applied sciences

MDPI

Article

Paint-CUT: A Generative Model for Chinese Landscape Painting Based on Shuffle Attentional Residual Block and Edge Enhancement

Zengguo Sun [1,2,*], Haoyue Li [2] and Xiaojun Wu [1,2]

[1] Key Laboratory of Intelligent Computing and Service Technology for Folk Song, Ministry of Culture and Tourism, Xi'an 710119, China; xjwu@snnu.edu.cn
[2] School of Computer Science, Shaanxi Normal University, Xi'an 710119, China; haoyuelee@snnu.edu.cn
* Correspondence: sunzg@snnu.edu.cn

Abstract: As one of the precious cultural heritages, Chinese landscape painting has developed unique styles and techniques. Researching the intelligent generation of Chinese landscape paintings from photos can benefit the inheritance of traditional Chinese culture. To address detail loss, blurred outlines, and poor style transfer in present generated results, a model for generating Chinese landscape paintings from photos named Paint-CUT is proposed. In order to solve the problem of detail loss, the SA-ResBlock module is proposed by combining shuffle attention with the resblocks in the generator, which is used to enhance the generator's ability to extract the main scene information and texture features. In order to solve the problem of poor style transfer, perceptual loss is introduced to constrain the model in terms of content and style. The pre-trained VGG is used to extract the content and style features to calculate the perceptual loss and, then, the loss can guide the model to generate landscape paintings with similar content to landscape photos and a similar style to target landscape paintings. In order to solve the problem of blurred outlines in generated landscape paintings, edge loss is proposed to the model. The Canny edge detection is used to generate edge maps and, then, the edge loss between edge maps of landscape photos and generated landscape paintings is calculated. The generated landscape paintings have clear outlines and details by adding edge loss. Comparison experiments and ablation experiments are performed on the proposed model. Experiments show that the proposed model can generate Chinese landscape paintings with clear outlines, rich details, and realistic style. Generated paintings not only retain the details of landscape photos, such as texture and outlines of mountains, but also have similar styles to the target paintings, such as colors and brush strokes. So, the generation quality of Chinese landscape paintings has improved.

Keywords: Chinese landscape painting generation; generative adversarial networks; shuffle attention; perceptual loss; edge loss

Citation: Sun, Z.; Li, H.; Wu, X. Paint-CUT: A Generative Model for Chinese Landscape Painting Based on Shuffle Attentional Residual Block and Edge Enhancement. *Appl. Sci.* **2024**, *14*, 1430. https://doi.org/10.3390/app14041430

Academic Editor: Michail Panagopoulos

Received: 13 January 2024
Revised: 2 February 2024
Accepted: 5 February 2024
Published: 9 February 2024

1. Introduction

Chinese landscape painting is an important branch of traditional Chinese painting, which primarily describes the natural landscape of mountains and rivers. As an important expression of Chinese culture, it possesses unique artistic charm and re-creation value [1]. However, creating landscape paintings is a difficult, time-consuming, and professional task. These problems are not conducive to the creation of landscape paintings. Therefore, researching the intelligent generation of Chinese landscape paintings from photos can not only promote the development of artistic creation but also benefit the inheritance of traditional Chinese culture, which has high theoretical and practical value.

Currently, the generation methods of Chinese landscape painting are based on traditional methods and deep learning methods. Traditional methods are divided into physical modeling and non-photorealistic rendering. Physical modeling is mainly used to establish a suitable model by analyzing the brush strokes, paper, and ink diffusion [2–4]. However,

modeling based on physical characteristics is complex and difficult to use to simulate an entire painting. Non-photorealistic rendering is based on the analogy of images, which is stylized by generating a two-dimensional image with artistic style [5]. However, it is difficult to simulate the rendering of different styles.

In recent years, artificial intelligence has flourished globally and reshaped the way people acquire cultural heritage and art. Digitalization has become one of the main ways to preserve and disseminate intangible cultural heritage [6,7]. Generating landscape paintings based on deep learning is widely used and is mainly divided into two categories: the convolutional neural network (CNN) and generative adversarial network (GAN). Inspired by the convolutional neural network, Gatys et al. [8] proposed neural style transfer (NST) in 2015, which uses the convolutional neural network to reproduce painting styles on natural pictures. Li et al. [9] proposed a neural style transfer method for traditional Chinese paintings. Sheng et al. [10] proposed a new model of Chinese painting style transfer, which combines four key features as restrictions into the CNN, including brush stroke, ink diffusion, space reservation, and yellowing. Li et al. [11] extracted the style features and transferred photos to Chinese paintings based on VGG.

The generative adversarial network (GAN) [12] is a framework proposed by Lan Goodfellow in 2014, which is widely used in industries and fields due to its great versatility. Alice Xue [13] proposed a model for generating landscape paintings named SAPGAN, which can generate landscape paintings with unconditional input. Lin et al. [14] proposed a generative adversarial network to generate Chinese paintings from sketches and photos. Gu et al. [15] proposed a layout adjustable simulated generation method for Chinese landscape paintings based on CGAN, which can generate landscape paintings from a layout label map. He et al. [16] proposed an end-to-end generative adversarial network named ChipGAN, which can transfer photos to Chinese ink wash paintings based on CycleGAN [17]. Zhou et al. [18] used CycleGAN to generate landscape paintings from sketches and develop an interactive system called Shanshui-DaDA. Zhang et al. [19] proposed a CycleGAN-AdaIN framework to transfer photos to Chinese ink paintings. Peng et al. [20] proposed an image style transfer framework to transfer landscape photos to landscape paintings. Zhu et al. [21] proposed a novel BiTGAN to perform better in the style transfer of Chinese ink paintings. At present, most of the researches on Chinese painting generation are based on generative adversarial networks. However, due to the unique artistic style and line features of landscape paintings, the present generated results may have some problems, including detail loss, blurred outlines, and poor style transfer. These problems still need to be solved. Therefore, when generating landscape paintings from photos, the following points need to be noticed: highlighting the style characteristics of landscape paintings, recovering the rich details, and reflecting the outline information of the scenery.

In conclusion, in order to solve the problems in landscape painting generation, the Paint-CUT based on contrastive unpaired translation (CUT) [22] is proposed to generate Chinese landscape paintings with unique styles and rich details from photos. CUT is an image translation model based on a generative adversarial network, which introduces contrastive learning into the image translation domain. However, Chinese landscape painting is a combination of line modeling, ink wash, and artistic conception. Due to the unique artistic style, the generated landscape paintings obtained from photos by using CUT directly are not good; they have poor style transfer, blurred details, and blurred outlines. To address these problems of CUT, the Paint-CUT is proposed. Specifically, in order to solve detail loss and highlight the main part of landscape paintings, the SA-ResBlock is constructed by combining shuffle attention [23] and residual block. By assigning feature attention weights, the main scene parts and their details can be better focused on. In order to solve poor style transfer and blurred outlines, perceptual loss and edge loss are proposed to constrain the model, which can reflect the characteristics of Chinese landscape painting (i.e., artistic conception and lines). To calculate the losses, a pre-trained VGG network is added to extract features for perceptual loss and an edge detection operator is used to

extract edge maps for edge loss. Therefore, generated landscape paintings have clearer outlines and structures, richer details, and realistic styles, which improves the generation quality of landscape paintings. In a word, the proposed Paint-CUT can generate landscape paintings with similar content to photos and a similar style to target paintings.

The rest of this paper is structured as follows: The proposed Paint-CUT is introduced in Section 2, including a general introduction, a generator based on the SA-ResBlock, and loss function. Section 3 shows the comparison and ablation experiment results and evaluations. Section 4 concludes this paper.

2. Proposed Paint-CUT Model

Based on the problems of CUT, the Paint-CUT is proposed to generate landscape paintings from photos. In this section, the proposed Paint-CUT is systematically introduced, including a general introduction of the model, a generator based on the SA-ResBlock, and the loss function.

2.1. General Introduction of Paint-CUT

Due to the unsatisfactory performance of basic CUT in generating landscape paintings, the Paint-CUT is proposed to ensure that the generated landscape paintings are consistent with the content of photos and the style of target landscape paintings. The code is available at https://github.com/haoyuelee/Paint-CUT.git, accessed on 2 February 2024.

The structure of the constructed Paint-CUT is shown in Figure 1. The main improvements are as follows: (1) Chinese landscape painting emphasizes layout and composition. In order to solve the problem of detail loss and maintain the texture details of the scenery, shuffle attention (SA) [23] is added to the residual blocks of the CUT generator. And then, the SA-ResBlock is constructed, which can better capture the main features of landscape photos and effectively extract the detailed features of photos. (2) In order to solve the problem of poor style transfer and ensure a generated landscape painting with a similar style to the target landscape painting, the pre-trained VGG is used to extract the content and style features to calculate the perceptual loss. (3) In order to solve the problem of blurred outlines, the Canny edge detection is used to generate edge maps. And then, the edge loss is calculated to constrain the model to generate landscape paintings with clear outlines and modeling. Finally, the landscape paintings generated by Paint-CUT have clear outlines and structure, rich details, and realistic style.

Specifically, the Paint-CUT consists of a generator G and a discriminator D, which generate landscape paintings from photos through unsupervised training. The model structure is shown in Figure 1. The input landscape photo is in Domain X and the landscape painting is in Domain Y. The generator G is used to generate landscape paintings $\hat{y}$ from photos x, i.e., $\hat{y} = G(x)$. The discriminator D is used to distinguish between the generated landscape paintings $\hat{y}$ and real landscape paintings. Both the generator G and discriminator D need to be trained simultaneously to convert the images in Domain X to Domain Y.

The generator consists of an encoder and a decoder. When images in Domain X are input, the convolutional encoder extracts deep features. And the SA-ResBlocks are used to extend the encoder structure and further accelerate model convergence. Then, the decoder generates fake images in Domain Y. Specifically, the landscape photo x is input into the encoder and is encoded as a feature map through three downsampling convolution blocks. The feature map contains the content and structure features of photos. Since the SA can use spatial and channel attention in parallel and combine the two types of attention, the SA is introduced into the residual blocks of the CUT generator to construct the SA-ResBlock. Then, nine SA-ResBlocks are adopted to better extract the main scene information and deeper features in landscape photos. Finally, the feature map with more complex features is input into two upsampling transpose convolution blocks and a convolution block to generate the landscape painting $\hat{y}$. The discriminator uses the 70×70 PatchGAN, which crops the image into 70×70 patches. For each patch, it predicts the true or false values and, finally, takes the average of all patch predictions to obtain the prediction value for the

entire image. In this way, the generated landscape painting is judged as true or false. The details of each module are described below.

