



Contour-enhanced CycleGAN framework for style transfer from scenery photos to Chinese landscape paintings

Xianlin Peng¹ · Shenglin Peng² · Qiyao Hu² · Jinye Peng² · Jiaxin Wang² · Xinyu Liu³ · Jianping Fan⁴

Received: 17 September 2021 / Accepted: 11 May 2022 / Published online: 9 June 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Image style transfer based on the generative adversarial network model has become an important research field. Among these generative adversarial network models, a distinct advantage of CycleGAN is that it can transfer between multiple domains when the data is not paired. To approximate the effects of the texturing method with the characteristics of traditional Chinese painting—“Cun method”, this paper proposes an image style transfer framework to realize the transfer from scenery photos to Chinese landscape paintings. We design a contour-enhancing translation branch, which effectively guides the transfer from photos to paintings with edge detection operators computing the gradient maps. Simulation results show that this method can convert real scenery photos to Chinese landscape paintings. The Inception Score shows that contour enhancement can make the generated set performs better on sensitivity to image edges. The Kernel Inception distance and Inception-based Structural Similarity between the generated image and the “Cun method” data set shows that contour enhancement can make the generated image closer to the “Cun method” effect. Compared with Kernel Inception distance and Frechet-Inception Distance, the Inception-based Structural Similarity proposed in this paper directly focuses on similarity, the similarities between the mean features of images generated by our model, and the “Cun method” set is 97.89%, and the composite similarity metric being 0.92. The method also performs better than the MUNIT, NiceGAN, CycleGAN, and U-GAT-IT reference models under the Neural Image Assessment metric. This indicates that the introduction of the edge operator makes the generated landscape paintings more aesthetic, especially in situations where scenery photos are rich in edge information.

Keywords GAN · Artificial intelligence art · Chinese painting · Edge-enhanced · Style transfer · Cun method · ISSIM

1 Introduction

Chinese painting has a long history and distinct national form and unique style in the world art field, which are painted on special rice paper or silk with brushes, ink, and Chinese pigments. The inheritance and innovation of Chinese painting art have become essential to cultural heritage and innovation. Recently, the artificial intelligence (AI) application that transfers scenery photos to Chinese painting has become popular and exciting for many art lovers and cultural relic protection. The role of AI technology in extracting typical features that reflect the style of artistic works and imitating paintings will help train

painting skills, innovate literary works, and inherit and protect traditional art.

The translation of scenery photos into Chinese landscape paintings is essentially an image style transfer. The traditional computational digital painting simulation methods can be divided into physical modeling, artistic effect decomposition, and non-photorealistic rendering.

- (1) The physical modeling methods [1] are based on the mechanism of landscape paintings: physical methods model brushes, paper, and ink. Simulation is achieved by simulating its inherent physical characteristics and dynamic interaction behavior. These methods are not ideal due to their intrinsic complexity [2].

Extended author information available on the last page of the article

- (2) In artistic effect decomposition methods [3, 4], the normal and real continuous painting processes are explained and described with a logical and virtual discrete step. The artistic effect decomposition methods can simulate various landscape painting effects, but it is difficult to achieve the blending and gradual change between multiple products, resulting in unsatisfactory overall outcomes.
- (3) The non-photorealistic rendering methods [5] are based on image analogy, mainly pursuing creativity to generate non-real image objects, such as synthetic rock textures. However, due to the ever-changing artistic effect of landscape paintings, each effect requires different rendering algorithms, which makes it difficult to simulate.

The generative adversarial network (GAN), proposed by Goodfellow et al. [6], is an effective method to realize image style transfer and has gradually become its mainstream method. The basic structure of GAN consists of two networks, one is called generator network, and the other is called discriminator network. These two networks can be convolutional neural networks, cyclic neural networks, or self-encoders. Training GANs is an unsupervised learning process. The generator network converts the input noise vector into an image and then sends the generated image to the discriminator network for classification. The goal of the generator network is to become perfect in developing realistic images, while the purpose of the discriminator network is to become perfect in judging whether the image provided to it is fake or real. This process is accomplished through multiple iterations.

