Research on the Principles and Applications of Generative Adversarial Networks in Image Style Transfer

Yubing Qu

College of Computer and Information Science, Southwest University School of Software, Southwest University, Chongqing, 400715, China

qyb0813@email.swu.edu.cn

Abstract. With the rapid development of machine learning, the field of image style transfer has received widespread attention, and scientists have gradually realized that this is an important field in computer vision. Image style transfer refers to the conversion of an image from one visual style to another and preserving the content. The technology of changing image style has a wide range of applications in various fields from art to virtual reality. Although current learning models have achieved some results in the study of image style transfer, they still face challenges in improving resolution and complex style transfer. This paper mainly introduces the background and characteristics of existing learning models that can achieve image style transfer. It introduces the principles, characteristics and limitations of GANs. In the basis of introduction, this paper also analyses the differences and different characteristics of GANs and summarizes the advantages and disadvantages of GANs. At the same time, this paper predicts future research directions, continuing to optimize these models, reducing training costs, and exploring hybrid methods to further advance image style transfer technology. These developments not only strengthen the theoretical research of image style transfer and GANs, but also have broad application potential in fields such as art, design and medical imaging.

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

1. Introduction

1.1. Background

With the development of science and technology, large models based on machine learning are emerging one after another. Among them, the field of computer vision has also attracted the attention of scientists because of its broad application prospects. Image style transformation is an important research direction in the field of computer vision. It can transform a picture from one visual style to another while retaining the content information of the picture. This technology will be widely used in art creation, advertising design and VR. However, many existing methods still have certain limitations in high resolution and complex style transfer. In recent years, with the rapid development of deep learning, generative adversarial networks represented by GAN models have been widely used and have become the mainstream tool chosen for image style transformation needs. This method can not only generate higher-resolution images, but also achieve complex image style conversion. Therefore, it has attracted widespread attention from scientists.

1.2. Research status

1.2.1. Application of CNN in Image Style Transfer

Alexander Waibel first proposed the concept of Convolutional Neural Networks (CNN). Among them, the Neural Style Transfer (NST) proposed by Gatys et al [1] can achieve image style conversion by optimizing the content loss function and the style loss function to fuse the original image content with the target style image. NST can also complete the image style conversion well without paired data as a training set. However, NST calculation is relatively complex, the result generation process is slow. Because it is very sensitive to image resolution, it may cause the loss or error of image content details during style conversion.

1.2.2. Application of Pix2Pix in Image Style Transfer

Isola et al proposed the Pix2Pix method to transfer image styles in their paper. Pix2Pix is an image-to-image conversion method based on a GAN network [2]. It is particularly suitable for some scenarios with paired images as training sets. After training, it can achieve accurate conversion from the original image to the target style. Due to the L1 loss and discriminator, the image style transfer completed by Pix2Pix has high detail fidelity. However, the training effect of Pix2Pix is not as good as expected when training single data, and the training model requires a lot of resources.

1.2.3. Application of VAE in Image Style Transfer

Variational Autoencoder (VAE) is an artificial neural network structure proposed by Diederik P. Kingma and Max Welling [3], which belongs to the probabilistic graph model and variational Bayesian method. VAE transforms image style by learning the latent space representation of the image and operating the latent variables. VAE can generate diverse style images with stable process, can process large-scale data, and is very suitable for the smooth conversion of images of different styles. However, the images generated by VAE are not detailed enough, and the style transfer is not natural enough, which is particularly obvious in images with more details.

1.3. Research significance

This paper will sort out the principles of GAN, analyze its basic role and limitations in image style transfer, and analyze the differences in the composition structure of different GANs networks to achieve image style transfer, as well as their advantages and disadvantages. It aims to provide subsequent researchers with relevant theoretical foundations and research ideas for image style transfer based on GANs, and provide research background and a wide range of solutions for the future application of image style transfer in various industries.

2. Application of GANs in image style transfer

Please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. This paragraph follows a section title so it should not be indented.

2.1. GAN model principle

Based on machine deep learning models, Generative Adversarial Networks (GAN)[4] is a neural network system with a wide range of potential applications. A generator that produces high-quality data and accomplishes user-requested tasks, like processing audio and images, is created by the GAN model's constant competition with the models produced by the discriminator and the generator, based on the principles of game theory.