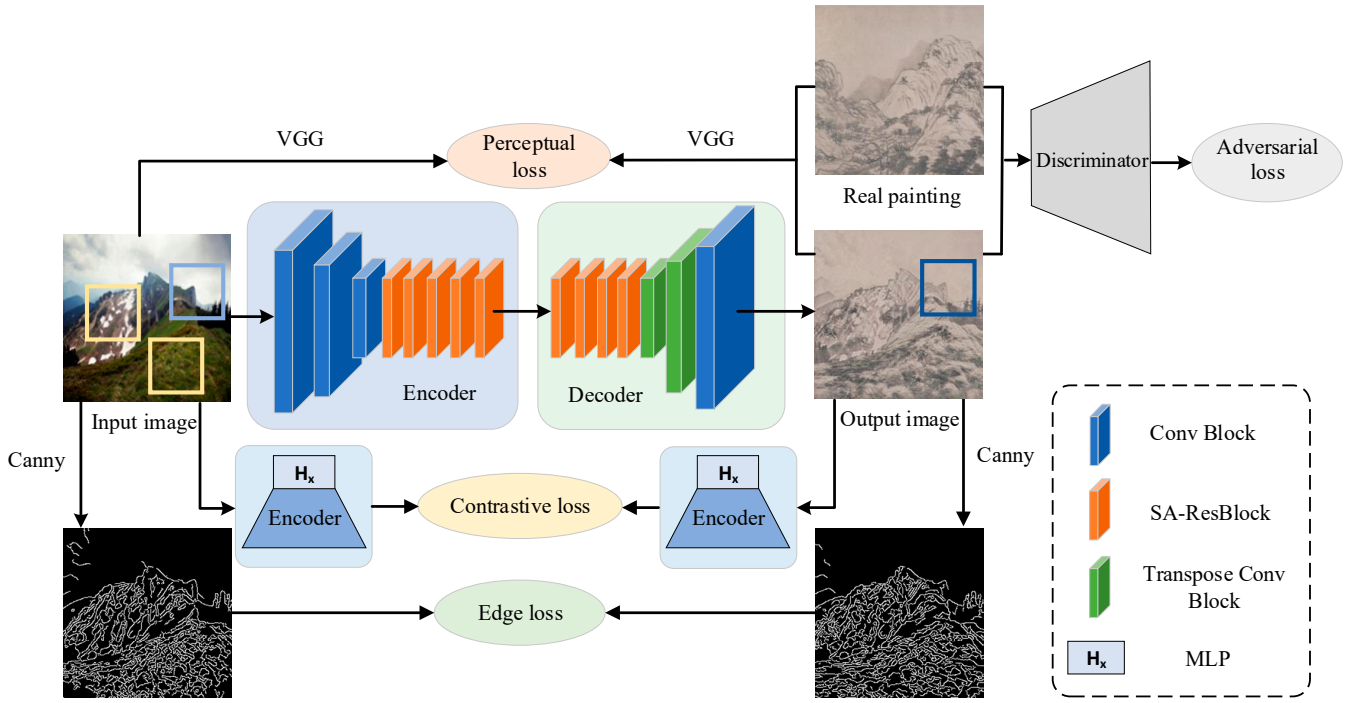


Figure 1. Constructed Paint-CUT structure (The blue box in the output image denotes query patch, the blue box in the input image denotes positive patch, and the yellow boxes in the input image denote negative patches).

2.2. Generator Based on the SA-ResBlock

Chinese landscape painting emphasizes modeling and composition and mainly describes the natural landscape. In order to solve the detail loss and highlight the scene information and texture characteristics, the shuffle attention residual block (SA-ResBlock) is constructed, which can better capture the main features of landscape photos and effectively extract the detailed features of photos. The scenery details can be generated by introducing the proposed SA-ResBlock, which reflects the improvement in terms of considering the scene features of Chinese landscape paintings.

At present, attention mechanisms can be mainly divided into two categories: spatial attention and channel attention, which capture pixel-level relationships and channel dependencies, respectively. The combination of them can perform better but will increase the computational costs. Shuffle attention (SA) [23] can efficiently combine spatial attention and channel attention to communicate information between different sub-features. The SA module is shown in Figure 2. The first step is feature grouping; the input features are divided into G groups along the channel dimension and, then, the sub-features of each group are processed in parallel. In the second step, the features of each group are evenly divided into two branches along the channel dimension. The attention weights are computed from the channel and spatial dimensions, respectively, and, then, the two branches within each group are concatenated. After that, all the sub-features are aggregated. Finally, the channel shuffle is used to recombine sub-features between different groups by channel dimension, which can promote the exchange of information between different channels. So, the diversity and expressiveness of the features can be enhanced.

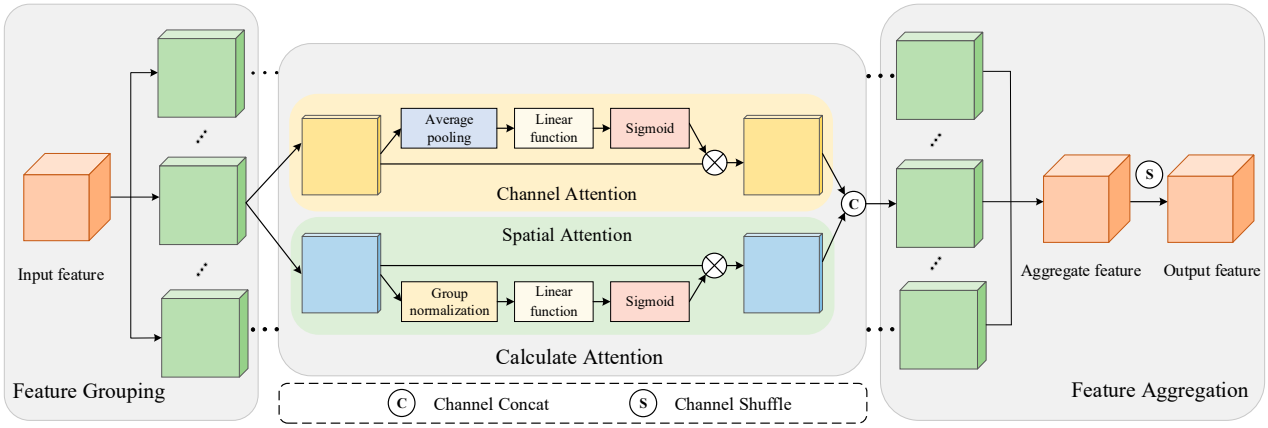


Figure 2. An overview of shuffle attention.

Chinese landscape paintings contain various scene information. The generated landscape paintings often lose details of scenes, such as rock texture. The residual blocks in the CUT generator can reduce the loss of features. Meanwhile, the SA can use spatial and channel attention in parallel and combine the two types of attention, which can be introduced to focus on relevant scene information. Therefore, the SA is introduced into the residual blocks in the CUT generator to construct the SA-ResBlock, which is used to enhance the generator's ability to extract the main scene information and texture features. Specifically, the generator of Paint-CUT consists of an encoder and a decoder, which is shown in Figure 3. The encoder consists of three convolution blocks and five SA-ResBlocks and the decoder consists of four SA-ResBlocks, two transpose convolution blocks, and one convolution block. The landscape photos are input into the generator and the content and structural features of photos are extracted by downsampling based on convolution blocks in the encoder. Then, the extracted features are input into the SA-ResBlock to further combine different features of landscape photos. Since the residual block can reduce the vanishing gradient, after adding SA, the features can be filtered purposely to reuse the useful features. After that, the features are input into the SA-ResBlocks in the decoder and the detail features are recovered and enhanced, which solves the problem of detail loss. Finally, the output features are upsampled by using transpose convolution blocks and input into a convolution block to generate the landscape paintings. After the above operations, the generated landscape paintings can better maintain the content and structure features of the input landscape photo and have rich details.

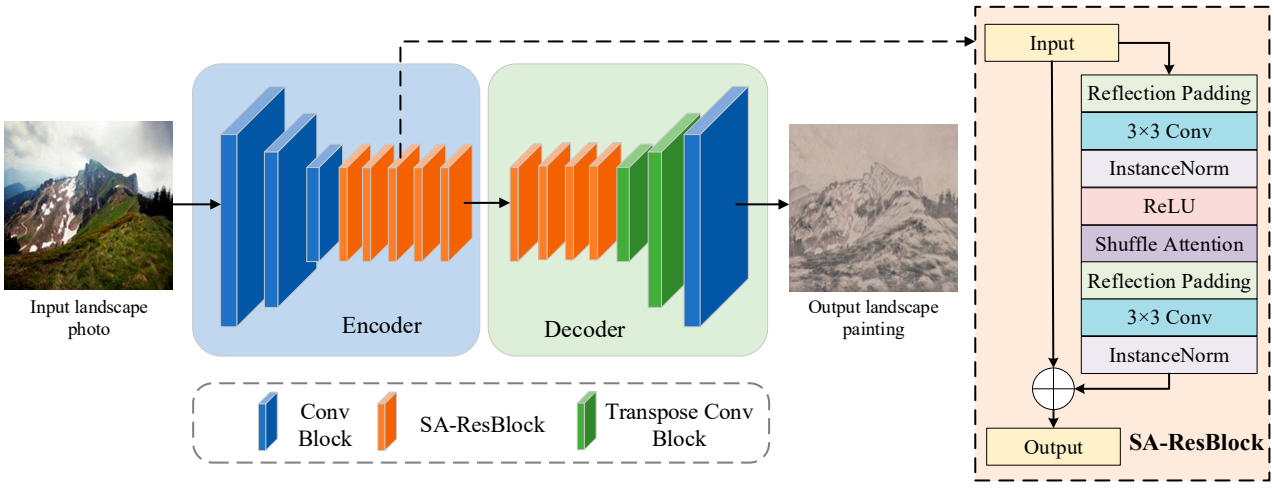


Figure 3. Generator of Paint-CUT based on the SA-ResBlock.

