Image Translation Based on Picture Ink Painting Conversion

Zixuan Guo
School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China

Keywords: Image translation, Loss Functions, Generative Adversarial Networks, Ink Wash Painting.

Abstract: With the continuous improvement of people's aesthetic level, people's requirements for the aesthetic quality of pictures are also increasing. It leads to the emergence of image style transfer technology, which meets people's requirements for the diversification of image style. At the same time, in the combination of art and technology, the vacancy of image translation in the expression of traditional Oriental art needs to be filled. Therefore, based on CycleGAN, this paper proposes an image style conversion network that can convert ordinary pictures into pictures with ink style, aiming at image translation between photos and ink paintings. The feasibility of the proposed transformation network is verified by experiments on the data set and the evaluation model. The experiment converted the photos into ink-and-wash paintings, and the resulting ink style image has clear stripes and high image quality.

1 INTRODUCTION

Image translation is an important research field of image processing. Image translation will accordingly have two images with different characteristics of the domain transformation. Nowadays, image translation technology has been widely applied to transfer image style. Image style is a generalization of the overall image style seen by human beings. It has a strong attraction to the human visual system and can improve people's ability to recognize and interest in images. Image style transfer refers to converting the image style to another style, which extracts content features from the original image. At the same time, limit the style synthesis, extract style features from the style image, generate the target image in order to retain the semantic content of the target image, such as generation of personalized ceramic patterns (Ning., Liu, Fan 2020), coordination of human faces (Fan, Li, Zhang 2021), and so on.

Generating adversarial networks is an unsupervised learning method that learns by making two neural networks play games. According to the corresponding relationship between the source image and the output image in the training data set, image translation can be divided into supervised and unsupervised (HUANG, YU, WANG 2018). Pix2pix is a typical supervised image translation method. The Pix2pix framework uses pairs of images for image translation. Two different styles of the same image are input, which can be used for style transfer.

This experiment aims to achieve image style transfer, to achieve the conversion of ordinary images to images with ink painting style. Ink and wash painting is a typical representative of traditional Chinese painting art. It has high aesthetic value (Zhang., Yu., Liao. Peng 2021) to create dry, wet and light ink colour by harmonizing ink and wash strokes. However, in the field of art and technology combined, many AI-generated works generate realism, post-modern, and even abstract expressionism, but rarely see AI in traditional Oriental art performance.

Generating adversarial networks is composed of a generation network and a discrimination network. It is a convolutional neural network algorithm based on deep learning which applies to image style conversion for the first time. It uses the neural representation to separate and reorganize arbitrary image content and style to generate art images (Gatys, Ecker, Bethge 2020). In the past, Gan was all unidirectional generation but broke through the limitation of one-to-one correspondence of images in datasets and adopted a bidirectional loop generation structure to retain the image content structure information, which can better establish the mutual mapping relationship between different image domains. In this paper, CycleGAN will be selected to

Guo, Z.
Image Translation Based on Picture Ink Painting Conversion.
DOI: 10.5220/0011768100003607
In *Proceedings of the 1st International Conference on Public Management, Digital Economy and Internet Technology (ICPDI 2022)*, pages 807-812
ISBN: 978-989-758-620-0

convert ordinary images to images with ink painting styles.

2 RELATED WORK

2.1 Generate Adversarial Networks

Many studies have proved generative adversarial networks capable of image style conversion (Chen 2020, Wu, Liu, Wang 2019). However, it is very challenging to train an image generation model that can synthesize the required image generation model to ensure that the generation model can learn the details of the image. A generative adversarial network is a generative model, mathematically expressed as a probability distribution p(x), and a generative model without constraints is an unsupervised model that will be given a simple prior distribution $\pi\ ((z)$ is mapped to the probability distribution p(x) of the pixel of the picture of the training set. An image with the characteristics of the training set following the distribution p(x) is output. The generative adversarial network consists of two neural networks, a generative network G and a discriminant network D. The function of the generator is to try to fit the random noise distribution into the actual distribution PDATA of the training data under the guidance of the discriminant by learning the characteristics of the training set data, to generate similar data with the characteristics of the training set. The discriminant is responsible for distinguishing between the input data is accurate or generated by the false generator data and feedback to the generator. The two networks are trained alternately, and their abilities are improved synchronously until the data generated by the generated network can be regarded as genuine and reach a certain balance with the ability to distinguish the network. In general, the input of G is a random noise vector z obtained from sampling a predefined potential space pz. The training optimization objective of the GAN network can be expressed as the following formula (Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, Bengio 2014),