GAN is powerful and has many practical applications, such as generating high-quality images [7], generating images from text [8], converting images from one style to another [9], etc. Many widely popular GAN-based architectures have emerged. Radford et al. proposed DCGAN [10]. This is the first application of convolutional neural networks in GANs, an important research milestone. Since then, a large number of GAN architectures have been introduced based on the DCGAN architecture. Brock et al. proposed BigGAN [11], in which the most important improvement is the orthogonal regularization of the generator to generate high-fidelity images. Karras proposed StyleGAN [12]. The core is style transfer or style mixing, which can generate images of different styles. Zhang et al. proposed StackGAN [13] to generate realistic images from text descriptions. Antipov et al. proposed Age-cGAN [14] for facial aging. Choi et al. proposed StarGAN [15] and realized the image conversion through an unsupervised learning method in multiple styles.

Kim et al. proposed the U-GAT-IT [16], which incorporates a new attention module and a new learnable

normalization function in an end-to-end manner. Huang et al. proposed MUNIT [17], which is a multimodal unsupervised image-to-image translation framework. To translate an image into another domain, they recombine its content code with a random style code drawn from the style space of the target domain. Chen et al. proposed NICE-GAN [18], which contends a novel role of the discriminator by reusing it for encoding the images of the target domain.

Although those methods can be applied to scenery photos and Chinese painting translation, some unique problems still need to be solved. Due to the particularity of painting materials and the extensive use of various painting techniques, strokes, and ink, the creative process of landscape paintings is complicated. The main content of Chinese landscape paintings is mountains, rocks, and trees, and painters often express these contents by the “Cun method” [19]. This has made the “Cun method” a vital expression language in landscape painting and an art form for painters’ aesthetic experience and aesthetic expression. Chinese paintings with the “Cun method” are quite different from other kinds of images and have a unique charm and artistic conception. Figures 1 and 2 are two typical “Cun method”. These “Cun methods” are mainly presented by lines, “imitate” the rough outlines of objects and scenes in terms of morphology. Therefore, when automatically generating landscape paintings from scenery photos, strengthening the contours of various objects to imitate the “Cun method” effect is a challenging topic worthy of in-depth study.

This paper developed an unsupervised and Contour-Enhanced(CE) image style transfer framework based on CycleGAN, called CE-CycleGAN. According to the characteristics of Chinese landscape paintings, it pays attention to the lines, sets constraints, highlights the edge features to realize the style transfer from scenery photos to landscape paintings with more Chinese painting characteristics.

The main innovations of this paper are summarized as follows:

- (1) We proposed an image style transfer framework from scenery photos to Chinese landscape paintings.



Fig. 1 Hemp-fiber strokes (Pima Cun)



Fig. 2 Axe-chopped strokes (Fupi Cun)

This framework adds a contour-enhancing translation branch with the edge detection operator to enhance edge information. The edge operator can highlight better the contour of objects such as rocks and trees and imitate specific texture effects of the “Cun Method”.

- (2) This paper proposes the Inception-based Structural Similarity (ISSIM) metrics to describe the similarity between two image sets. The Fréchet-Inception Distance (FID) and Kernel Inception distance (KID) metric scores are lower only the more similar the generated image set is to the original image set and the more diverse the generated images are at the same time in only one metric. The ISSIM metric set is defined directly from similarity and can provide richer information than the FID metric for judging the quality of the generated image set. Unlike the FID and KID, the ISSIM set presents similarity and diversity separately in multiple metrics, and a single ISSIM metric can be very intuitive to reflect the similarity. Therefore, the ISSIM metric is more applicable to tasks where similarity is more important than diversity. From the analysis in the later section, it is indicated that it can be judged whether the output of the model collapses to a certain direction through the ISSIM metrics, if there is a data set in that direction.
- (3) To evaluate the artistry of the generated landscape paintings, we introduce an aesthetic evaluation metric, NIMA [20], to evaluate the performance of the method from an artistic point of view, which can effectively assess the improvement of the aesthetic performance of the generated paintings in the introduction of the edge operator.