2.3. Loss Function

The loss function can constrain the model and choosing an appropriate loss function is crucial for model training. Chinese landscape painting mainly uses ink techniques and has its own unique style features. Additionally, lines are used to draw the outlines and details of scenery. However, the landscape paintings generated by existing methods have the problems of poor style transfer, detail loss, and blurred outlines. Other than the adversarial loss, contrastive loss, and identity loss of CUT, perceptual loss and edge loss are introduced to Paint-CUT to constrain the model. These losses will be described specifically as follows.

2.3.1. Adversarial Loss

The landscape painting in Domain Y is generated from a photo in Domain X. The adversarial loss is used to ensure that the generated landscape paintings are similar to the style of the target paintings and have the same content as the landscape photos. The adversarial loss of the generator G and the discriminator D is defined as follows:

$$L_{GAN}(G, D, X, Y) = E_{y\sim Y}[\log D(y)] + E_{x\sim X}[\log(1 - D(G(x)))] \tag{1}$$

where $E_{x\sim X}[f(x)]$ represents the expectation of $f(x)$ about X when the random variable x satisfies the probability distribution of X. $E_{y\sim Y}[f(y)]$ represents the expectation of $f(y)$ about Y when the random variable y satisfies the probability distribution of Y.

2.3.2. Contrastive Loss

The idea of contrastive learning is introduced, which aims to maximize mutual information between input and output image patches by constructing positive and negative samples. The contrastive loss is calculated by the Noise Contrastive Estimation (NCE) loss, which makes the query close to the positive and stays away from the negatives. The NCE loss is defined as follows:

$$\ell\left(z, z^{+}, z^{-}\right) = -\log\left[\frac{\exp(z \cdot z^{+}/\tau)}{\exp(z \cdot z^{+}/\tau) + \sum_{n=1}^{N} \exp\left(z \cdot z_{n}^{-}/\tau\right)}\right] \tag{2}$$

where z, z^{+}, and z^{-} are the K-dimensional vectors of the query sample, positive sample, and negative samples, respectively. N is the number of negative samples and τ is the proportional hyperparameter.

The goal of contrastive loss is to retain the features of landscape photos in Domain X maximally. After the image is divided into patches, the feature maps are extracted from L layers of the generator encoder to obtain the feature vectors. In addition, a two-layer MLP network is added after each layer to further extract features in different dimensions. Finally, a stack of features $\{z_l\}_L$ is produced. The contrastive loss is calculated by summing the NCE loss of each layer feature, which is defined as follows:

$$L_{PatchNCE}(G, H, X) = E_{x\sim X} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell\left(\hat{z}_l^s, z_l^s, z_l^{S\setminus s}\right) \tag{3}$$

where H is a two-layer MLP network and L is the number of feature map layers. S_l is the number of spatial locations in l layer for each spatial location $s \in \{1, 2, \ldots, S_l\}$. $\hat{z}_l^s$ is the generated landscape painting features, z_l^s is the corresponding features of landscape photos, and $z_l^{S\setminus s}$ is the noncorresponding features of landscape photos.

2.3.3. Identity Loss

When generating images from Domain X to Domain Y, for each input image y, the identity loss is used to ensure that the generated image is consistent with the original

y. For instance, after inputting a landscape painting, the generated result should also be a landscape painting. By constructing positive and negative samples, identity loss still uses the idea of contrastive learning to maximize the mutual information between the input and output image patches. The identity loss is defined as follows:

$$L_{PatchNCE}(G, H, Y) = E_{y \sim Y} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell\left(\hat{z}_l^s, z_l^s, z_l^{S \setminus s}\right) \tag{4}$$

where H is a two-layer MLP network and L is the number of feature map layers. S_l is the number of spatial locations in l layer for each spatial location $s \in \{1, 2, \ldots, S_l\}$. $\hat{z}_l^s$ is the generated landscape painting features, z_l^s is the corresponding features of landscape paintings, and $z_l^{S \setminus s}$ is the noncorresponding features of landscape paintings.

2.3.4. Perceptual Loss

Chinese landscape painting focuses on the use of ink techniques and has its unique style characteristics. Other than the above CUT loss functions, perceptual loss is introduced to solve the problem of poor style transfer, retain the content of the landscape photo, and generate a similar style to the target landscape painting. Landscape photos contain rich information about scenery and the target landscape paintings contain specific style features. The main idea of perceptual loss is to quantify the similarity between the generated image and the target image by calculating the feature differences between them. Specifically, the layers of the pre-trained convolutional neural network are used to map the generated paintings and landscape photos and target paintings to the feature space, respectively. These features have better semantic information and, then, the similarity of the style and content is measured by comparing the difference between the feature maps. The perceptual loss is added so that the generated landscape paintings have rich details and realistic style characteristics. The pre-trained VGG16 network is used to extract the feature maps of landscape photos, target landscape paintings, and generated landscape paintings. And then, the style loss ℓ_{style} and content loss $\ell_{content}$ are calculated, respectively. The formulas are defined as follows:

$$\ell_{style} = \sum_j \frac{1}{C_j H_j W_j} \|\phi_j(y) - \phi_j(\hat{y})\|_2^2 \tag{5}$$

$$\ell_{content} = \sum_j \frac{1}{C_j H_j W_j} \|\phi_j(x) - \phi_j(\hat{y})\|_2^2 \tag{6}$$

where ϕ_j denotes the output feature map of the j layer in the VGG16 network; C_j, H_j, and W_j are the number of channels, height, and width of the feature map, respectively; x is the input landscape photo; y is the target landscape painting; and $\hat{y}$ is the generated landscape painting.

The perceptual loss is calculated by a weighted summation of style loss and content loss, which is defined as follows:

$$L_{per}(G, X, Y) = \alpha \ell_{style} + \beta \ell_{content} \tag{7}$$

where hyperparameter α is a weight that controls the style loss and hyperparameter β is a weight that controls the content loss. Better results are obtained when α and β are set to 0.6 and 0.4, respectively.

2.3.5. Edge Loss

Chinese painting focuses on artistic conception and lines, which use lines to describe the outline and layout of scenery. For a landscape painting, it should have rich details and clear outlines, i.e., the trend of mountains and the waves of water. Therefore, it is important to generate landscape paintings with clear lines and distinct scenery. The lines and details

can be generated by introducing the edge loss, which reflects the improvement in terms of considering the line features of Chinese landscape paintings.

In order to solve the problem of blurred outlines, edge loss is proposed to generate landscape paintings with clear lines and details of scenery. The edges of landscape paintings usually concentrate on much information, which is important to recognize and use to generate the scenery in paintings. The Canny edge detection [24] is used to extract the edges from input photos and generated landscape paintings. In order to generate high-quality edge maps, the image is preprocessed using Gaussian filtering and other methods to reduce the noise. And then, the edges are detected by the Canny edge detector. The edge loss is obtained by calculating the mean square error of edge images, which is defined as follows:

$$L_{edge}(G, X, Y) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[y_{edge}(i,j) - x_{edge}(i,j)\right]^2 \tag{8}$$

where m and n are the length and width of the image, respectively, $y_{edge}(i,j)$ is the pixel value of edge maps of generated landscape paintings at (i,j), and $x_{edge}(i,j)$ is the pixel value of edge maps of input landscape photos at (i,j).

2.3.6. Total Loss of Paint-CUT

Combining all the above losses, the total loss of Paint-CUT is defined as follows:

$$\begin{aligned} L_{total}(G, D, X, Y) \quad &= L_{GAN}(G, D, X, Y) + \lambda_x L_{PatchNCE}(G, H, X) \\ &+\lambda_y L_{PatchNCE}(G, H, Y) + L_{per}(G, X, Y) + L_{edge}(G, X, Y) \end{aligned} \tag{9}$$

where λ_x and λ_y are hyperparameters that constrain the contrastive loss and identity loss. Better results are obtained when λ_x and λ_y are set to 1.

3. Experiment and Result Analysis

In this section, qualitative and quantitative analysis of generated landscape paintings will be conducted. The effectiveness of the proposed Paint-CUT will be demonstrated by analyzing the generated landscape paintings and evaluation metrics. We compare the generated results of Paint-CUT with MUNIT [25], NICE-GAN [26], U-GAT-IT [27], CycleGAN [28], and the basic CUT [22]. Evaluation metrics are calculated for quantitative analysis. Furthermore, ablation experiments will be conducted to compare the effects of each module added in the proposed Paint-CUT and quantitative analysis will be carried out. The specific details of the experiment will be presented below.

3.1. Construction of Dataset

In order to better study the intelligent generation of Chinese landscape paintings, we train and test the proposed Paint-CUT on the constructed landscape painting dataset. This dataset includes landscape photo samples, as well as two of the top ten painting samples, i.e., the samples of Dwelling in the Fuchun Mountains and A Thousand Li of Rivers and Mountains. In order to generate high-quality results, a series of processing methods are applied to the samples. The construction process of landscape painting samples, using A Thousand Li of Rivers and Mountains as an example, is shown in Figure 4. Landscape photos cover various elements, such as mountains, water, trees, rocks, and houses, which are common and represent elements of Chinese landscape paintings. These photos are consistent with the Chinese landscape paintings in content, which can be used to generate landscape paintings. The size of the sample is 256 × 256, including 5388 samples of Dwelling in the Fuchun Mountains, 5025 samples of A Thousand Li of Rivers and Mountains, and 3253 landscape photos. In addition, the entire dataset is randomly divided into a training dataset (80%) and a test dataset (20%). A detailed overview is shown in Table 1. This dataset can satisfy the requirements of experiments; examples of samples are shown in Figure 5.