The generator and discriminator are the two most crucial parts of the generative adversarial network. The detailed fundamentals of these two networks will be given below.

An input layer, a hidden layer, and an output layer make up the generator's three primary components. A Gaussian distributed vector will be fed into the generator as random noise throughout the GAN model's operation, and it will be produced as pseudo data. This Gaussian distributed vector is often

mapped using a deep neural network (such a convolutional neural network). The generator's goal is to produce samples that closely resemble the original data in order to "fool" the discriminator.

A neural network model for binary classification serves as the discriminator. Its basic components are an input layer, an output layer, and a radial basis hidden layer, making it structurally comparable to the generator. The generator generates actual or pseudo data, which is received by the input layer. The radial basis hidden layer then classifies the incoming data, and the output layer outputs a probability value. To determine if the input data is actual data or fake data produced by the generator, one uses the probability value.
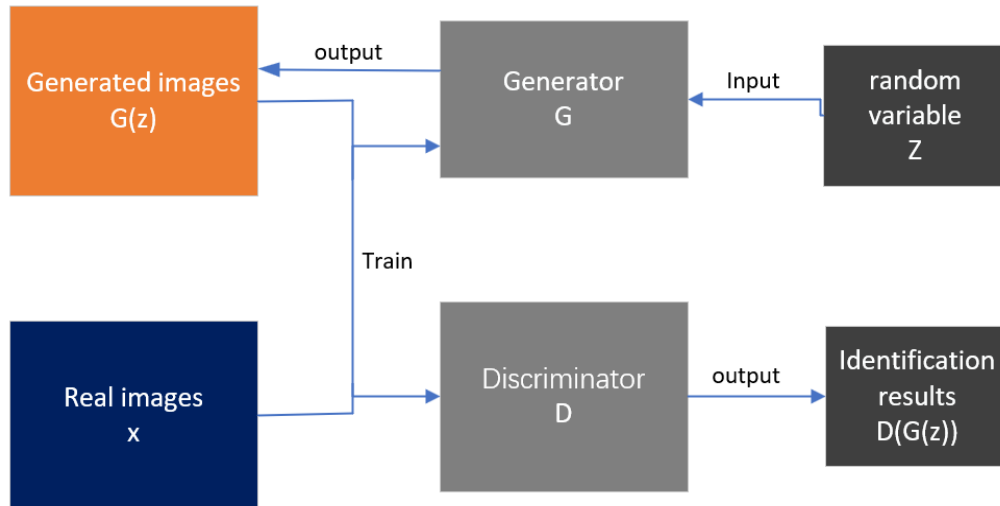

Figure 1. Figure with short caption (caption centred).

As illustrated in figure 1, the training procedure of the GAN model involves the following steps: Initially, the generator uses random noise to produce fictitious samples, or pseudo data. Following the entry of these samples, the discriminator will produce a probability value indicating whether the judgment is true or incorrect. The adversarial process of the GAN model refers to how the generator and discriminator keep becoming better at each other over repeated rounds. The generator eventually produces perfect data samples.

The GAN model may produce clearer and more realistic samples than other models, and it is an unsupervised learning and training approach. However, despite its numerous advantages in the data production process—such as the elimination of complicated Markov chains—it still has certain limitations.

Because the GAN model uses adversarial training, which aims to produce realistic pictures rather than precisely altering image style, it lacks precise control over image style. The generator does not include a technique to directly optimize style aspects; instead, it concentrates on producing realistic visuals. The lack of a single standard makes it impossible to assess the quality of samples produced by the GAN model, and its great sensitivity to hyperparameter variation makes it difficult to make sophisticated adjustments. Furthermore, the GAN model exhibits flaw in its handling of discrete input and is vulnerable to mode collapse during training.

In the study of image style changes, the GANs method is the most widely used. The traditional GAN original model was proposed by Goodfellow et al. in 2014. It consists of a generator and a discriminator and generates high-quality images through adversarial training. However, due to its limitations in image style transformation, based on the traditional GAN, scientists have continuously improved and developed various more advanced derivative models to address its shortcomings and broaden its range of applications. they also can achieve image style transformation. The following lists a number of popular models.