The rest of this paper is structured as follows. Section 2 introduces the CycleGAN model and related work, Sect. 3 details the CE-CycleGAN method proposed in this paper, Sect. 4 gives experimental results and discussions, and

Sect. 5 summarizes the full text and introduces future work ideas.

2 The CycleGAN model

The two models, Pix2pix [21] and CycleGAN [9], are worthy of attention in image style transfer from scenery photos to Chinese landscape paintings. In 2016, the Pix2pix model proposed by Isola et al. [21] provides a general framework for transforming images from one style to another, that is, to map labels to photos, map edges to objects, convert night images to day images, color for black and white images, convert sketches to images, etc. However, the model must require pairwise data (paired data), but it is usually challenging to obtain training sample pairs of photos and landscape paintings in photo-to-painting conversion. In 2017, the CycleGAN model proposed by Zhu et al. [9] has good performance in geometry, color, and style transfer and can achieve some exciting image conversions, such as converting photos into paintings. The outstanding advantage of this model is that it can be trained in unpaired data. Based on CycleGAN, Zhang et al. [22] proposed a CycleGAN-AdaIN framework to convert real photos to Chinese ink paintings. In this model structure, the author uses one cycle-consistency loss to replace the two cycle-consistency losses. In addition, multi-scale structural similarity metric loss is added to the reconstruction loss to generate more detailed images.

As shown in Fig. 3, CycleGAN is essentially two mirror-symmetric GANs that form a ring network. The two GANs share two generators, and each carries a discriminator, i.e., two discriminators and two generators. One one-way GAN has two losses, and two one-way GANs have a total of four losses. In short, the model works by acquiring an input image from the domain Discriminator A, D_A . This input image is passed to the first generator, G_{AB} , transforming the given image from the domain Discriminator A to the target domain Discriminator B. This newly generated image is then passed to another generator, G_{BA} , whose task is to convert the image rec_A back to the original domain Discriminator A. Here, a comparison can be made with the autoencoder. This output image must be similar to the original input image and define meaningful mappings that did not originally exist in the unpaired dataset (Table 1).

3 The proposed CE-CycleGAN method

CycleGAN is characterized by performing unpaired image-to-image translation and is suitable for style transfer from scenery photos to landscape paintings. Landscape paintings pay attention to freehand brushwork rather than other

