Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230118

The style transfer of photos and landscape paintings based on CycleGAN combined with nested edge detection module

Yueling Jin[1, 4, †], Zeyang Li[2, †], Tiantian Lu[3, †]

[1]The department of engineering, Shanghai Ocean University, Shanghai, 201306, China
[2]The department of engineering, University of California Santa Barbara, Santa Barbara, 93117, United States
[3]The department of engineering, University of California Irvine, Irvine, 92697, USA

[4]2052604@st.shou.edu.cn
†These authors contributed equally.

Abstract. Chinese landscape paintings, especially Chinese landscape ink paintings have always been the treasure of the Chinese, or even the world culture. Due to the advancement of the science and knowledge, there are always interests for finding ways to produce Chinese landscape paintings through the way of the technology. In this paper, based on the original CycleGAN model, we are trying to transfer a real-world photograph to a typical Chinese landscape painting. To be specific, the brushwork in Chinese ink paintings contains enormous numbers of distinct and technical brush strokes which makes it extremely hard for CycleGAN to recognize and transfer Chinese ink painting. We improved the brushstroke effect of Chinese ink painting still by incorporating the Holistically Nested Edge Detection (HED) method into the CycleGAN model. The HED module puts the source image and the corresponding generated image into the convolution layer of five stages to generate their respective edge maps. The balanced cross-entropy loss calculated with edge maps is added to the total loss of CycleGAN to train the generator which is confronted with discriminator to improve output results. The addition of HED enables the extraction of edge features from the images, preserving the structural information and enhancing the accuracy of the style transfer, so the detailed of the Chinese ink painting can be transferred better in this sense. Experimental process with model building, training, validating, and predicting concludes that, complementing HED method into the original CycleGAN models well preserved distinct features in the conversion processes from real-world photographs to traditional Chinese landscape paintings.

Keywords: Chinese landscape painting, style transfer, CycleGAN, HED.

1. Introduction

Chinese landscape painting, one of the representative arts in the Chinese history, has played the essential role in Chinese cultural influence. It is a style of painting that incorporates calligraphy and muted color in the brushstrokes. Nowadays, technology is capable to realize this thousand-year-old cultural masterpiece. Machine learning is currently gaining popularity as a result of the modern era's rapid technological progress. For instance, some learning models e.g. InceptionNet, ResNet, and DenseNet achieved high accuracy in image classification. In terms of generating images with certain types, some

DOI: 10.54254/2755-2721/8/20230118

famous model like Convolutional Neural Network (CNN) is used for the artistic style conversion of images [1]. CNN using the features extracted by the pre-trained VGG network to recombine the content of any given picture and the style of the artistic picture to complete the style transfer [1]. The produced image's quality was displayed in a room with real painters. In general, transferring a real-world photograph to a typical Chinese landscape painting is possible with the use of a machine learning model.

For the development history of Generative Adversarial Network (GAN), it was first inspired by the noise-contrastive estimation, as it contains the same loss function as GANs. Some similar ideas were also raised at that time. For instance, Olli Niemitalo introduced an idea involving adversarial networks and published it in 2010, but he never implemented the adversarial network and late was named as conditional GAN [2]. Formally, GAN, as machine learning frameworks, was formally designed and invented by Ian Goodfellow in June 2014[4]. To be specific, the primary purpose of GAN is to generate new information or data with providing statistics by the training set. For example, a GAN can generate new photos by training photos datasets that are remarkably similar to the trained photo, even authentic to human observation. GAN was helpful for unsupervised learning, but also can be worked in environments of semi-supervised learning, supervised learning, or reinforcement learning.

CycleGan is a variant of the GAN network that focuses on the image style conversion function and does not require the corresponding appearance of two styles of images [5]. For instance, it can be utilized to train pictures to achieve style transformation. Notably, it can be used to transform and train pictures without large amounts of the database, such as forming Vincent Van Gogh's styles' arti pieces as there is a limited number of drawing created by him, resulting in the limited data set. This problem cannot be achieved by pix2pixGAN as it needs a pairwise dataset in order to perform the training process [6], but this is not required by the CycleGAN. It is very powerful and practical to convert two types of images. We hope to use CycleGAN to realize the transfer of real landscape photos into Chinese landscape paintings through style transfer.