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[log\, D\,(x)] \\ + E_{Z\sim p_z(z)}[log(\,1 - D(G(z)))]$$

V (D, G) represents the degree of difference between the generated sample and the actual sample, and the cross quotient loss of classification can be used. $\min_G V(D,G)$ indicates that the parameter of discriminant D is updated by maximizing the cross quotient loss $V(D,G)$ with the generator fixed. $\min_G \max_D V(D,G)$ indicates that the generator should minimize the cross quotient loss under the condition that the discriminator maximizes the cross quotient loss V (D, G) of true and false images.

2.2 Cycle Generate Adversarial Networks

Cyclic generate adversarial networks, namely CycleGAN (Zhu, Efros 2017), one of the GaN models, which can also be used for image style transfer tasks. Unlike GaN in the past, which were all generated in one direction, CycleGAN broke through the restriction of one-to-one correspondence of data set images and adopted the structure of bidirectional cycle generation, so it was named CycleGAN. CycleGAN also learns the probability distribution of images pixels in the data set to generate images through the confrontation training of discriminator and generator. In order to complete the image style transfer from domain X to domain Y, it is required that the GAN network should not only fit the pixel probability distribution of the image in domain Y but also maintain the corresponding features of the image in domain X.

3 METHOD

The design of this paper is based on the CycleGAN network structure, which can complete unsupervised image translation tasks without the use of one-to-one paired training data sets. On this basis, a model of image style transfer is designed to formalize the mapping process of existing images to ink painting styles. CycleGAN is structured as follows:

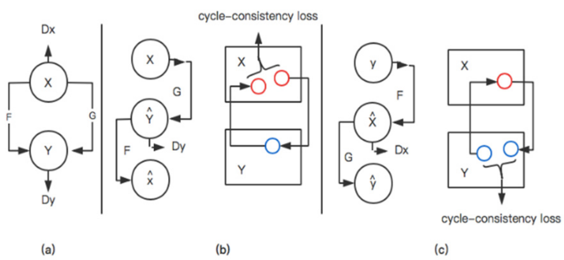

Figure 1: Process of CycleGAN.

CycleGAN structure has four networks, and the first network is the generation (transformation) network named G: X→Y; The second network is the generation (transformation) network named F: Y→X; The third network is named DX for antagonistic network, which identifies whether the input image is X or not. The fourth network, named Dy for the antagonistic network, identifies whether the input image is Y.

Set the actual picture as X and the ink painting picture as Y. G network converts the real picture into the ink painting picture. F network transforms ink painting pictures into real pictures; DX network identifies whether the input image is actual or not; DY network identifies whether the input image is an ink painting image. These four networks have only two network structures. G and F are generative (transformation) networks with the same network structure, and DX and DY are adversarial networks with the same network structure.

The structure of CycleGAN is shown on the next page. The figure shows the placement of two pairs of discriminators and generators. The upper part is the training process of generator G and discriminator DY for x2y, and the lower part is generator F and discriminator DX for y2x. For X and Y domains, two images of the area can pass the generator. Let X generated a Y X - fake image in the domain. Meanwhile, joining a discriminant, judge X-ray images of the fake, if Y domain, the process generally use layer lower generation network, when generating network layer deep, input and output gap will be huge. (Ren 2020) There is no one-to-one correspondence between the two sets of images sent into CycleGAN, so CycleGAN adds loop generation and optimizes consistency loss instead of using constraints to restrict the generator from retaining the image features of the original domain. CycleGAN is composed of adversarial loss and consistency loss.