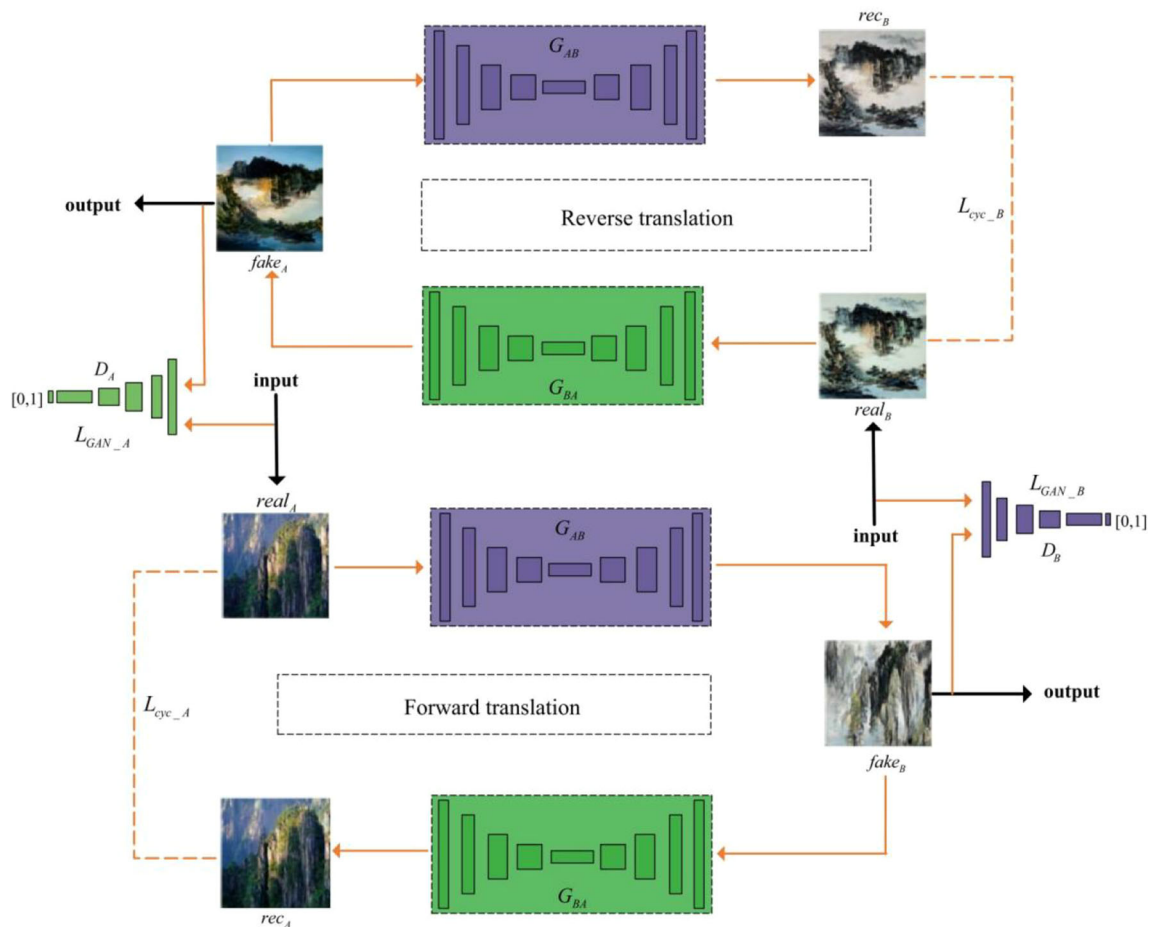


Fig. 3 Diagram of the CycleGAN network structure

details, but some contrasting edge information is still essential to the whole work. Generally, the human eye is more sensitive to solid edges than weak edges. According to this visual feature of the human eye, it is necessary to selectively enhance the edge information in the image, that is, to retain the strong edge with more considerable contrast instead of the weak edge with minor contrast. Thus, by adding a contour-enhancing translation branch to the CycleGAN network, we designed a gradient guidance method to effectively guide the style transfer from photos to paintings with gradient information.

Furthermore, to enhance the landscape painting style, the edge detection operator is used to extract the strong edge of the grayscale image. In this paper, we use the simple Sobel edge detection operator [23] to test whether the introduction of the edge detection operator is useful. The Sobel operator is a typical edge detection operator based on the first derivative. Since this operator introduces a similar local average operation, it smooths the noise and eliminates its influence. It has a good detection effect on rough edges. This feature meets the requirements of the method for the simulation effect of landscape paintings.

The whole network with a contour-enhancing translation branch is called CE-CycleGAN, and its structure is shown in Fig. 4. We define scenery photos as domains A and landscape paintings as domains B. Using the characteristics of cyclic consistency, the network is designed as forwarding translation and reverse translation. The forward translation first transfers the scenery photos into landscape paintings, namely $G_{AB} : A \rightarrow B$. Then, the landscape paintings are translated back to the original scenery photos, namely $G_{BA} : B \rightarrow A$. The reverse translation will transfer landscape paintings into scenery photos through G_{BA} and then from scenery photos to the original landscape paintings through G_{AB} . Both forward and reverse translation contain two branches: the painted translation branch and the contour-enhancing translation branch.