Despite the success of CycleGAN in transforming images between different styles, there is still room for improvement in the fine details, particularly in transforming images into Chinese painting styles. Compared to Western oil paintings, Chinese ink paintings possess numerous distinctive features, such as a monochromatic palette, blank spaces for imaginative interpretation, and the intricate and varied brushwork. The brushwork in Chinese ink paintings holds a potent allure that can express the atmosphere and vitality of nature through the variations in line thickness, bendiness, and the contrast between black and white hues. It is considered as one of the quintessences of traditional Chinese art. Consequently, only relying on CycleGAN presents difficulty in clearly recognizing and outlining objects in Chinese ink paintings due to the distinctiveness and intensity of the brush strokes. To tackle this issue, further improvement of models is required.

This research aims to improve the brushstroke effect of Chinese ink painting style transfer to a certain extent by incorporating the Holistically-Nested Edge Detection (HED) into the original CycleGAN model [7]. The addition of HED enables the extraction of edge features from the images, preserving the structural information and enhancing the accuracy of the style transfer. The loss of CycleGAN has been augmented with a re-designed cross-entropy loss calculated with the extracted edge maps of the real and fake images, effectively addressing the issue of blurred edges in ink paintings.

2. Method

2.1. Dataset description and preprocessing

The dataset used in this study consists of two parts, namely photographs of actual landscapes (i.e. dataset-A) and photographs of Chinese landscapes (i.e. dataset-B). Prior to the experiment, we obtained the real-world landscape images from the Baidu Web browser using a web crawler to complete the dataset-A. We located 155 examples of Chinese landscape paintings from the Princeton University Art Museum for dataset-B[8]. Some sample data from dataset- A and dataset-B can be observed in Figure 1 and Figure 2. In terms of preprocessing, first we used the BICUBIC technique to resize the entire image. We converted all of the Chinese landscape paintings from their original, variable-sized dataset to

DOI: 10.54254/2755-2721/8/20230118

256x256. The image was then randomly cropped and horizontally flipped. In order for the three-channel value of the image to fall inside the range of 0 to 1, it is also necessary to scale the image's R, G, and B elements using the ToTensor function. Using the value after scaling, we normalized data by subtracting 0.5 as the average and dividing 0.5 as the standard derivation to limit the range in -1 to 1.

(a) **(b)**

Figure 1. Some samples in the dataset-A.

(a) **(b)**

Figure 2. Some samples in the dataset-B.

2.2. CycleGAN combined with HED module

2.2.1. Introduction for CycleGAN. In terms of the Generative Adversarial Networks (GANs) method, it generates the new dataset that represents the training data. For instance, the generative adversarial network can generate photos that look exactly like human faces, while these generated photos are not any real person's photo. And GANs' method to generate such images and data instances is using a generator and a discriminator and pairing them together: while the generator works to create the target output, and the discriminator works to identify the real data, real human faces in this instant, from the generator output [9].

Then, for the method of CycleGAN, different from the traditional GAN, CycleGAN consists of two generators and two discriminators. It is a type of unsupervised image-to-image translation model that is designed to learn a mapping between two different image domains without using any paired training data. Specifically, generator G learns to generate images in the target domain Y from the source domain X, while generator F learns to generate images in the source domain X from the target domain Y. CycleGAN introduces cycle consistency loss used to ensure that the generator produces images that are consistent when translated back and forth between the two domains. To produce images that are as close to the real images as possible, adversarial loss is used to train the discriminators, Dx and Dy, to distinguish between real and fake images in the target domain. In addition, identity loss is used to ensure

DOI: 10.54254/2755-2721/8/20230118

that the generator does not modify the input image when it is already in the target domain. This loss computes the difference between the input image and the generated image.

For the correspond steps described above, it can be divided into 3 steps: 1) The data x in X field is output by generator G to get Y', and the discriminator Dy is used to calculate the additive loss. 2) Y' obtains the reconstruction result x' through the output of the inverse generator F. It is expected to be completely consistent with the input content. The cycle consistency loss is calculated between x and x'. 3) Exchange the source domain and target domain, and calculate the cycle-consistent loss between y and y' [5].

