine whether the image belongs to the domain. The goal of both generators is to generate as many images of each other's fields as possible to "fool" each other's discriminators. In its previous approach to solving visual and graphical problems such as image style transfer, the main task is to learn the mapping between input and output images using pairs of training sets of images. But for many tasks, matching training data is difficult. However, Cyclegan can perfectly solve this problem in image style transfer. Cyclegan provides a new training mechanism (cycle training, exchange source domain training) and introduces the cycle consistency loss function into the image-to-image conversion, which makes it get rid of the limitation of paired data sets and significantly reduces the pressure of data set preparation. It can present qualitative results on several tasks that do not have paired training data, including collection style transfer, object deformation, seasonal transfer, photo enhancement, etc. [3]. DualGAN[4] is another technology at the same time. Like cyclegan, DualGAN is responsible for unsupervised image-to-image translation. The two articles have the same idea. Both take two sets of unlabeled images as input, each representing an image domain, while learning two reliable image converters from one domain to another, so that a variety of image-to-image translation tasks can be handled.

2.4. Stylegan

StyleGAN is the most advanced high-resolution image synthesis method available and has been proven to work reliably on a variety of data sets. StyleGAN generators can synthesize very realistic images while also enabling image-to-image translation. StyleGAN architecture is noted for its semantically rich, non-entangled, and well-behaved potential space [5].One of the main advantages of StyleGan is the control of the resulting image through style blending, i.e. providing different potential W to different layers when reasoning. In practice, style modulation can magnify some feature maps by an order of magnitude or more. For style blending to work, we must explicitly counteract this amplification on a per-sample basis, otherwise subsequent layers will not be able to manipulate the data in a meaningful way. StyleGAN is not perfect, however. The most obvious drawback is that the resulting images sometimes contain blotchy artifacts. The researchers at NVIDIA released StyleGAN2, an upgrade to Stylegan that focuses on fixing artifacts issues and further improving the quality of the generated images. After StyleGAN2, a new improvement, StyleGAN3, has been proposed successively. StyleGAN3 has added the universal Mapping network style mix, which improves the linear separability of w's training space perception path length.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0023

2.5. AnimeGAN

AnimeGAN is a style transfer that transforms real photos into cartoons [6].It was preceded by a number of reality-to-anime technologies, such as CartoonGan, but they all had the same problem.(1) The animation style texture of the generated image is not very obvious;(2) Some content of the generated image is lost;(3) Parameters in the training require large memory capacity and high requirements for equipment. AnimeGAN is an improvement on CartoonGAN and proposes a more lightweight generator architecture. At the same time, three new loss functions are proposed to improve the visual effect of stylized animation. These three loss functions are gray style loss, gray counter loss and color reconstruction loss.(1) gray scale adversarial loss: make the generated photo texture clear, while maintaining the color saturation;(2) Grayscale Estyle loss: make the generated image style close to the animation style;(3) colorre construction loss: maintain the color of the original drawing.

2.6. U-GAT-IT

U-GAT-IT is a new unsupervised method of image conversion, which introduces a new attention module and a new learnable normalization method from end to end. The attention module will guide the model to obtain the attention map according to the auxiliary classifier, focusing on the different regions between the source domain and the target domain [7].At the same time, an auxiliary classifier is used to identify the difference between the two domains. Then, using the mechanism of CAM, the weight reflecting these differences is taken as the weight of the corresponding feature map to extract and enhance the feature that mainly distinguishes the two domains. CAM serves two functions: it provides the discriminator with an attention mechanism that allows the discriminator to focus locally; It provides a global discriminant to make up for the lack of global discriminant ability required by patchD. Modification Differences (AdaLIN) : The differences found in the first step are used to guide AdaLIN to adaptively modify textures and shapes.

3. CONCLUSION

3.1. Challenges and future research directions

Evaluation indicators. There is no single evaluation index for style transfer. Neurostyle transfer is a kind of artistic creation. Different people may have different or even diametrically opposed views on the results of the same migration. Therefore, it is an urgent problem to put forward a more accurate evaluation index, adopt a unified standard, and build a standardized and universal scientific evaluation system.

Performance versus cost tradeoffs. CycleganGAN, StyleGAN and others have proved that GAN model can be improved in performance. For example, StyleGAN3 can achieve image style migration well, but the huge model requires TPU or multiple Gpus for training, and the cost is high. One possible approach is to compress the model while maintaining performance, using methods such as model quantization, but this is still a serious challenge.

There is still a big gap between the style migration with and without paired data sets, so there must be a certain choice between the two. It remains a serious challenge to achieve unpaired datasets and effects comparable to those using matched datasets.

Compared with image style transfer, video style transfer is a promising research field. Studying advanced video processing technology to achieve high-quality video style transfer will further promote the development of style transfer.

3.2. Conclusion

This paper introduces and summarizes the main models of generative adversarial network to realize style transfer in terms of their principles, existing challenges and future research

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202212_8(12).0023

directions, and injects new vitality into the field of style transfer through generative adversarial model, especially in unsupervised learning, and provides a new algorithm framework by using the adversarial idea of generative adversarial network. It has promoted the development of this field. Although there are still many unsolved problems in style transfer, the way to realize style transfer with generative adversarial network is still being explored. We believe that we can break through these problems and create a better model in the future.

ACKNOWLEDGMENTS

This paper was supported by the Innovation Fund of Sichuan University Of Science and Engineering, the project number is (Y2022169).

REFERENCES

[1] Luo Jia, Huang Jinying: Review of Generative adversarial networks [J]. Chinese Journal of Scientific Instrument, 2019 (3): 74-84.

[2] Xu Zhehao, Chen We: Image style migration based on Generative adversarial network [J]. Software Guide, 2018, 17(6): 207-209.

[3] Zhu J Y, Zhang R, Pathak D, et al: Toward multimodal image-to-image translation[J]. Advances in neural information processing systems, 2017, 30.

[4] Yi Z, Zhang H, Tan P, et al. Dualgan: Unsupervised dual learning for image-to-image translation[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2849-2857.

[5] Karras T, Laine S, Aila T: A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.

[6] Chen J, Liu G, Chen X. Animegan: A novel lightweight gan for photo animation[C]//International symposium on intelligence computation and applications. Springer, Singapore, 2020: 242-256.

[7] Kim J, Kim M, Kang H, et al: U-GAT-IT: An unsupervised attention generation network with Adaptive layer instance normalization for image-to-image translation [J]. 2020.