For the forward translation, the mapping of G_{AB} is first performed. The painted translation branch takes the real scenery photos $real_A$ as the generator G_{AB} input and the resulting image enhancement features as the input of the fusion module. The contour-enhancing translation branch takes the real image gradient map $real_{A_edge}$ as the input of the generator G_{BA} , and the generated edge-enhanced

Table 1 A list of terms for all equations and their corresponding meanings

rec_{B_edge}	Image gradient map
rec_B	Generator of contour-enhancing translation branches in the forward translation process
$G_{AB}G_{AB}$	Turning landscape photos into landscape paintings
Fuse	Integration
$fake_{A_edge}$	Landscape photo edge enhancement gradient map
$fake_A$	Landscape photo final edge enhancement map
$fake_{B_edge}$	Landscape painting edge enhancement gradient map
$fake_B$	The final edge-enhanced view of the landscape painting
G_{AB_edge}	Generator of contour-enhancing translation branches in the forward translation process
$real_{A_edge}$	Landscape photo real edge enhancement gradient map
$fake_{B_edge}$	Landscape painting edge enhancement gradient map
$L_{GAN_B_edge}$	Generative adversarial loss function for gradient advection branching
L_{GAN_B}	Generative adversarial loss function for image translation branching
D_B	Image panning branch of landscape painting discriminator
D_{B_edge}	Landscape painting discriminator with gradient panning branch
E_{real_B}	Discriminator loss function for image translation branch
$E_{real_B_edge}$	Discriminator loss function for gradient advection branching
E_{fake_B}	Generator loss function for image translation branching
$E_{fake_B_edge}$	Generator loss function for contour-enhancing translation branches
$real_B$	Landscape painting real picture
$real_{B_edge}$	Landscape painting real edge enhancement gradient map
rec_{A_edge}	Image gradient map
rec_A	Original landscape photos
G_{BA_edge}	Generator of contour-enhancing translation branches in the reverse translation process
$G_{AB} G_{AB} G_{AB}$	Converting landscape paintings to landscape photos
$fake_{B_edge}$	Landscape painting edge enhancement gradient map
$fake_{AB}$	The final edge-enhanced view of the landscape painting
$L_{cyc_A_edge}$	Cyclic consistency loss of the reduced image gradient map and the real image gradient map
L_{cyc_A}	Loss of circular consistency between restored scene photos and real scene photos
$L_{GAN_A_edge}$	Generative adversarial loss function for gradient advection branching
L_{GAN_A}	Generative adversarial loss function for image translation branching
D_A	Image panning branching for landscape photo discriminator
D_{A_edge}	Landscape photo discriminator with gradient panning branching
E_{real_A}	Discriminator loss function for image translation branch
$E_{real_A_edge}$	Discriminator loss function for gradient advection branching
E_{fake_A}	Generator loss function for image translation branching
$E_{fake_A_edge}$	Generator loss function for contour-enhancing translation branches
$real_A$	Scenic photo real picture
$real_{A_edge}$	Landscape photo real edge enhancement gradient map
$L_{cyc_B_edge}$	Cyclic consistency loss of the reduced image gradient map and the real image gradient map
L_{cyc_B}	Loss of circular consistency between restored scene photos and real scene photos
$L(G, D)$	The overall objective loss function of the model
λ	Relative weights of generative adversarial loss and cyclic consistency loss

gradient map $fake_{B_edge}$ and the real edge-enhanced gradient map $real_{B_edge}$ in the reverse translation are input into the discriminator D_{B_edge} for authenticity judgment. The

$fake_{B_edge}$ is also be spliced with the output features of the image translation branch and sent into the fusion module, which provides the guidance of the gradient information for the fusion module to generate the final edge-enhanced