2.2.2. Introduction for HED. Holistically-Nested Edge Detection (HED) is a classical edge detection algorithm that performs pixel-level edge detection on input images to identify edges and contours within them [10]. The first advantage of HED is that its training and prediction are performed end-to-end, meaning that the network's input is the original image, and the output is a binary image of the detected edges. The second advantage is that it extracts rich features for intermediate details through different scales and outputs the results after each convolutional layer. This allows the model to gradually optimize the final result by continually inheriting and learning from the accurate edge detection results, resulting in better edge detection results than directly using the output layer.

The HED network uses the VGG16 network as its backbone and makes two modifications. The first modification is to reduce training time by removing the last pooling layer and subsequent fully connected layers, while retaining the first five stages (Figure 3).

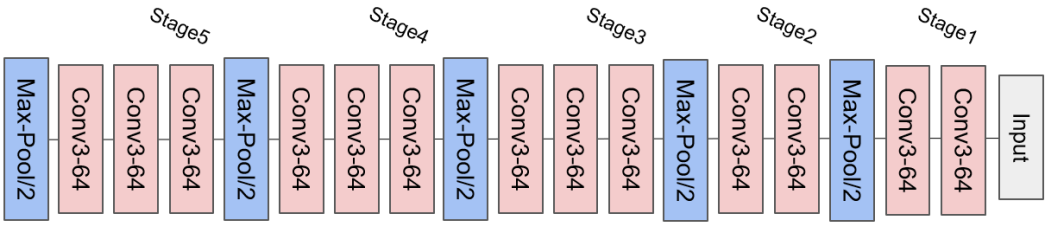

Figure 3. The structure of HED.

Secondly, side output layers are incorporated after the last convolutional layer in each stage of the network to obtain edge maps at different stages. Unlike conventional edge detection that only performs pixel-level detection, we use a multi-level edge detector to obtain five different types of strokes through regression training [7]. In each of the five stages in the HED network, there is a side output layer after the last convolutional layer, and each side output is upsampled using bilinear interpolation to restore the image to its original size. The five outputs are then fused and added according to the learned weights during training to obtain the final output of the model [10].

2.2.3. The structure of the CycleGAN with HED. In this study, we apply the CycleGAN framework to perform image style transfer, with the inclusion of the HED technique. The CycleGAN network contains two generators and two discriminators to achieve image style transfer. The generators consist of an encoder, a converter, and a decoder. The encoder is a three-layer convolutional network that uses ReflectionPad to pad all boundaries of the input image to enhance image resolution. After each layer of convolution, Instance Normalization is used to normalize each individual channel of a single image. The converter component takes in 256x256 images and uses nine residual blocks, which allows the CycleGAN network to train deep models with high precision by skipping unhelpful layers [11]. The decoder uses two transposed convolutional layers and one convolutional layer to restore the image to its original size. The discriminator contains five convolutional layers, followed by average pooling and the final output of the discriminator indicates the decision regarding the input image.

DOI: 10.54254/2755-2721/8/20230118

The CycleGAN framework is modified by embedding the HED technique (Figure 4), which includes an edge detection branch that is connected to the CycleGAN generator to retain the original image's fine details.

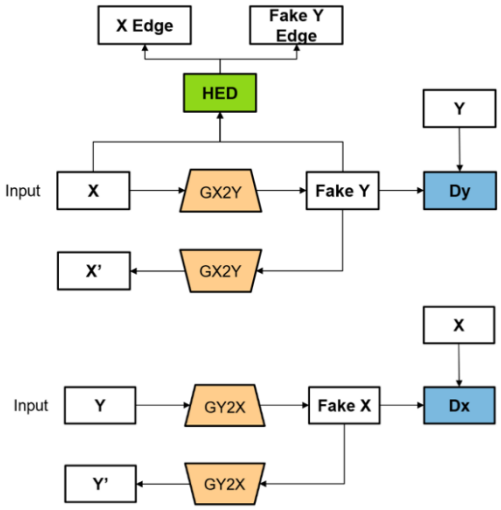

Figure 4. The structure of CycleGAN combined with HED module.