Adversarial Loss

$$L_{GAN}(G, D_y, X, Y) = E_{y\sim P_{data}(y)}[log\, D_y\,(y)] \\ +E_{X\sim P_{data}(x)}[log(1 - D_Y(G(x))]$$

Consistency Loss

$$L_{CYC}(G, F) = E_{y\sim P_{data}(x)}[||F(G(x)) - x||_l] \\ +E_{X\sim P_{data}(y)}[||G(F(y)) - y||l]$$

Total Loss

$$L(G, F, D_x, D_y) = L_{GAN}(G, D_{y,}X, Y) \\ +L_{GAN}(F, D_x, X, Y) + \mu L_{CYC}(G, F)$$

Where $L_{GAN}(G,\ D_y, X, Y)$ refers to the Adversarial Loss of the X2Y Process. $L_{GAN}(F, D_x, X, Y)$ Adversarial loss refers to the Y2X process. $\mu L_{CYC}(G, F)$ refers to the loss of loop consistency for generators G and F, where the scaling coefficient of consistency loss is a hyperparameter.

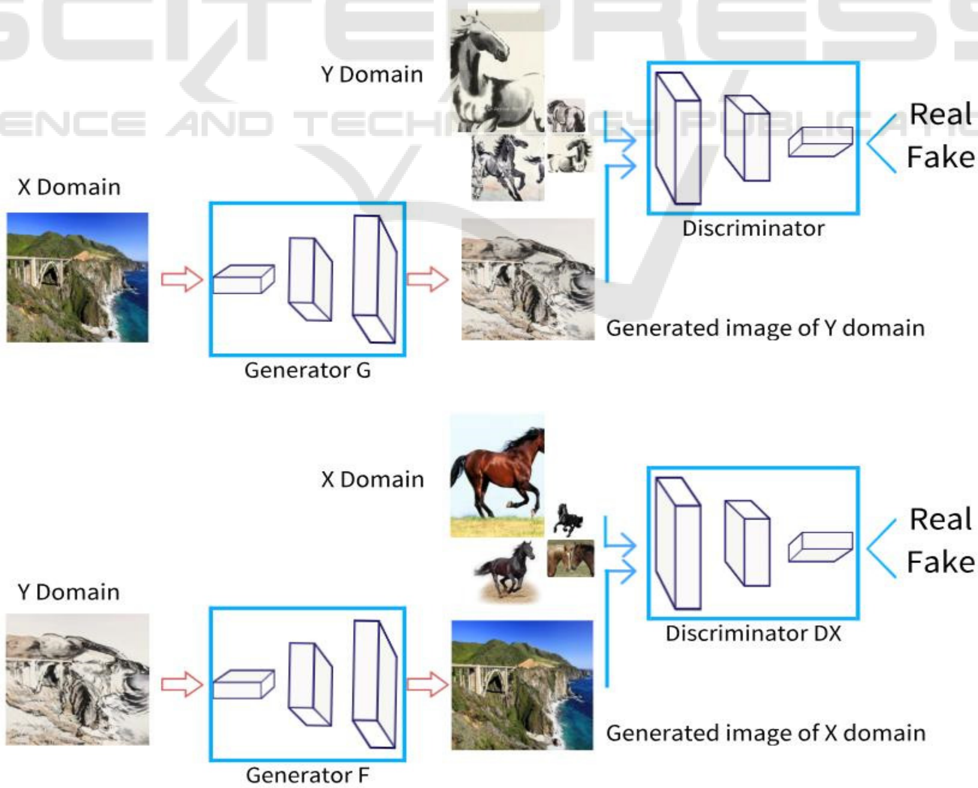

Figure 2: Structure of CycleGAN.

During CycleGAN's normal training, the generator G enters X to generate Y ^. When the ontology mapping loss of generator G is calculated, the generator G enters Y to generate Y ^, and then the L1loss of Y and Y ^ is used as G's Identity Loss. Correspondingly, Identity Loss of generator F is the input X and the generated L1 Loss of X ^. When CycleGAN is optimized, these two parts are added to the total loss of the model if Identity Loss is enabled. Like the cyclic consistency loss, the scale factor super parameter controls its proportion in the total loss.

4 EXPERIMENTS

4.1 Experimental Detail

This article borrowed the WASH_INK_DATASET_FIN dataset from the University of Science and Technology of China.

The environment in which the experiment was configured is as follows:

Table 1: Experimental Environment Configuration.

Hardware configuration	Software configuration
Intel (R) Core (TM) i5-8250UCPU@1.60Ghz 1.80GHz	CUDA Version 8.0
Win10 Professional Edition	PyTorch 0.3.1

The input and output image sizes are set to a 256×256 pixel size. The network structure refers to the generated network structure constructed by Johnsons et al. (Johnson, Alahi and Fei-Fei 2016). The generator is composed of three parts, which are encoder, converter and decoder, respectively.