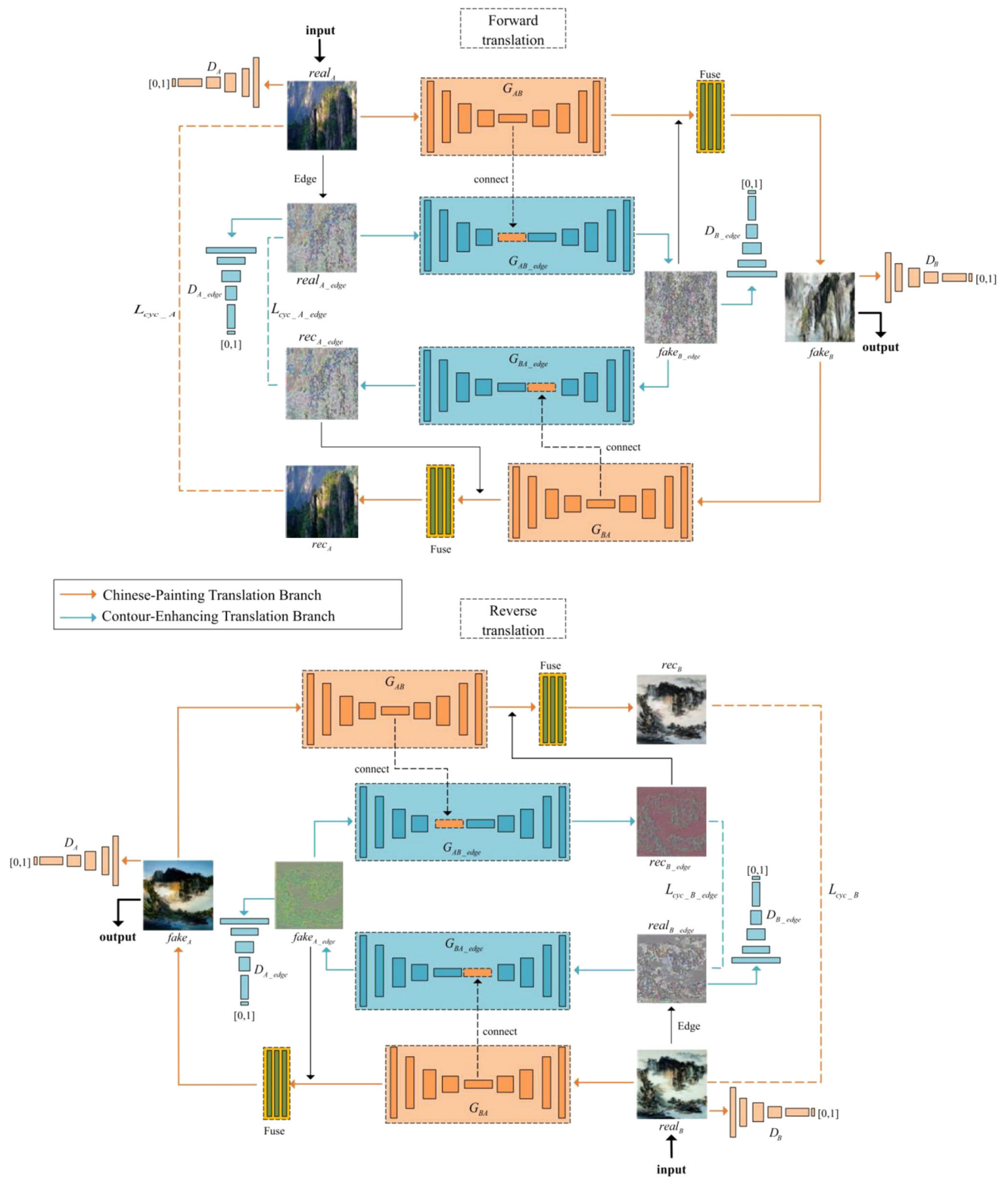


Fig. 4 Diagram of the proposed CE-CycleGAN network structure

map $fake_B$. The generated $fake_B$ will be input to the discriminator D_B with the real edge-enhanced map $real_B$ in reverse translation for authenticity discrimination. The process is described by formulas (1) and (2) as:

$$fake_{B_edge} = (G_{AB_edge}(real_{A_edge})) \tag{1}$$

$$fake_B = Fuse(G_{AB}(real_A), fake_{B_edge}) \tag{2}$$

The corresponding generative adversarial losses also come from the contour-enhancing translation branch and the image translation branch, which are expressed as (3) and (4):

$$L_{GAN_B_edge}(G_{AB_edge}, D_{B_edge}) = eE_{real_{B_edge}}[\log D_{B_edge}(real_{B_edge})] + E_{fake_{B_edge}}[\log(1 - D_{B_edge}(fake_{B_edge}))] \tag{3}$$

$$L_{GAN_B}(G_{AB}, D_B) = eE_{real_B}[\log D_B(real_B)] + E_{fake_B}[\log(1 - D_B(fake_B))] \tag{4}$$