2.2. Application of CycleGAN in Image Style Transfer

Cycle-Consistency Generative Adversarial Network (CycleGAN) [5] is more appropriate for style transfer across different picture domains than GAN. A variation of the conventional GAN model is called CycleGAN. In addition to generating fresh sample data like GAN does, CycleGAN can also transform the sample input, something the GAN model is unable to accomplish. Two generators and discriminators are introduced by CycleGAN in order to accomplish the aforementioned functions. The generator in GAN and generator G serve the same purpose. The two discriminators, respectively, decide if the data created by G and D are real, and the extra generator F is in charge of turning the pseudo data generated by G into source data. CycleGAN has acquired the ability to transform data from two sources—the self-selection of researchers who supply training data sets—based on the aforementioned technological assistance. The primary distinction between CycleGAN and GAN is the novel addition of cycle consistency loss, which guarantees that, upon conversion, the output of the two generators may be returned to its initial state, preserving the image's consistency.

2.3. Application of StarGAN in Image Style Transfer

Image style transfer is possible over multiple domains in addition to single domains. The primary advantage of multi-domain image style transfer network (StarGAN) [6] over the GAN model is its ability to transfer styles across several target domains as opposed to depending solely on one domain to finish picture conversion. StarGAN may effectively minimize the model's complexity and training cost because it only requires one generator and discriminator to handle numerous targets. StarGAN is more effective than GAN, which needs a different generator and discriminator for every domain training. To further enable StarGAN to regulate the style domain of the produced picture and achieve more precise and flexible image style transfer, StarGAN also includes domain labels through the generator and discriminator.

2.4. Application of StyleGAN in Image Style Transfer

The addition of style blocks is Style Generative Adversarial Networks (StyleGAN) [7] most significant enhancement over GAN. A mapping network may be used to translate the input noise to the style space, after which each level of the created picture will get a different set of style parameters. Compared to GAN, StyleGAN's picture can more precisely manage the image style. Furthermore, StyleGAN employs a progressive normalization technique, which entails using progressive training for the training data set and gradually enhancing low-resolution images throughout the training phase. The generator can better catch details thanks to this training process, which raises the caliber and consistency of the images it generates.

3. Comparison and analysis

Through the basic explanation of the above GAN model and its variants, scientists have found that each model has its own characteristics. Table 1 compares the advantages and disadvantages of GANs.

As science and technology have advanced recently, researchers have created a range of generative adversarial networks which could deal with transferring image style, in addition to the more well-known GAN and its derivative networks in the preceding table. For instance, ARGAN, it suggested three loss functions. The issue that classic GAN requires paired training data is successfully resolved, and the network increases the efficiency of picture style transfer. It also does not require paired data during training, and the training speed is quicker. JoJoGAN introduced a Shot Face Image Style Transfer Method with randomly mixed potential coding, it can provide more precise paired data sets [9]. In addition to Shape-Matching GAN, which requires only one image to generate a style and significantly lowers the training cost. PDD-GAN processes data through a band-stop filter, decouples blurred images at the structural level, and suggests a dehazing GAN network based on priors and with decoupling capabilities, which can make blurred images clearer [10].

Table 1. Table of GANs advantages and disadvantages.

	Advantages	Disadvantages
GAN	• The basis of generative adversarial networks. GAN's generation of high-quality samples is a huge advancement in machine learning. • The simple of GANs, Easy to understand and suitable for generating images of various styles.	• May not be able to focus on this style. • No clear style control mechanism, so the style of the generated images cannot be specified. • May prone to mode collapse.
CycleGAN	• Does not require paired data for training. • image style transfer has consistency so that the generated image can be restored to the original image, avoiding unnatural image style conversion. • Can transfer the style of images in different domains, maintaining the diversity of images.	• Unstable during training and prone to mode collapse • Requires higher costs, including longer time and a lot of computing resources When training higher-resolution images. • Will lose some image details, making the image blurry or generate unrealistic images. • The model parameter adjustment is more complicated.
StarGAN	• Can convert between multiple target domains. Only one model is needed to handle multiple domains and multiple styles. It highly efficient and flexible. • Network structure is simpler, which reduces the model's overhead when training existing data sets. • Still maintains the advantages of unsupervised learning.	• The generation quality may not be as good as that of a single conversion task when dealing with complex image style conversion tasks. • Requires complex training and tuning when dealing with multiple domains.
StyleGAN	• Good at generating high-quality, realistic images, especially in processing facial images. • Controls the adjustment and generation of image styles more accurately. More stable during training, when compared with traditional GAN.	• A large amount of computing resources is required, and some artifacts may be generated during the generation process. • Low flexibility in image style transformation. • Highly dependent on data.