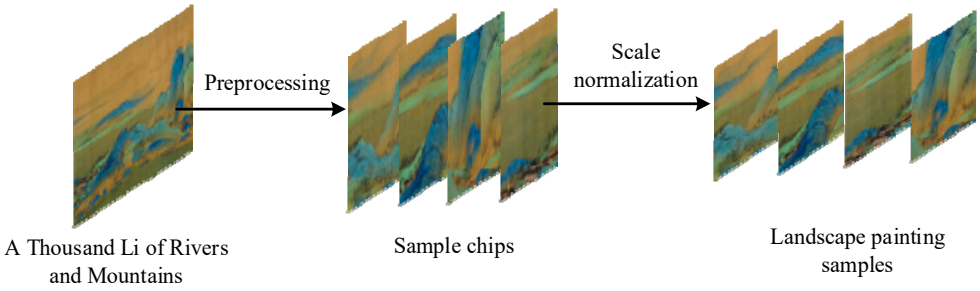


Figure 4. Construction process of landscape painting samples.

Table 1. Number of landscape paintingdataset.

Sample Types	Training Samples	Test Samples	Total
Dwelling in the Fuchun Mountains	4310	1078	5388
A Thousand Li of Rivers and Mountains	4020	1005	5025
Landscape photo	2602	651	3253
Total	10,932	2734	13,666

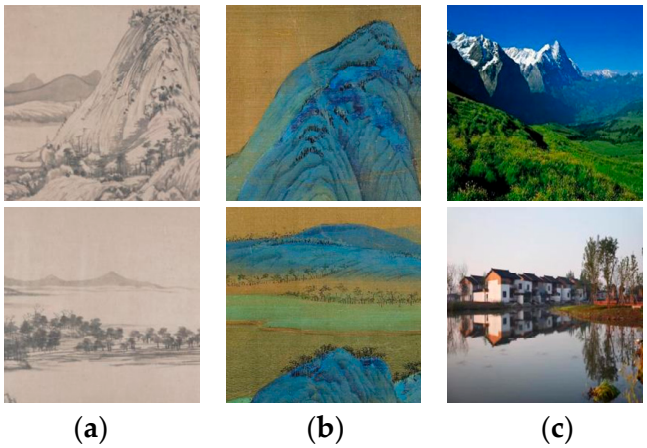


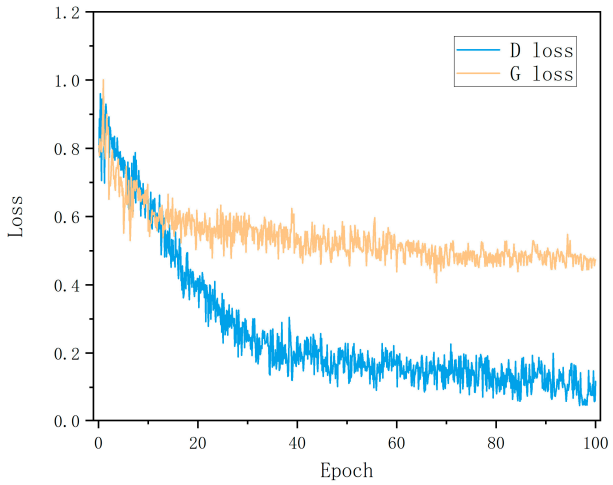
Figure 5. Samples of the constructed dataset ((**a**) is the samples of Dwelling in the Fuchun Mountains, (**b**) is the samples of A Thousand Li of Rivers and Mountains, and (**c**) is the landscape photo samples).

Meanwhile, in order to prove that the proposed Paint-CUT has good generalization ability, the ChipPhi dataset [16] is also used to train and test, which contains landscape photos and Chinese paintings, including 1630 horse photos, 912 horse paintings by Xu Beihong, 1976 landscape photos, and 1542 landscape paintings by Huang Binhong. The landscape photos of the ChipPhi dataset contain various representative elements of Chinese landscape paintings, such as mountains and rivers, which can be used to generate landscape paintings. The ChipPhi dataset is divided into a training dataset (90%) and a test dataset (10%). A total of 1774 photos and 1388 paintings are used to train and 202 photos and 154 paintings are used to test.

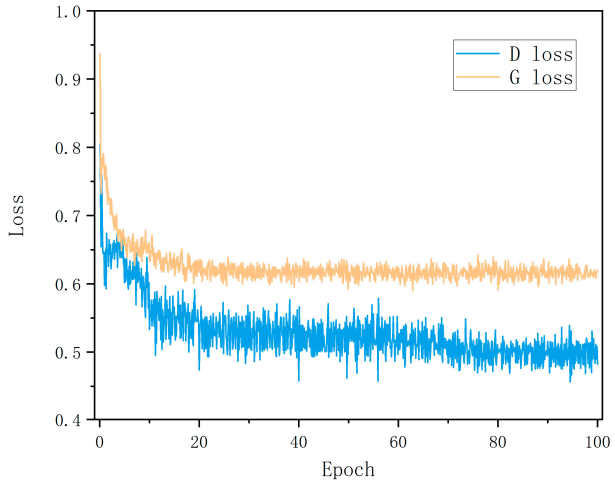
3.2. Training Process

The experimental environment is listed as follows: Linux system 4.18.0, Pytorch framework 1.8.0, and NVIDIA GeForce RTX 3080Ti 12GB GPU, which is manufactured by NVIDIA in Santa Clara, CA, USA. Hyperparameter tuning is the process of selecting the optimal hyperparameters for the model. Hyperparameters are tuned to generate better results in different datasets. During the model training, hyperparameters are set as follows: the size of all input samples is 256 $\times$ 256, the batch size is 1, and the number of Epochs is 100. The Adam optimization algorithm is used to update the weights. The constructed landscape painting dataset and ChipPhi dataset are selected to train models, respectively.

The initial learning rate is set to 0.0001, 0.00015, and 0.0001 for the samples of Dwelling in the Fuchun Mountains, the samples of A Thousand Li of Rivers and Mountains, and the ChipPhi dataset, respectively. However, generative models are difficult to train and challenges, such as model collapse and vanishing gradient, may be encountered during training. In order to demonstrate the dynamic changes in the model training, the loss curves with Epochs of the generator and discriminator on both datasets are shown in Figure 6. The model has converged when the curve has stabilized. Taking the training process of samples (i.e., Dwelling in the Fuchun Mountains) shown in Figure 6a as an example, we can observe the loss curve during training. When the Epoch is small, the generator loss (G loss) function oscillates violently and decreases rapidly. As the Epoch continues to increase, the amplitude of the generator loss function gradually decreases and eventually stabilizes. When the Epoch is small, the discriminator loss (D loss) function also decreases rapidly. As the Epoch continues to increase, the discriminator loss function still tends to stabilize. The loss curve of the ChipPhi dataset shown in Figure 6b also follows a similar trend. In conclusion, with the increase in training Epochs, the proposed Paint-CUT can rapidly tend to a stable state.



(**a**)



(**b**)

Figure 6. Curves of loss function with Epochs during the training process. (**a**) Samples of Dwelling in the Fuchun Mountains. (**b**) Samples of ChipPhi dataset.

3.3. Experimental Results

The proposed Paint-CUT is trained and tested on the constructed landscape painting dataset and ChipPhi dataset, including comparison experiments and ablation experiments.

The IS and FID of the model results are used to evaluate model performance. The specific experiments are as follows.

3.3.1. Comparison Experiments

(1) Qualitative Analysis

In order to verify the effectiveness of the proposed Paint-CUT in generating landscape paintings, MUNIT [25], NICE-GAN [26], U-GAT-IT [27], CycleGAN [28], and CUT [22] are selected to compare the generated results with the proposed Paint-CUT.

MUNIT is a multimodal unsupervised image-to-image translation model, which samples from both a content space and style space and reconstructs them to generate final results. NICE-GAN contends a new role of the discriminator by reusing it to encode the images of the target domain. U-GAT-IT incorporates a new attention module and a learnable normalization parameter. The attention module guides the model to focus on important regions, distinguishing between source and target domains. In addition, AdaIN is introduced to control shape and texture changes. CycleGAN is a generative adversarial network for image translation with unpaired data, which consists of two generators and two discriminators in a ring network. In this paper, we compare the above models with the proposed Paint-CUT on the constructed dataset and ChipPhi dataset, respectively, and analyze their generated results.

Firstly, experiments are conducted on the samples of Dwelling in the Fuchun Mountains in the constructed dataset, which generates landscape paintings with the style of Dwelling in the Fuchun Mountains from landscape photos. In total, 1000 landscape photos and 1000 landscape paintings with the style of Dwelling in the Fuchun Mountains are selected as training samples from the constructed dataset. The learning rate is set to 0.0001. Another 100 photo samples are selected for testing. The comparison experiment results of generated landscape paintings are shown in Figure 7. The samples of Dwelling in the Fuchun Mountains have elegant ink with an appropriate layout of mountains and water. As shown in Figure 7b, the generated results of MUNIT have a similar ink wash style to Dwelling in the Fuchun Mountains. However, the content is quite different from the input landscape photo, which only has a general outline and loses specific texture features. MUNIT samples from both a content space and style space and reconstructs them to obtain the final generated results. Due to the complex structural features and style features of landscape paintings, the generated results of MUNIT are not ideal. The generated paintings of NICE-GAN are shown in Figure 7c. Although there is some improvement in content and style compared with MUNIT, some generated results still have incomplete content and generated errors, such as a large blank area in the mountains of the fourth row in Figure 7c. The generated results of U-GAT-IT, which incorporate attention and maintain the basic structure and detailed information of the generated landscape paintings, are shown in Figure 7d. However, for some photos with indistinct boundaries in Figure 7d, the generated landscape paintings by U-GAT-IT have obvious missing parts and the details are not recovered well. The generated results of CycleGAN, which uses cycle consistency loss to ensure the accuracy of translation results, are shown in Figure 7e. The landscape paintings generated by CycleGAN are similar to the style of Dwelling in the Fuchun Mountains; however, some details are still missing in the landscape paintings. The generated results of CUT, which uses contrastive loss to replace cycle consistency loss and recovers the main content of the landscape paintings, are shown in Figure 7f. However, the detailed information is still missing and the style is not similar enough to the Dwelling in the Fuchun Mountains. The generated results of the proposed Paint-CUT are shown in Figure 7g. The SA is used to construct the SA-ResBlock in the Paint-CUT, which enables the model to better capture the internal structural features and effectively retain the texture details of the landscape photos. At the same time, perceptual loss and edge loss are added to learn the style features and modeling of landscape paintings. From the visual results, the generated landscape paintings of ours are similar to the original photos in terms of content, such as

the outline of mountains and the texture of rocks. And the results have rich details and the realistic style of Dwelling in the Fuchun Mountains.