The HED module takes in the source image (X) and the corresponding generated image (Fake Y), generating edge maps for both images. The edge detection map of the source image is taken as the ground truth to calculate the balanced cross-entropy loss [7], which is added to the overall loss of the CycleGAN generator for backpropagation and updating the model parameters. The loss function is as:

$$\mathcal{L}_{Brushstroke}(G,X) = E_{x\sim p_{data}(x)}[-\frac{1}{N}\sum_{i=1}^{N}\mu E(x)_i log E\big(G(x)\big)_i + (1-\mu)(1-E(x)_i)\log(1-E\big(G(x)\big)_i)] \quad (1)$$

2.2.4. Implementation details. All experiments were carried out based on the PyTorch framework with RTX A4000 GPU. We chose 200 for the training epoch, 0.0002 for the learning rate, and 1 for the batch size as the important parameters. Our beta values are 0.5 and 0.999, and the decay epoch is 36. Adam optimizer is being used. When the decay epoch arrives, the interface under PyTorch LambdaLR is utilized to help modify the learning rate.

The total loss of the model is the sum of the identity loss, adversarial loss, consistency loss, and edge loss. Traditional CycleGAN do not have the Edge loss. With the addition of the HED, edge loss can calculate balanced cross entropy loss so that generator can generate better brush strokes.

$$\mathcal{L}_{Total} = \mathcal{L}_{GAN}(G,D_Y,X,Y) + \mathcal{L}_{GAN}(F,D_X,Y,X) + \lambda\mathcal{L}_{cyc}(G,F) + \beta\mathcal{L}_{Identity}(G,F,X,Y) + \gamma\mathcal{L}_{Brushstroke}(E,G,X) \quad (2)$$

$$\mathcal{L}_{\text{Identity}}(G,F) = E_{y\sim Pdata(y)}[||G(y)-y||_1] + E_{x\sim Pdata(x)}[||F(x)-x||_1] \quad (3)$$

$$\mathcal{L}_{\text{GAN}}(G,Dy,X,Y) = E_{y\sim Pdata(y)}[\log D_Y(y)] + E_{x\sim Pdata(x)}[\log(1-D_Y(G(x)))] \quad (4)$$

$$\mathcal{L}_{\text{GAN}}(F,Dx,X,Y) = E_{x\sim Pdata(x)}[\log D_X(x)] + E_{y\sim Pdata(y)}[\log(1-D_X(F(x)))] \quad (5)$$

$$\mathcal{L}_{\text{cyc}}(G,F) = E_{x\sim Pdata(x)}[||F(G(x))-x||_1] + E_{y\sim Pdata(y)}[||G(F(y))-y||_1] \quad (6)$$

DOI: 10.54254/2755-2721/8/20230118

$$\mathcal{L}_{\text{Brushstroke}}(\mathrm{G}, \mathrm{X}) = \mathrm{E}_{\mathrm{x}\sim\mathrm{Pdata(x)}}[-\frac{1}{N}\sum_{i=1}^{N}\mu E(x)_{\mathrm{i}}\log\mathrm{E}(\mathrm{G(x)})_{\mathrm{i}} + (1-\ \mu)(1-\mathrm{E(x)_i})\log(1-\mathrm{E}(\mathrm{G(x)})_{\mathrm{i}})] \quad (7)$$