The GAN model has made great progress in image style transfer, but it still has problems such as unstable training, being prone to mode collapse, strong dependence on data, and high requirements for training resources. In addition, the quality of GAN generation in complex scenes is unstable, and details of images may be lost or artifacts may appear. Therefore, there is still room for improvement in GAN.

GAN about image style transfer has a wide range of potential applications in the future. For instance, it may improve, fix, and enhance photos used in the medical industry. It may produce creative style pictures, carry out style transfer, and produce artistic masterpieces in the realm of design. In the realm of game development, it may provide virtual characters and sceneries of excellent quality to improve the user experience. It can also be utilized in the field of autonomous driving. The robustness of image recognition systems may be increased and simulated driving situations can be produced via GAN.

Although GAN and its derivative models offer a lot of promise, there are still several issues that need to be resolved via research and development.

4. Conclusion

This research mainly conducts an in-depth discussion on the GANs model in the field of image style transfer. Through the analysis of traditional GAN, CycleGAN, StarGAN and StyleGAN, the characteristics of each in the field of image style transfer are discussed. The CycleGAN model can handle unpaired style transfer. Effectively maintaining content consistency, StarGAN can handle multiple style transfers on one model, expanding the scope of application. The pictures processed by StyleGAN have high resolution and precise details, which is especially suitable for the processing of style transfer of pictures with more details and high requirements for restoration.

With the current level of technological development, it is of great significance to study the conversion of image styles. The research on these GANs not only promotes the development of image style transformation, but also provides effective tools for future AI art creation design and detailed scene transformation of virtual reality. At the same time, GANs expand the team of machine learning models, alleviate the limitations of traditional methods, and enable machine learning models to be applied in a wider range of fields. These research results not only enrich the development of theoretical research, but also promote the progress of related industries and improve people's quality of life.

For GANs models, future research directions on image style changes may focus on further optimizing the performance of the model, reducing training costs, and training to analyse image content details to improve the accuracy of complex image style conversion. In addition, multiple models can be combined to explore the field of hybrid models, which may also produce breakthrough progress.

References

[1] Gatys L A, Ecker A S and Bethge M 2016 Image style transfer using convolutional neural networks *the IEEE conf. on computer vision and pattern recognition* pp 2414-23

[2] Isola P, Zhu J Y, Zhou T and Efros A A 2017 Image-to-image translation with conditional adversarial networks *the IEEE conf. on computer vision and pattern recognition* pp 1125-34

[3] Kingma D P and Welling M 2013 Auto-Encoding Variational Bayes *CoRR*, abs/1312.6114.

[4] Yu Z and Zhang H 2021 Research and practice of image technology based on artificial intelligence (GAN) *the 2021 1st Int. Conf. on Control and Intelligent Robotics* pp 291-4

[5] Liao Y and Huang Y 2022 Deep Learning‐Based Application of Image Style Transfer *Mathematical Problems in Engineering* 2022(1) 1693892.

[6] Xu X, Chang J and Ding S 2022 Image Style Transfering Based on StarGAN and Class Encoder *Int. Journal of Software & Informatics*, 12(2).

[7] Richardson E, Alaluf Y, Patashnik O, Nitzan Y, Azar Y, Shapiro S and Cohen-Or, D. 2021 Encoding in style: a stylegan encoder for image-to-image translation *IEEE/CVF conf. on computer vision and pattern recognition* pp 2287-96

[8] Zenoozi A D, Navi K and Majidi B 2022, February Argan: fast converging gan for animation style transfer *2022 Int. Conf. on Machine Vision and Image Processing (MVIP)* pp 1-5

[9] Chong M J and Forsyth D 2022, October Jojogan: One shot face stylization *European Conf. on Computer Vision* pp 128-5

[10] Chai X, Zhou J, Zhou H and Lai J 2022, October PDD-GAN: Prior-based GAN network with decoupling ability for single image dehazing *30th ACM Int. Conf. on Multimedia* pp 5952-60