(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**)

Figure 7. Comparison results of generating landscape paintings with the style of Dwelling in the Fuchun Mountains (Red boxes are used for comparison of details). (**a**) Landscape photo, (**b**) MUNIT, (**c**) NICE-GAN, (**d**) U-GAT-IT, (**e**) CycleGAN, (**f**) CUT, and (**g**) Ours.

In conclusion, the proposed Paint-CUT solves the problem of detail loss, poor style transfer, and blurred outlines in present generated results and can generate landscape paintings with the style of Dwelling in the Fuchun Mountains from photos. The comparison results in Figure 7 show that our model generates better results compared with others. The generated landscape paintings not only maintain the layout and content of the target photos but also reflect the characteristics of Dwelling in the Fuchun Mountains with ink wash style, which has better-generated results.

Secondly, experiments are conducted on the samples of A Thousand Li of Rivers and Mountains in the constructed dataset, which generates landscape paintings with the style of A Thousand Li of Rivers and Mountains from landscape photos. A total of 500 landscape photos and 500 landscape paintings with the style of A Thousand Li of Rivers and Mountains are selected as training samples from the constructed dataset. The learning rate is set to 0.00015. Another 100 photo samples are selected for testing. The comparison experiment results of generated landscape paintings are shown in Figure 8. The samples of A Thousand Li of Rivers and Mountains are meticulous with the style of blue and green, which describes the beauty of the southern scenery. The generated results of MUNIT are shown in Figure 8b. The content of the paintings is quite different from the landscape photos and fails to recover the main scenery in the picture. At the same time, the color is also different from the style of blue and green. Since A Thousand Li of Rivers and Mountains emphasizes delicate brush strokes and the style of blue and green, MUNIT fails to generate satisfactory results. As shown in Figure 8c, the generated results of NICE-GAN have more obvious style features than MUNIT; however, the generated landscape paintings miss some elements. The background color of the first row in Figure 8c is incorrect and it cannot generate landscape paintings with the style of blue and green well.

As shown in Figure 8d, some generated results of U-GAT-IT miss a large range of content. For example, the generated scenery in the third row of Figure 8d cannot be distinguished. The generated result loses essential structural information and local details also cannot be generated. As shown in Figure 8e, the generated results of CycleGAN are similar to the input photos in terms of content. However, the details and lines are blurred and the distribution of color is also unreasonable. The generated results of CUT are better than the above models, which are shown in Figure 8f. However, the details are not clear enough and the style is not apparent. The generated results of the proposed Paint-CUT are shown in Figure 8g. It can be seen that the proposed Paint-CUT can generate landscape paintings with the style of A Thousand Li of Rivers and Mountains from photos. From the visual results, the generated landscape paintings of ours are similar to the original photos in terms of content. As shown in the second row in Figure 8g, the details of the trees in the original photo are maintained in the generated result. And the generated landscape paintings not only maintain the layout and content of the target photos but also reflect the style characteristics of blue and green, which have better-generated results. In conclusion, the proposed Paint-CUT solves the problem of detail loss, poor style transfer, and blurred outlines in generated landscape paintings and can generate landscape paintings with the style of A Thousand Li of Rivers and Mountains from photos. The comparison results in Figure 8 show that our model generates better results. The generated landscape paintings are more similar to the original photos in content and target paintings in style.

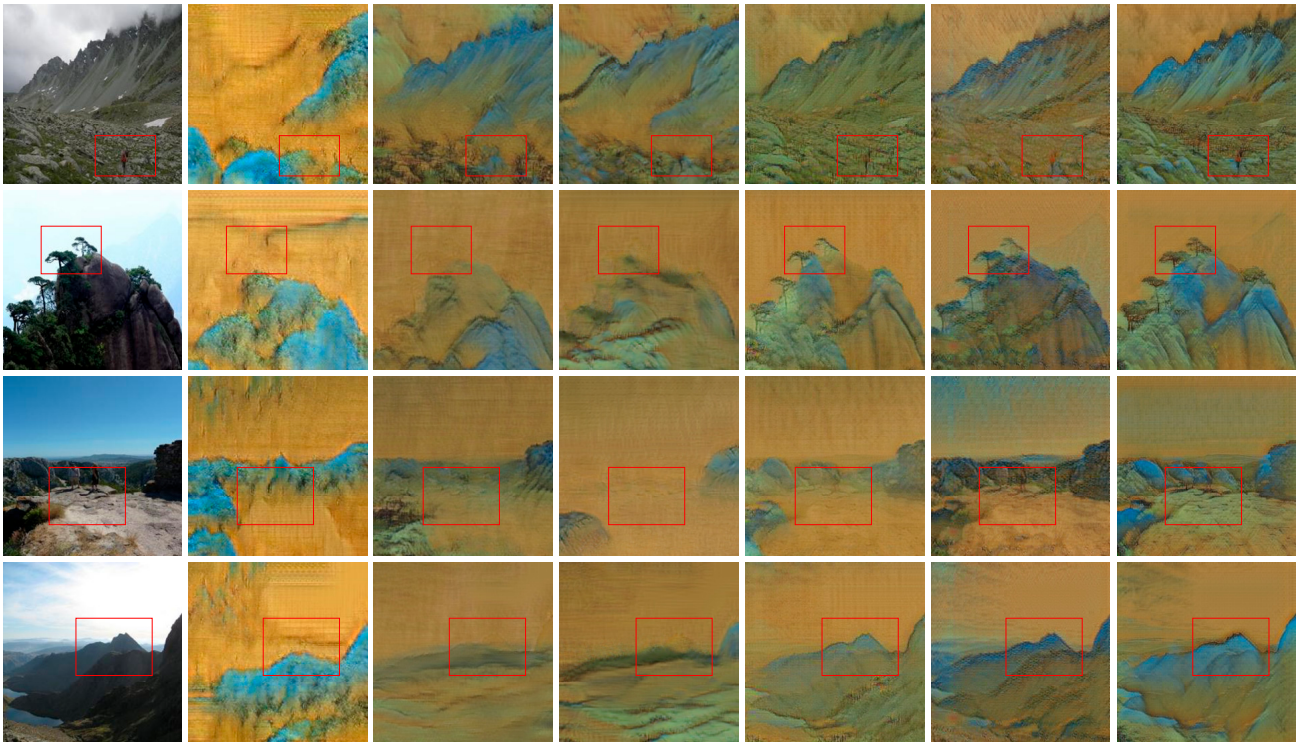


Figure 8. Comparison results of generating landscape paintings with the style of A Thousand Li of Rivers and Mountains (Red boxes are used for comparison of details). (**a**) Landscape photo, (**b**) MUNIT, (**c**) NICE-GAN, (**d**) U-GAT-IT, (**e**) CycleGAN, (**f**) CUT, and (**g**) Ours.

Finally, experiments are conducted on the samples of the ChipPhi dataset to demonstrate that the proposed Paint-CUT has good generalization ability. In total, 1000 landscape photos and 1000 landscape paintings are selected as training samples from the ChipPhi dataset. The learning rate is set to 0.0001. Another 100 photo samples are selected for testing. The comparison experiment results of generated landscape paintings are shown in Figure 9. The scene in the landscape paintings of the ChipPhi dataset is described with an

ink brush. As shown in Figure 9b, the generated results of MUNIT are similar to the ink wash style of ChipPhi. However, the content of the landscape paintings is deformed and is different from the landscape photos. This suggests that MUNIT cannot generate good results. As shown in Figure 9c, the generated results of NICE-GAN basically acquire the ink wash style of ChipPhi; however, the detailed information is not recovered well. Part of the scenery is generated incorrectly due to inconsistent recognition. The generated results of U-GAT-IT are more similar to the ink wash style of ChipPhi, as shown in Figure 9d. However, the outline of generated landscape paintings is unclear, and some areas are generated incorrectly. The generated results of CycleGAN are shown in Figure 9e; it can be seen that the results of CycleGAN are consistent with the content of the photos and, basically, have an ink wash style. However, the results still lose details (i.e., the mountains in the fourth row). As shown in Figure 9f, the generated results of CUT are similar to the input photos in terms of content and have basic scene information. But there are still some problems, such as simple texture, blurred outlines (i.e., distant mountains), and inconsistent overall color. Figure 9g shows the generated results of the proposed Paint-CUT; it can be seen that Paint-CUT can generate landscape paintings with the style of the ChipPhi dataset from photos. From the visual results, the generated landscape paintings of ours are similar to the original photos in terms of the content, which maintains the layout and content of the original photos. Therefore, the model leads to better-generated results. In conclusion, the proposed Paint-CUT solves the problem of detail loss, poor style transfer, and blurred outlines in generated landscape paintings and can generate landscape paintings with the style of the ChipPhi dataset from photos. The comparison results in Figure 9 show that our model generates better results both in content and style.