3. Result and discussion

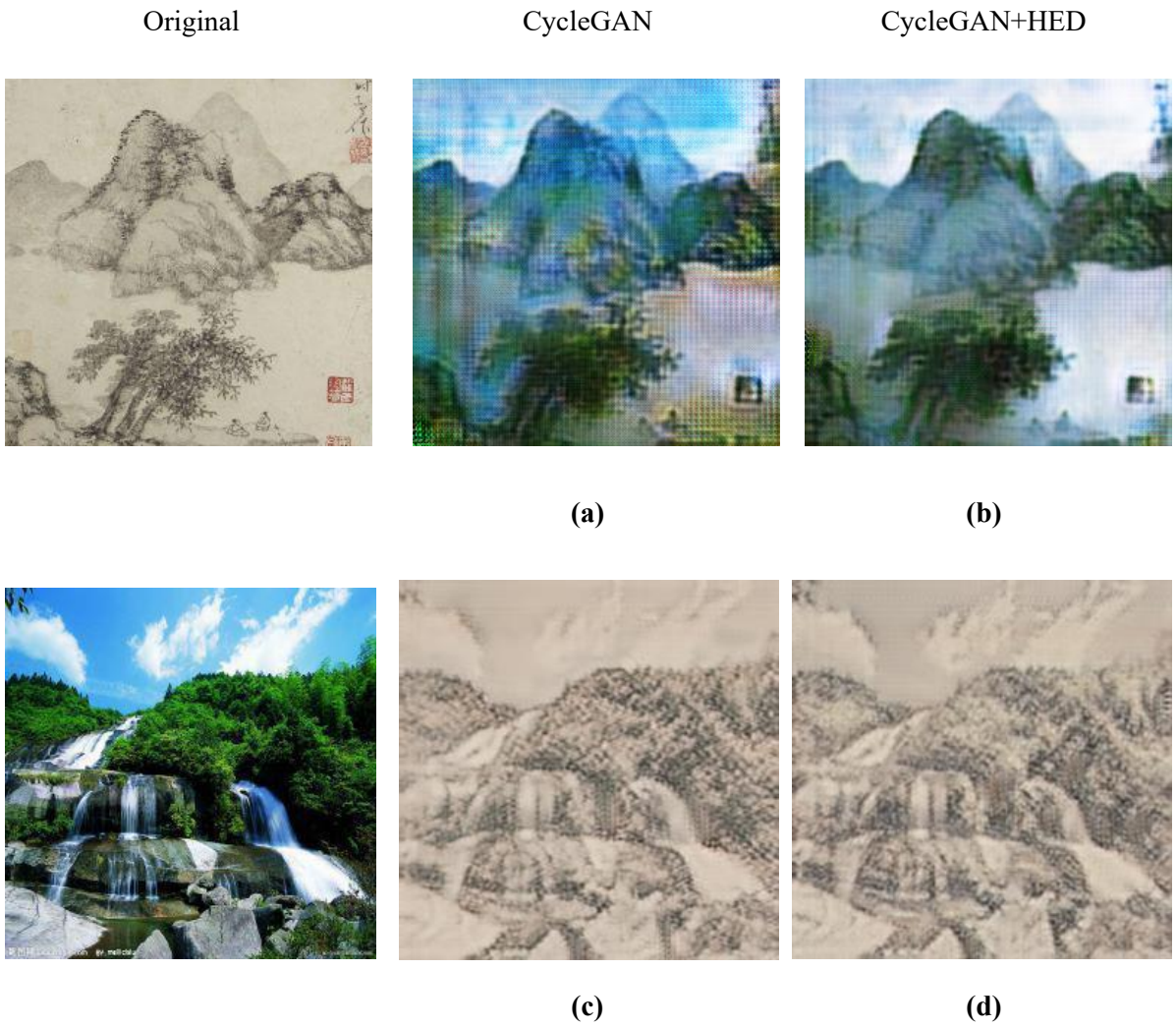

Figure 5. The generated images based on the training with 50 samples.