The encoder consists of four convolutional layers, including two stride-2 causes and two 1/2-strided convolutions. The residual block can make the network deeper and smoother. In addition to weakening the gradient and disappearing, it is also a kind of adaptive depth. The network can automatically adjust the depth. According to the size of the picture, the number of residual blocks used by the model is 9 (Zhu, Park, Isola et al, 2017), and the input and output sizes of the residual network are consistent.

Deconvolution and convolution layers are used in the encoder. After the residual structure of Tensor, the first, second and third layers of deconvolution are successively passed through, and finally, a convolution layer is passed through to get a 256*256 image of three channels. And then it is mapped to -1,1 by tanh.

The Discriminator network uses 70×70 patch-gans, aiming to identify 70*70 overlapping image patches are actual or not. Also, the four-layer convolutional network is used to reduce the number of channels to 1, and finally, the reshape is carried out after pooling averaging. The final output is the discriminant result, which can be applied to images of any size.

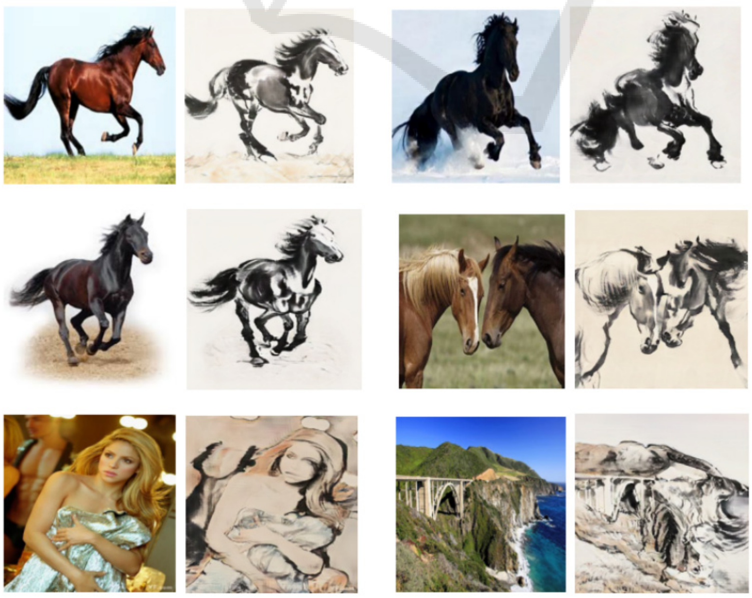

Figure 3: Results of the model.

In order to reduce the oscillation of the model, (Nips 2016) the model will retain an image buffer to store the generated images (Shrivastava, Pfister, Tuzel, Susskind, Wang and Webb 2016).

4.2 Qualitative Evaluation

In this paper, an experiment is conducted on the WASH_INK_DATASET_FIN dataset to convert ordinary photos to ink paintings, and in the asymmetric image style conversion, the experiment is carried out on the current CycleGAN network with better network performance (Ren 2020). Since the suggestion of CycleGAN, the processing of conditional information in discriminant networks is mainly shown in the following ways: firstly, the conditional information is combined with the input in the input layer; secondly, conditional information is connected with the feature vector in the hidden layer of the discriminant network; finally, the discriminant is used to reconstruct conditional information instead of conveying the conditional information to the discriminant. Therefore, the discriminant should learn to judge the authenticity of the image and perform the additional task of image classification. When evaluating the quality of the generated images, the evaluation methods can be divided into two categories: subjective evaluation and objective evaluation. Subjective evaluation refers to evaluating the generated image by the experimental personnel or the third party according to their subjective feelings. The objective evaluation chooses the evaluation index to evaluate the quality of the generated image.

Subjective Analysis

CycleGAN image ink under the framework of the results, as shown in the Figure, in the proposed framework after image translation, the image generated by the model in this paper can retain more picture details and only convert the colour, painting style and other features related to the target domain. It also does not appear deformation. Shading level changes nature, and stroke smooth and orderly, dry wet comparative harmony, picture overall layout space feels good, able to cope with the composition, ink, and paintings three elements of change.