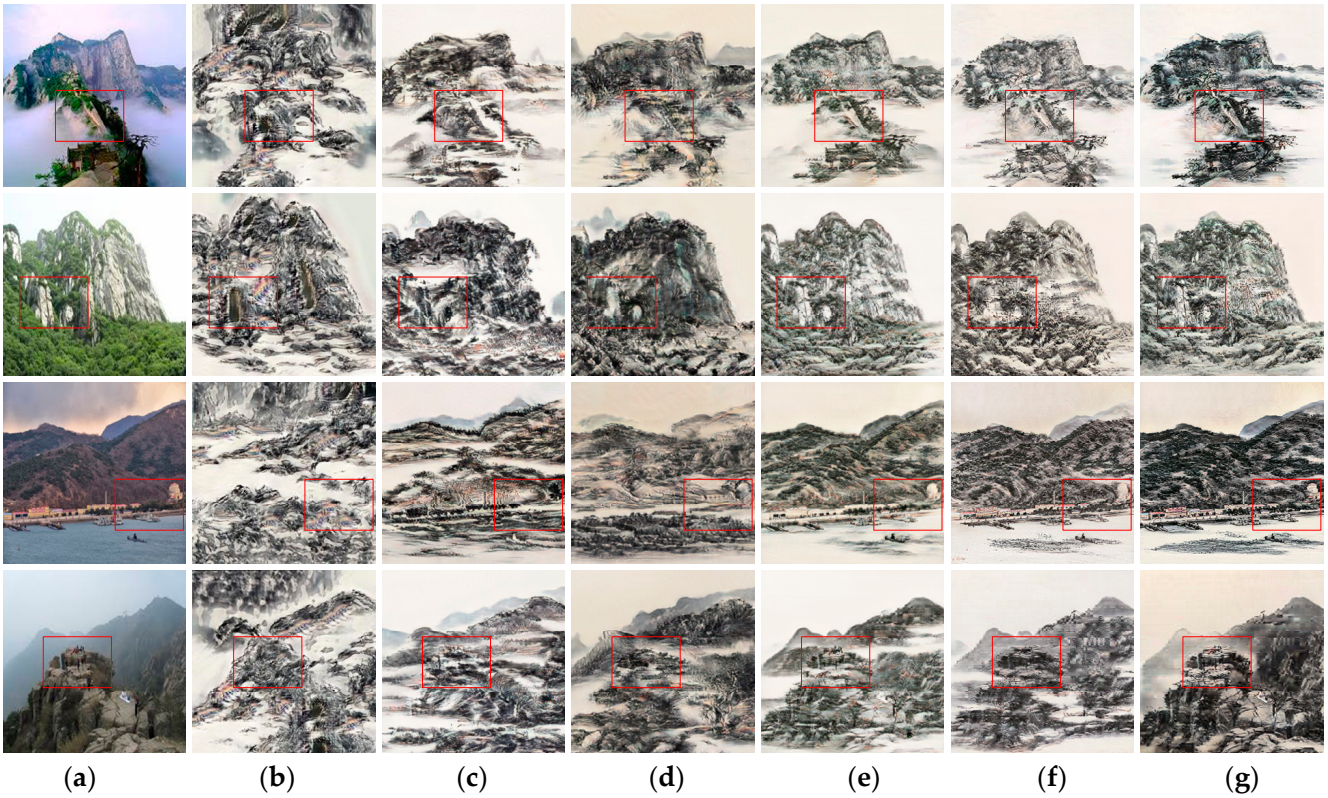


Figure 9. Comparison results of generating landscape paintings with the style of the ChipPhi dataset (Red boxes are used for comparison of details). (**a**) Landscape photo, (**b**) MUNIT, (**c**) NICE-GAN, (**d**) U-GAT-IT, (**e**) CycleGAN, (**f**) CUT, and (**g**) Ours.

In a word, the analysis of generated results shows that the proposed Paint-CUT, based on the constructed dataset and the ChipPhi dataset, solves the problems of detail

loss, poor style transfer, and blurred outlines in the present generated results. The SA is used to construct the SA-ResBlock, which enables the model to better capture the internal structural features and effectively retain the texture details of the photos. At the same time, perceptual loss and edge loss are added to learn the style features and modeling of landscape paintings. The generated landscape paintings have rich details (i.e., stone texture), clear outlines (i.e., outlines of distant mountains), and realistic style. In addition, the proposed Paint-CUT generates better landscape paintings both on the constructed dataset and the public ChipPhi dataset compared with other models, which indicates that our model has good generalizability.

(2) Quantitative analysis

Inception score (IS) [29] and Fréchet Inception Distance (FID) [30] are two evaluation metrics for generative models to measure the quality and diversity of generated results. IS calculates the KL-Divergence between the probability distribution of generated images and the real images. A high IS indicates that the generated results have higher quality and diversity. The IS is defined as follows:

$$\mathrm{IS}(G) = \exp\left(\mathbb{E}_{\mathrm{x}\sim p_g} D_{KL}(p(\mathrm{y}|\mathrm{x}) \parallel p(\mathrm{y}))\right) \tag{10}$$

where $\mathbb{E}$ is expectation, p_g is a distribution encoded by generative model G, $\mathrm{x} \sim p_g$ indicates that x is an image sampled from p_g, $D_{KL}(p \parallel q)$ is the KL-Divergence between distributions p and q, $p(\mathrm{y}|\mathrm{x})$ is the conditional class distribution denoting the probability that image x belongs to class y, and $p(\mathrm{y}) = \int_x p(\mathrm{y}|\mathrm{x}) p_g(\mathrm{x})$ denotes the marginal distribution of class y.

FID calculates the Wasserstein-2 distance between original images and generated images in a feature space. A low score of FID indicates that the results have better quality. The FID is defined as follows:

$$\mathrm{FID} = \|\mu_r - \mu_g\|^2 + Tr\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right) \tag{11}$$

where μ_r is the mean of the real image feature vector, μ_g is the mean of the generated image feature vector, Tr is the trace of the matrix, Σ_r is the covariance matrix of the real image feature vector, and Σ_g is the covariance matrix of the generated image feature vector.

The generated results of various models in Figures 7–9 are compared by calculating the IS and FID. And the comparison results are shown in Table 2, Table 3, and Table 4, respectively. The analysis of these quantitative metrics further supports the visual comparison results in Figures 7–9.

Table 2. Comparison of various models in Figure 7.

Model	IS ↑	FID ↓
MUNIT	1.145	226.136
NICE-GAN	1.559	220.375
U-GAT-IT	1.696	216.852
CycleGAN	1.710	193.157
CUT	1.739	190.513
Ours	2.357	161.584

Table 3. Comparison of various models in Figure 8.

Model	IS ↑	FID ↓
MUNIT	1.279	260.315
NICE-GAN	1.399	240.544

Table 3. *Cont.*

Model	IS ↑	FID ↓
U-GAT-IT	1.576	234.586
CycleGAN	1.590	230.647
CUT	1.743	227.163
Ours	2.469	225.458

Table 4. Comparison of various models in Figure 9.

Model	IS ↑	FID ↓
MUNIT	1.154	245.977
NICE-GAN	1.546	209.554
U-GAT-IT	1.649	198.231
CycleGAN	1.688	195.785
CUT	1.784	191.346
Ours	1.950	153.791

The evaluation metrics are evaluated to compare generated landscape paintings on samples of Dwelling in the Fuchun Mountains; the comparison results are shown in Table 2. MUNIT samples from both a content and style space and reconstructs them to generate final results. Due to the complex structure and style features of landscape paintings, the generated results are not satisfactory. MUNIT has the lowest IS and highest FID. The generated landscape paintings of NICE-GAN are similar to the ink wash style; however, the details are not recovered well. Although there are some problems, the generated landscape paintings have improvements both in style and content compared to MUNIT; thus, the NICE-GAN has a higher IS and lower FID than MUNIT. U-GAT-IT incorporates attention to maintaining the basic structure and detailed information of the generated landscape paintings, which can generate paintings with more details than NICE-GAN. Although some brush strokes are still lost, the generated landscape paintings are better than NICE-GAN in terms of content and style. The IS of U-GAT-IT is higher than NICE-GAN and the FID is lower. CycleGAN uses cycle consistency loss to ensure the accuracy of translation results. The generated landscape paintings are similar to the content of landscape photos and have an ink wash style; however, some brush strokes and details are lost. The CycleGAN has a higher IS and lower FID than U-GAT-IT. CUT is used as a baseline model, which uses contrastive loss instead of cycle consistency loss. And the CUT only uses a generator and a discriminator to achieve the image translation. The generated results of CUT are better than other comparison models both in content and style and has a higher IS and lower FID than others. However, the details of the generated landscape paintings are still blurred and the style is unclear by directly using CUT. The Paint-CUT proposed in this paper uses the SA to construct the SA-ResBlock so that the model better captures the internal structural features and effectively preserves the texture details of landscape photos. At the same time, perceptual loss and edge loss are added to learn style features and the modeling of landscape paintings. And then, it can generate landscape paintings of a better and higher quality. Therefore, the generated landscape paintings of Paint-CUT have the highest IS and lowest FID, which is consistent with the qualitative analysis. The comparison results indicate that the proposed Paint-CUT generates better results, which can retain the content of original landscape photos and recover the style of target landscape paintings.

Similarly, the IS and FID of the comparison experiments on samples of A Thousand Li of Rivers and Mountains and the ChipPhi dataset are shown in Tables 3 and 4, respectively. It can be seen that the proposed Paint-CUT has the highest IS and lowest FID. In conclusion, the proposed Paint-CUT introduces the SA to construct the SA-ResBlock and adds

perceptual loss and edge loss to generate landscape paintings. On the samples of Dwelling in the Fuchun Mountains, samples of A Thousand Li of Rivers and Mountains, and the ChipPhi dataset, our model has the highest IS and lowest FID compared to other models, which is consistent with the qualitative analysis. And it indicates that the constructed Paint-CUT has a stronger ability of feature learning and can generate landscape paintings with a reasonable layout, clear modeling, and a similar style to target landscape paintings.

3.3.2. Ablation Experiments

Chinese landscape painting focuses on artistic conception and lines, which mainly describe the natural landscape of mountains and rivers. In order to evaluate the resulting improvement by shuffle attention (SA), perceptual loss, and edge loss, ablation experiments are conducted on the proposed Paint-CUT. The SA, perceptual loss, and edge loss are sequentially added to the baseline CUT to study improvements in detail, style, and outlines. The ablation experiment results on the constructed dataset and ChipPhi dataset will be analyzed in detail in this section.