As a result, we realized the translation from scenery photos to landscape paintings. We then need to transfer from landscape paintings back to the original scenery photos to achieve cyclic consistency, the mapping of G_{BA} . The generated edge-enhanced map $fake_B$ is input into the generator G_{BA} of the painted translation branch. A set of restored image features is obtained as the input of the fusion module for restoration. The generated edge-enhanced gradient map $fake_{B_edge}$ is then input into the generator G_{BA_edge} of the contour-enhancing translation branch to generate the image gradient map rec_{A_edge} for restoration. Finally, rec_{A_edge} is spliced with the restored image feature of the painted translation branch and sent to the fusion module to guide gradient information to restore the original scenery photo rec_A . The process can be described as formula (5) and formula (6):

$$rec_{A_edge} = G_{BA_edge}(fake_{B_edge}) \tag{5}$$

$$rec_A = Fuse(G_{BA}(fake_B), rec_{A_edge}) \tag{6}$$

We set the cyclic consistency loss of the restored image gradient map and the real image gradient map, which facilitates the contour-enhancing translation branch to provide more accurate gradient information for the image translation branch to restore the image. At the same time, we also need to ensure the consistency of restored scenery photos and real scenery photos. The L1 distance loss defines the cyclic consistency loss of these two parts as formula (7) and formula (8):

$$L_{cyc_A_edge}(G_{AB_edge}, G_{BA_edge}) = E[||rec_{A_edge} - real_{A_edge}||_1] \tag{7}$$

$$L_{cyc_A}(G_{AB}, G_{BA}) = E[||rec_A - real_A||_1] \tag{8}$$

So far, the forward translation process of our method is completed. The reverse translation is the opposite process to the forward translation. For the $G_{BA} : B \rightarrow A$ mapping, the process can be described as formula (9) and formula (10):

$$fake_{A_edge} = (G_{BA_edge}(real_{B_edge})) \tag{9}$$

$$fake_A = Fuse(G_{BA}(real_B), fake_{A_edge}) \tag{10}$$

The corresponding generative adversarial loss function is defined as formula (11) and formula (12):

$$L_{GAN_A_edge}(G_{BA_edge}, D_{A_edge}) = eE_{real_{A_edge}}[\log D_{A_edge}(real_{A_edge})] + E_{fake_{A_edge}}[\log(1 - D_{A_edge}(fake_{A_edge}))] \tag{11}$$

$$L_{GAN_A}(G_{BA}, D_A) = eE_{real_A}[\log D_A(real_A)] + E_{fake_A}[\log(1 - D_A(fake_A))] \tag{12}$$

For the $G_{AB} : A \rightarrow B$ mapping, the process can be described as expressions (13) and (14):

$$rec_{B_edge} = G_{AB_edge}(fake_{A_edge}) \tag{13}$$

$$rec_B = Fuse(G_{AB}(fake_A), rec_{B_edge}) \tag{14}$$

The corresponding cyclic consistency loss is defined as expressions (15) and (16):

$$L_{cyc_B_edge}(G_{BA_edge}, G_{AB_edge}) = E[||rec_{B_edge} - real_{B_edge}||_1] \tag{15}$$

$$L_{cyc_B}(G_{BA}, G_{AB}) = E[||rec_B - real_B||_1] \tag{16}$$

The overall objective loss function of our model is:

$$L(G, D) = L_{GAN_B_edge} + L_{GAN_B} + L_{GAN_A_edge} + L_{GAN_A} + \lambda(L_{cyc_A} + L_{cyc_A_edge} + L_{cyc_B} + L_{cyc_B_edge}) \tag{17}$$

where λ denotes the relative weight of the generative adversarial loss and the cyclic consistency loss.