Quantitative Analysis

FID was selected as the image evaluation index to evaluate the quality of the translated image. It is a measure to calculate the distance between the actual image and the eigenvector of the generated image and comprehensively represents the distance between the actual image and the Inception feature vector of the generated image in the same domain. FID has a good discriminant ability. The smaller FID is, the closer the feature distribution of the generated object is to the target feature distribution, and the better the generator effect is [16]. On the contrary, the higher the score, the worse the quality and the linear relationship. This paper uses GAN as the baseline. Table 2 shows the results of FID scoring for the GAN model

And CycleGAN model on the image style transfer task. The results indicate that CycleGAN has a lower FID value than GAN, and it can be considered that CycleGAN has an excellent performance in completing the task of style transformation.

Table 2: FID for GAN and CycleGAN.

Model	GAN	CycleGAN
FID	52.6906	48.2173

5 CONCLUSION

The methods of image style conversion by neural network emerge in an endless stream, and the applications in various directions and fields are also being explored constantly. In this paper, a CycleGAN framework for the style transfer of ordinary image ink painting is proposed. The deep neural network is used to learn the cross-band mapping relationship of images without the need for training pairs of images. The innovation of this paper lies in the combination of the modern network structure model and the traditional ink painting, which realizes the one-click transfer from the actual image to the ink painting. The result meets the requirements and has a certain artistic quality.

Similarly, this study also has some shortcomings, which will be the optimization direction of future research: 1. More scale training sets should be added to solve the problems of a single colour and poor transition of the generated image. 2. Although the model has been proved to be feasible to a certain extent, the structure and training methods have not been modified. Therefore, it has certain limitations.

REFERENCES

A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. arXiv preprint arXiv:1612.07828, 2016. 3, 4, 5

Chen, JC "Image style transfer of Chinese painting based on neural network", [D]. Hanzhou Electronic Science and Technology University, 2020.

Fan LL, Li Y, Zhang XX, "Key face contour region cartoon stylized generation algorithm[J]", Journal of graphics, 2021, 42(01):44-51.

HeuselM, RamsauerH, UnterhinerT, etal. "GAN strained by a two time-scale update rule converge to a local Nash equilibrium" [C]. Advancesin Neural Information Process Sing Systems, December 4-9,2017, long Beach, CA, USA. Network: Curran Associates, 2017: 6626-6637.

HUANG H, YU P S, WANG C. "An introduction to image synthesis with Generative Adversarial Nets" [J/OL]. arXiv e-prints.2018-03-12. https://arxiv.org/abs/1803.04469v2.

I. Goodfellow. Nips 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160, 2016. 2, 4

JY Zhu, T. Park, Phillip Isola Alexei A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", 2017 IEEE International Conference on Computer Vision.

J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, pages 694–711. Springer, 2016. 2, 3, 4

L. Gatys; A. Ecker; M. Bethge "A Neural Algorithm of Artistic Style" Journal|[J]Journal of VisionVolume 16, Issue 12. 2016. PP 326-326

Ian J. Goodfellow, Jean Pouget-Abadie∗, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair†, Aaron Courville, Yoshua Bengio‡ "Generative Adversarial Nets", arXiv:1406.2661v1 [stat.ML] 10 Jun 2014.

Ning. HY, Liu,J, Fan, YB, "Application of generative adversarial network algorithm in personalized ceramic pattern generation[J]" ceramics. 2020(02):24-27.

Wu.HM, Liu.RX, Wang.YH, "Face image translation based on generative adversarial network" [J]. Journal of Tianjin University (Natural Science and Engineering), 2019,52(03):306-314.R.

Zhang. JJ, Yu.JH, Liao. YW. Peng.R, "Adaptive Computational Aesthetic Evaluation of Ink Painting Based on Deep Learning" [J/OL]. Journal of Computer Aided Design and Graphics:1-12 [2021-06-16]. http://kns.cnki.net/kcms/detail/11.2925.tp.20210531.2105.009.html.

ZQ. Ren, "Image Style Conversion Based on GaN Network", [D]. Beijing Jiaotong University, 2020. P15-P16

Zhu, J.Y., Park, T., Isola et al, 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. [C] In Proceedings of the IEEE International Conference on Computer Vision (pp. 2223-2232)

ZQ. Ren, "Image Style Conversion Based on GaN Network", [D]. Beijing Jiaotong University, 2020. P35.