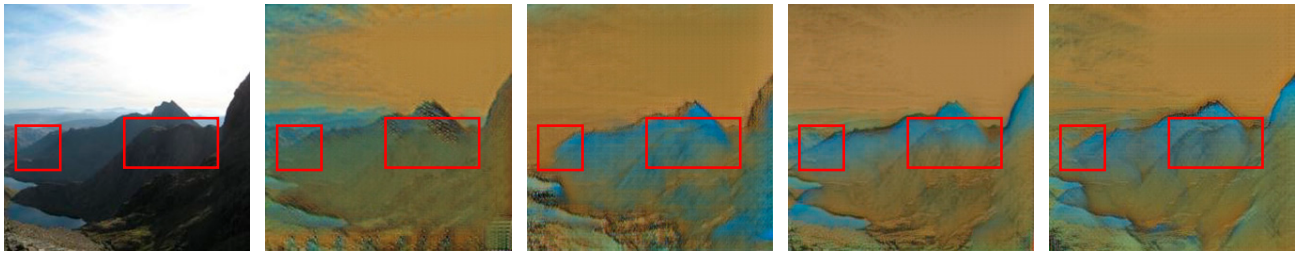
(1) Qualitative analysis

The ablation experiment results on the constructed dataset of Dwelling in the Fuchun Mountains are shown in Figure 10. The baseline CUT generates landscape paintings with the basic style of Dwelling in the Fuchun Mountains in Figure 10b; however, the details of the scenery are unclear. For example, the houses and outlines of mountains in the red box are blurred and the ink wash style is not obvious. The generated result is unsatisfactory. Shuffle attention (SA) can capture the detailed features of the main regions in photos. In order to solve the problem of detail loss and focus on the main scenery in landscape painting, SA is added to the generator to construct the SA-ResBlock. As shown in Figure 10c, the generated landscape painting has more detailed information than the result of the baseline model, i.e., the texture of mountains has a variation of ink wash and the details of houses are richer. However, these details are still blurred and cannot change according to the scenery characteristics. The perceptual loss (L_{per}) can better constrain the content and style information of generated landscape paintings. In order to solve the problem of unclear style, we continue to add perceptual loss (L_{per}) so that the generated landscape paintings are highly consistent with the content of the input photos and the details of scene information become richer. At the same time, the generated results have variations of ink wash. As shown in Figure 10d, the ink of the mountains and houses is consistent with the light and dark areas of photos. In addition, lines can guide the generation of landscape paintings. More outline information can be obtained based on the lines of scenery, such as the trend of mountains and waves of water. Meanwhile, rich line information can generate landscape paintings with more detailed features. In order to further improve the quality of landscape paintings and solve the problem of blurred outlines and details, we continue to add edge loss (L_{edge}) and then construct the proposed Paint-CUT. As the edge loss (L_{edge}) can constrain the line information of landscape paintings, the generated results have clearer outlines and richer details. As shown in the red box in Figure 10e, the details of mountains are clearer and the doors and windows of houses are visible. In conclusion, compared with the baseline model, the ablation experiments in Figure 10 prove that the proposed Paint-CUT improves the generation quality of landscape paintings. Specifically, the SA is used to construct the SA-ResBlock, which enables the model to better capture the internal structural features of the landscape photos and effectively retain the texture details. At the same time, the perceptual loss and edge loss are added so that the model can learn the style characteristics and modeling of landscape paintings. Finally, the generated landscape paintings retain the content of photos and have the similar style to Dwelling in the Fuchun Mountains.

(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 10. Generating results of ablation experiments on samples of Dwelling in the Fuchun Mountains. (**a**) Landscape photo, (**b**) CUT, (**c**) CUT + SA, (**d**) CUT + SA + L_{per}, and (**e**) CUT + SA + L_{per} + L_{edge} (Paint-CUT).

The ablation experiment results on the constructed dataset of A Thousand Li of Rivers and Mountains are shown in Figure 11. The landscape painting generated by the baseline CUT in Figure 11b basically possesses the style of blue and green; however, as shown in the red box, the outlines of the generated landscape painting are not clear and a lot of basic brush strokes are lost while the colors are also vague. SA can capture detailed features of landscape photos. As shown in the red box in Figure 11c, after adding the SA, the mountain trend becomes clearer and the texture details become richer. However, some details are still lost. The style of blue and green becomes clearer, but the colors of various scenes are still not accurate enough. The perceptual loss (L_{per}) can better constrain the content and style information of generated landscape paintings. As shown in the red box in Figure 11d, after adding the perceptual loss (L_{per}), the lost brush strokes in the mountains are recovered and the color distribution is more reasonable. However, the outlines and texture details of mountains are still unclear. The edge loss (L_{edge}) can constrain the line information of landscape paintings and the generated results have clearer outlines and richer details. The proposed Paint-CUT is finally constructed after adding the edge loss (L_{edge}). As shown in the red box in Figure 11e, the outlines of the mountains become clear and the texture of mountains is generated. In conclusion, based on the proposed Paint-CUT, the ablation experiments in Figure 11 prove that the generated landscape paintings retain the content of photos and have a similar style to A Thousand Li of Rivers and Mountains, which improves the generation quality.

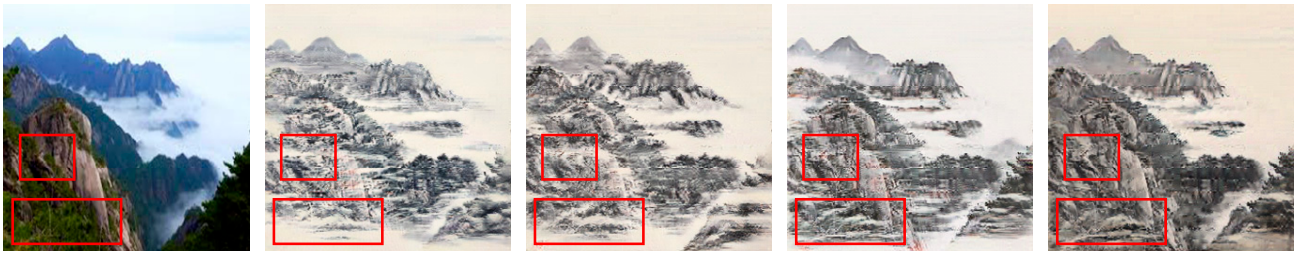


(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 11. Generating results of ablation experiments on samples of A Thousand Li of Rivers and Mountains. (**a**) Landscape photo, (**b**) CUT, (**c**) CUT + SA, (**d**) CUT + SA + L_{per}, and (**e**) CUT + SA + L_{per} + L_{edge} (Paint-CUT).

In addition, ablation experiments on the ChipPhi dataset are conducted to demonstrate that the SA, perceptual loss, and edge loss in the proposed Paint-CUT are equally necessary for generating landscape paintings on different datasets. The ablation experiment results on the ChipPhi dataset are shown in Figure 12. The generated landscape painting of CUT in Figure 12b is similar to the content of the input landscape photo. But, as shown in the red box, the generated result has some missing parts in mountains and the overall color of the painting is weak. The SA can capture detailed features of the main regions

in photos. As shown in the red box in Figure 12c, after adding the SA, the vacancy of mountains is generated and the texture details are richer; however, there are still some details missing. The perceptual loss (L_{per}) can better constrain the content and style information of the generated landscape paintings. As shown in the red box in Figure 12d, after adding the perceptual loss (L_{per}), the missing part in the mountains is recovered and the generated landscape painting has richer details. However, some textures and lines are still missing. The edge loss (L_{edge}) can constrain the line information of landscape paintings; the generated results have clearer outlines and richer details. The proposed Paint-CUT is finally constructed after adding the edge loss (L_{edge}). As shown in the red box in Figure 12e, the outlines of mountains become clear and the whole painting is highly similar to the content of the input landscape photo. In a word, based on the proposed Paint-CUT, the ablation experiments in Figure 12 prove that the generated landscape paintings retain the content of photos and have a similar style to the target landscape paintings, which improves the generation quality.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 12. Generating results of ablation experiments on samples of the ChipPhi dataset. (**a**) Landscape photo, (**b**) CUT, (**c**) CUT + SA, (**d**) CUT + SA + L_{per}, and (**e**) CUT + SA + L_{per} + L_{edge} (Paint-CUT).

In conclusion, the proposed Paint-CUT uses shuffle attention (SA) to construct the SA-ResBlock, which enables the model to better capture the internal structural features and effectively retain the texture details of the landscape photos. At the same time, perceptual loss and edge loss are added to enable the model to learn the style characteristics and modeling of landscape paintings. In summary, the proposed Paint-CUT can generate landscape paintings with clear outlines, rich details, and realistic style, which improves the quality of generated landscape paintings.

(2) Quantitative analysis

In order to further verify the effects of shuffle attention (SA), perceptual loss, and edge loss on generating landscape paintings, the generated results of various models in Figures 10–12 are compared by calculating the IS and FID. And the comparison results are shown in Table 5, Table 6, and Table 7, respectively. A higher IS indicates that the generated results have higher quality and diversity. A lower FID indicates that the generated results have better quality.

Table 5. Comparison of ablation experiments in Figure 10.

Model	IS ↑	FID ↓
CUT	1.739	190.513
CUT + SA	2.176	186.810
CUT + SA + L_{per}	2.215	178.639
CUT + SA + L_{per} + L_{edge}	2.357	161.584

Table 6. Comparison of ablation experiments in Figure 11.

Model	IS ↑	FID ↓
CUT	1.743	227.163
CUT + SA	2.313	226.522
CUT + SA + L_{per}	2.396	225.907
CUT + SA + L_{per} + L_{edge}	2.469	225.458

Table 7. Comparison of ablation experiments in Figure 12.

Model	IS ↑	FID ↓
CUT	1.784	191.346
CUT + SA	1.830	186.886
CUT + SA + L_{per}	1.888	169.068
CUT + SA + L_{per} + L_{edge}	1.950	153.791

As shown in Table 5, the baseline CUT has the lowest IS and highest FID on the samples of Dwelling in the Fuchun Mountains, which indicates that the CUT cannot generate ideal landscape paintings. Qualitative analysis shown in Figure 10 suggests that the generated landscape paintings of CUT have unclear details and blurred outlines and the variation of ink is not obvious. SA can capture the detailed features of the main regions in photos. In order to highlight the details of the scenery, SA is added to the generator to construct the SA-ResBlock. The IS increases and the FID decreases, which indicates that the addition of SA is effective in generating landscape paintings. The perceptual loss (L_{per}) can better constrain the content and style information of the generated landscape paintings. In order to make the generated landscape paintings highly consistent with the content of input landscape photos, and have the variation of ink, the perceptual loss (L_{per}) is added on the basis of SA. After adding L_{per}, the IS increases while the FID decreases, which indicates that the addition of L_{per} improves the results of landscape paintings. Lines can guide the generation of landscape paintings and rich line information can generate landscape paintings with richer detailed features. The proposed Paint-CUT is finally constructed after adding the edge loss (L_{edge}). And the generated landscape paintings have the highest IS and lowest FID, which is consistent with the qualitative analysis. The experimental results show that the proposed Paint-CUT has a stronger feature learning ability and better-generated results, which can largely retain the content of landscape photos and, at the same time, learn the style of target landscape paintings.