DOI: 10.54254/2755-2721/8/20230118

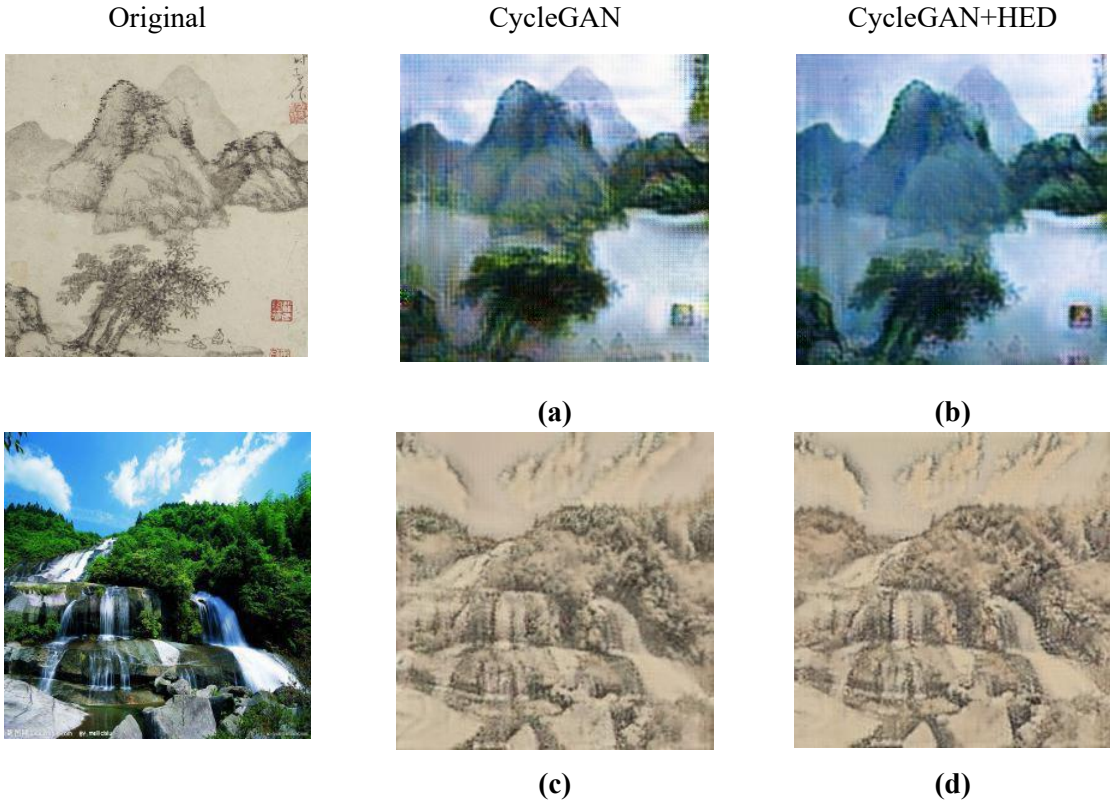

Figure 6. The generated images based on the training with 100 samples.

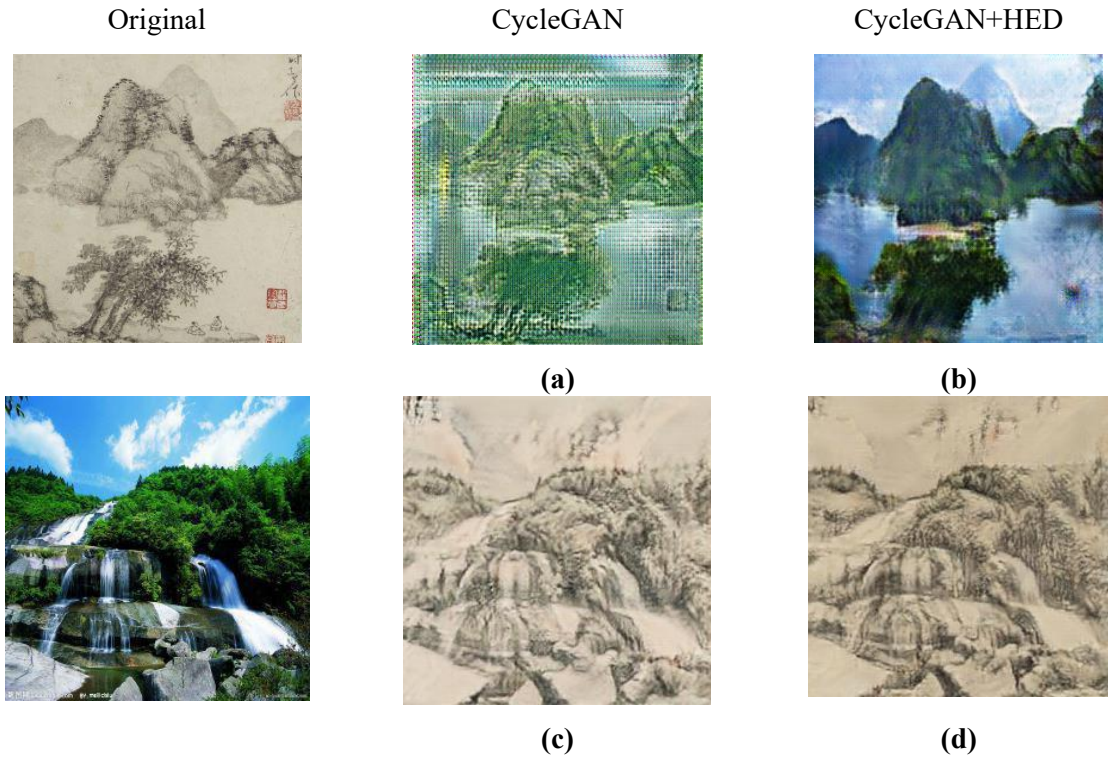

Figure 7. The generated images based on the training with 155 samples.