Similarly, the evaluation metrics of the ablation experiments on samples of A Thousand Li of Rivers and Mountains and the ChipPhi dataset are shown in Tables 6 and 7, respectively. It can be seen that after adding the SA to construct the SA-ResBlock, the model better captures the internal structural features and effectively retains the texture details of the landscape photos. At the same time, the perceptual loss (L_{per}) and edge loss (L_{edge}) are added to construct the proposed Paint-CUT, which can learn the style characteristics and modeling of landscape paintings. The generated results retain the content of landscape photos and have a similar style to the target landscape paintings, which improves the generation quality. The results of our model have the highest IS and lowest FID compared to other models, which is consistent with the qualitative analysis. In conclusion, the constructed Paint-CUT can generate landscape paintings with clear outlines, rich details, and a similar style to target landscape paintings, which improves the quality of generated landscape paintings.

4. Conclusions

The generated landscape paintings of traditional CUT often have the problems of detail loss, blurred outlines, and poor style transfer. In this paper, we propose a generation model (Paint-CUT) to generate landscape paintings from input landscape photos. The key findings and contributions of Paint-CUT are introducing the SA-ResBlock, perceptual loss, and edge loss to address the above problems. Specifically, in order to solve the problem of detail loss, based on the layout and texture features of landscape paintings, the SA-ResBlock with shuffle attention is proposed to extract the features. In order to solve the problem of poor style transfer, based on the unique style of landscape paintings, the perceptual loss is introduced to constrain the model both in content and style, which guides the transfer from landscape photos to landscape paintings. In order to solve the problem of blurred outlines and generate landscape paintings with clear lines and edges, the edge loss is constructed to calculate the error between the edge map of the landscape photo and the edge map of the generated landscape painting. After the analysis of the training process, the qualitative and quantitative analysis of comparison and ablation experiments, the results prove that the landscape paintings generated by Paint-CUT are not only consistent with the content of the landscape photos but also consistent with the style of the target landscape paintings. The generated landscape paintings have clear lines and rich details, which improve the quality of landscape paintings.

The generated landscape paintings based on the proposed Paint-CUT are consistent with the content of the photos and the style of the target landscape paintings. The intelligent generation of landscape paintings can effectively preserve the artistic value of landscape paintings and improve modern people's interest in traditional culture. Popularization of traditional Chinese culture can be promoted and public awareness of cultural protection can be enhanced. Consequently, the main purpose of this paper is to further integrate the culture and technology and provide a strong technological assistance for the inheritance and development of traditional Chinese culture.

However, as one of the precious cultural heritages, Chinese landscape paintings mainly describe the characteristics of diverse landscape scenes, embody a unique style and mood, and use lines and special techniques to depict the outlines and details of scenes. The task of generating landscape paintings is more difficult than natural images due to the different scales of scenes and special painting techniques in handling different styles. The extraction ability of multi-scale information and generation quality of complex strokes still need to be improved, which are potential challenges of the proposed model. And the generated figures should present better results, such as clearer lines and a more rational layout.

In the future, we will continue to address limitations and improve the generation quality of our model according to landscape painting characteristics. Firstly, we will focus on the special painting techniques of landscape paintings and learn how to better extract multi-scale features. Secondly, evaluation metrics specifically for landscape painting will be proposed to better evaluate the generated results. Finally, software or systems about Chinese landscape painting generation can be developed for exhibition and aesthetic education to promote the integration of culture and technology.

Author Contributions: Conceptualization, Z.S.; methodology, Z.S.; validation, H.L. and X.W.; investigation, H.L.; data curation, H.L.; writing—original draft preparation, Z.S. and H.L.; writing—review and editing, Z.S., H.L. and X.W.; supervision, Z.S.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (No. 2017YFB1402102), the National Natural Science Foundation of China (No. 62377033), the Shaanxi Key Science and Technology Innovation Team Project (No. 2022TD-26), the Xi'an Science and Technology Plan Project (No. 23ZDCYJSGG0010-2022), and the Fundamental Research Funds for the Central Universities (No. GK202205036, GK202101004).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, J.; Wang, Q.; Li, S.; Zhong, Q.; Zhou, Q. Immersive traditional Chinese portrait painting: Research on style transfer and face replacement. In Proceedings of the 4th Chinese Conference on Pattern Recognition and Computer Vision, Beijing, China, 29 October–1 November 2021; pp. 192–203.
2. Wang, Y.; Li, W.; Zhu, Q. Ink wash painting style rendering with physically-based ink dispersion model. *J. Phys. Conf. Ser.* **2018**, *1004*, 012026. [CrossRef]
3. Tang, F.; Dong, W.; Meng, Y.; Mei, X.; Huang, F.; Zhang, X.; Deussen, O. Animated construction of Chinese brush paintings. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 3019–3031. [CrossRef] [PubMed]
4. Bin, Y.; Sun, J.; Bai, H. Simulation of diffusion effect based on physically modeling of paper in Chinese ink wash drawing. *J. Syst. Simul.* **2005**, *17*, 2305–2309.
5. Yeh, J.W.; Ouhyoung, M. Non-Photorealistic rendering in Chinese painting of animals. *J. Syst. Simul.* **2002**, *14*, 1220–1224.
6. Ma, X.; Tu, L.; Xu, Y. Development status of the digitization of intangible cultural heritages. *Sci. Sin. Informationis* **2019**, *49*, 121–142. [CrossRef]
7. Geng, G.H.; He, X.L.; Wang, M.L.; Li, K.; He, X.W. Research progress on key technologies of cultural heritage activation. *J. Image Graph.* **2022**, *27*, 1988–2007.
8. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
9. Li, B.; Xiong, C.; Wu, T.; Zhou, Y.; Zhang, L.; Chu, R. Neural abstract style transfer for Chinese traditional painting. In Proceedings of the 14th Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; pp. 212–227.
10. Sheng, J.; Song, C.; Wang, J.; Han, Y. Convolutional neural network style transfer towards Chinese paintings. *IEEE Access* **2019**, *7*, 163719–163728. [CrossRef]
11. Li, Z.; Lin, S.; Peng, Y. Chinese painting style transfer system based on machine learning. In Proceedings of the 2021 IEEE International Conference on Data Science and Computer Application, Dalian, China, 25–27 October 2021; pp. 38–41.
12. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
13. Xue, A. End-to-end Chinese landscape painting creation using generative adversarial networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 3862–3870.
14. Lin, D.; Wang, Y.; Xu, G.; Li, J.; Fu, K. Transform a simple sketch to a Chinese painting by a multiscale deep neural network. *Algorithms* **2018**, *11*, 4. [CrossRef]
15. Gu, Y.; Chen, Z.J.; Chen, C. Layout adjustable simulated generation method for Chinese landscape paintings based on CGAN. *Pattern Recognit. Artif. Intell.* **2019**, *32*, 844–854.
16. He, B.; Gao, F.; Ma, D.; Shi, B.; Duan, L.Y. Chipgan: A generative adversarial network for Chinese ink wash painting style transfer. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 1172–1180.
17. Bao, F.; Neumann, M.; Vu, N.T. CycleGAN-Based emotion style transfer as data augmentation for speech emotion recognition. In Proceedings of the INTERSPEECH, Graz, Austria, 15–19 September 2019; pp. 2828–2832.
18. Zhou, L.; Wang, Q.F.; Huang, K.; Lo, C.H. An interactive and generative approach for Chinese shanshui painting document. In Proceedings of the 2019 International Conference on Document Analysis and Recognition, Sydney, Australia, 22–25 September 2019; pp. 819–824.
19. Zhang, F.; Gao, H.; Lai, Y. Detail-preserving CycleGAN-AdaIN framework for image-to-ink painting translation. *IEEE Access* **2020**, *8*, 132002–132011. [CrossRef]
20. Peng, X.; Peng, S.; Hu, Q.; Peng, J.; Wang, J.; Liu, X.; Fan, J. Contour-enhanced CycleGAN framework for style transfer from scenery photos to Chinese landscape paintings. *Neural Comput. Appl.* **2022**, *34*, 18075–18096. [CrossRef]
21. He, X.; Zhu, M.; Wang, N.; Wang, X.; Gao, X. BiTGAN: Bilateral generative adversarial networks for Chinese ink wash painting style transfer. *Sci. China Inf. Sci.* **2023**, *66*, 119104. [CrossRef]
22. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.Y. Contrastive learning for unpaired image-to-image translation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 319–345.
23. Zhang, Q.-L.; Yang, Y.-B. Sa-net: Shuffle attention for deep convolutional neural networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 2235–2239. [CrossRef]
24. Lu, Y.; Duanmu, L.; Zhai, Z.; Wang, Z. Application and improvement of Canny edge-detection algorithm for exterior wall hollowing detection using infrared thermal images. *Energy Build.* **2022**, *274*, 112421. [CrossRef]
25. Huang, X.; Liu, M.Y.; Belongie, S.; Kautz, J. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 172–189.

26. Chen, R.; Huang, W.; Huang, B.; Sun, F.; Fang, B. Reusing discriminators for encoding: Towards unsupervised image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8165–8174.
27. Kim, J.; Kim, M.; Kang, H.; Lee, K. U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019; pp. 1–19.
28. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
29. Barratt, S.; Sharma, R. A note on the inception score. *arXiv* **2018**, arXiv:1801.01973.
30. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6629–6640.