DOI: 10.54254/2755-2721/8/20230118

From the Figure 5, Figure 6 and Figure 7, it can be observed that the more samples we trained, the higher quality of generated images. Figure 5. b's coloration and brightness are superior to Figure 5. a's with the addition of the HED during the stage of 50 train samples. The brushstrokes in Figure 5.d are thicker in the top region and left side compared to Figure 5.c, while they are lighter on the right side. In general, the differences are not great, and the hue of the paintings created in the Chinese style varies.

As we raised the number of training dataset to 100, the resulting image quality improved. The colors and brightness are still better in Figure 6.b than they are in Figure 6.a. When compared to the figure in the section for 50 samples, the checkerboard artifacts, or blur effect, on the image, diminished.

Figure 7.a, the generated image without HED, has been entirely driven by the checkerboard artifacts over 155 samples. Figure 7.b, on the other hand, has the best effect when HED is added as compared to the previous section. The amount of checkerboard artifacts has been minimized.

In image style transfer, models need to convert source images into target images. However, due to differences in feature representation between the source and target images, it is difficult for the model to convert all features in the source image to the target image. In this conversion process, some detailed features, such as edge features, are often ignored or lost, as edges are high-frequency details in the image that change abruptly. In practical implementation, edge detection techniques are usually realized through convolutional neural networks (CNN). CNN typically detects edge features by using specific filters, which are pre-trained convolutional kernels that can recognize changes in brightness and color in the image, thus helping to identify edges. Edge detection techniques can recognize and capture edge features in the source image and integrate these features into the conversion process to preserve the details of the image.

4. Conclusion

In general, transferring the real-world image to Chinese painting style using CycleGAN is achievable. Typically, there are significant feature discrepancies between source and target images, making it challenging to accurately convert all of the source images' features into the target images. We integrated HED into the original CycleGAN to better resolve the brushstrokes effect of Chinese ink painting style. HED allows the model to extract the edge features of the image to enhance the accuracy of the transfer process. With the addition of the HED, the quality of the generated image has improved compared to the traditional CycleGAN. The generated images have better brightness and coloring, less checkboard effect that decreases the clearness of the image. The resulting image still leaves a lot of room for improvement in sharpness. Future plans include using the model to make more intricate transformations and improving the algorithm to produce more finely defined images.

References

[1] Zhang G Chen J L Song J et al 2020 Chinese Landscape Painting Automatic Generation Model Based on Adversarial Generation Network (in Chinese), Phase 3 Computer and Telecommunications p 6

[2] Zhao J Li F F 2023 A GAN-based Lightweight ink Painting Style Transfer Model (in Chinese) Volume 36 Issue 2 Electronic Science and Technology p 6

[3] Niemitalo O 2010 A method for training artificial neural networks to generate missing data within a variable context Internet Archive (Wayback Machine). Archived from the original on March 12 2012 Retrieved February 22 2019

[4] Goodfellow I and Pouget-Abadie J and Mirza M et al 2014 Generative Adversarial Nets (Massachusetts:MIT Press/Neural Information Processing Systems)

[5] Zhu J Y and Park T and Isola P et al 2017 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks (IEEE)

[6] Isola P and Zhu J Y and Zhou T et al 2016 Image-to-Image Translation with Conditional Adversarial Networks Proceedings of the IEEE conference on computer vision and pattern recognition pp1125-1134

[7] He B and Feng G and Ma D et al 2018 ChipGAN: A Generative Adversarial Network for Chinese

DOI: 10.54254/2755-2721/8/20230118

Ink Wash Painting Style Transfer (ACM/Multimedia Conference)
[8] Xue A 2020 End-to-End Chinese Landscape Painting Creation Using Generative Adversarial Networks Proceedings of the IEEE/CVF Winter conference on applications of computer vision pp 3863-3871
[9] Google 2022 Introduction | Machine Learning | Google Developers https://developers.google.com/machine-learning/gan#:~:text=Generative%20adversarial%20networks%20(GANs)%20are,belong%20to%20any%20real%20person
[10] Xie S and Tu Z 2016 Holistically-Nested Edge Detection IEEE/International Conference on Computer Vision
[11] He K and Zhang X and Ren S et al 2016 Deep Residual Learning for Image Recognition (IEEE)