4 The proposed ISSIM metrics

One challenge in comparing models is that quantitative comparisons are difficult. For a good model, in addition to generating realistic images (clear), the generated image set must also have both sufficient similarity and diversity. That is, the images in the generated set should be similar to the training set and be representative of the general distribution at the same time. However, while ensuring similarity, overfitting should be avoided. Networks tend to overfit with the deepening of networks and the increase of parameters [24]. The smaller the data set, the more severe the overfitting phenomenon. In the extreme case, all images

in the generation set are copied from the training set. Simultaneously, insufficient sample diversity in generating image sets, or mode collapse [25], are common problems occurring during training a GAN resulting in a generator that outputs a single data point for the different modes of the domain. In the extreme case, the generator collapses to a single output mode which makes the model practically useless since it can only generate extremely similar patterns regardless of the input. Note that, if the data in the training set is realistic (clear) enough, then the generated set with high similarity must also be realistic (clear) enough. So, in general, it is good enough if a model can generate image sets that satisfy both diversity and similarity. In order to evaluate the quality of the generated samples of Chinese landscape paintings reconstructed by our models, the IS [26], FID [27], and KID [28] metrics are introduced in this paper.

The IS [26] metric is the KL-Divergence between conditional and marginal label distributions over the generated data, and the higher is, the better. This score correlates somewhat with the human judgment of sample quality on natural images, which determines whether the image is real (clear) or not. That is, in Inception Net-V3's "world view", any data that does not look like ImageNet is not real (clear). It knows nothing about the desired distribution for the model and cannot evaluate the similarity between the generated set and the real set [28].

The FID [27] is the Wasserstein-2 distance between multi-variate Gaussians fitted to data embedded into a feature space. As an unbiased alternative to FID, the KID [28] measures the dissimilarity between two probability distributions using samples drawn independently from each distribution. Smaller FID and KID values represent better feature distributions in the generated images and thus indicate they are closer to real images [29]. Both of these metrics incorporate the effects of similarity and diversity, whose values will become large due to model collapse [30]. However, neither IS, nor FID can catch the overfitting and underfitting [31], and the same goes for KID. Meanwhile, since both FID and KID take into account similarity and diversity, it is not possible to distinguish which values are specifically influenced by which one.

To overcome these shortcomings, the ISSIM set is proposed in this paper, which can provide richer information. Based on the 2048-dimensional feature vector output of Inception Net-V3, ISSIM for evaluating individual image similarity is proposed referring to the form of traditional structural similarity (SSIM) [32]. Unlike the SSIM, the samples of ISSIM are 2048-dimensional feature vectors $\{x, y\}$ instead of matrixes of image pixel values. The formula for ISSIM is:

$$\text{ISSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (18)$$

where μ_x is the mean of x , μ_y is the mean of y , σ_x^2 is the variance of x , σ_y^2 is the variance of y , and σ_{xy} is the covariance of x and y . The constants C_1 and C_2 are defined to avoid system instability when the denominator is close to 0, which takes $C_1 = 0.0013$, $C_2 = 0.0117$. This ISSIM metric is used to calculate the similarity of two images directly.

Based on this, the other three ISSIM metrics for two image sets were proposed to reflect the similarity between the generated images and the real images in different aspects. For a sample set Y with Inception Net-V3 feature vectors y of m generated images, and a sample set X with feature vectors x of n real images used for comparison, the mean value of the feature vectors can be calculated first, and then ISSIM of two sets is calculated to obtain:

$$\text{ISSIM}_0 = \text{ISSIM}(\bar{X}, \bar{Y}) \quad (19)$$

ISSIM_0 corresponds to the similarity between the mean values of the features of the generated image set and the real image set, which represents whether the centers of the two distributions are close. The maximum ISSIM value of the generated image feature vector Y_i and the feature vectors of the sample set X can also be accumulated and then averaged to obtain:

$$\text{ISSIM}_1 = \frac{1}{m} \sum_i \max_j (\text{ISSIM}(X_j, Y_i)) \quad (20)$$

ISSIM_1 corresponds to the average of the similarity between the features of any image in generated set and the most similar samples that can be found in the real image set, which represents whether the images in the generated set can be found similar in the real set. Similarly, ISSIM_2 is defined as:

$$\text{ISSIM}_2 = \frac{1}{n} \sum_i \max_j (\text{ISSIM}(X_i, Y_j)) \quad (21)$$

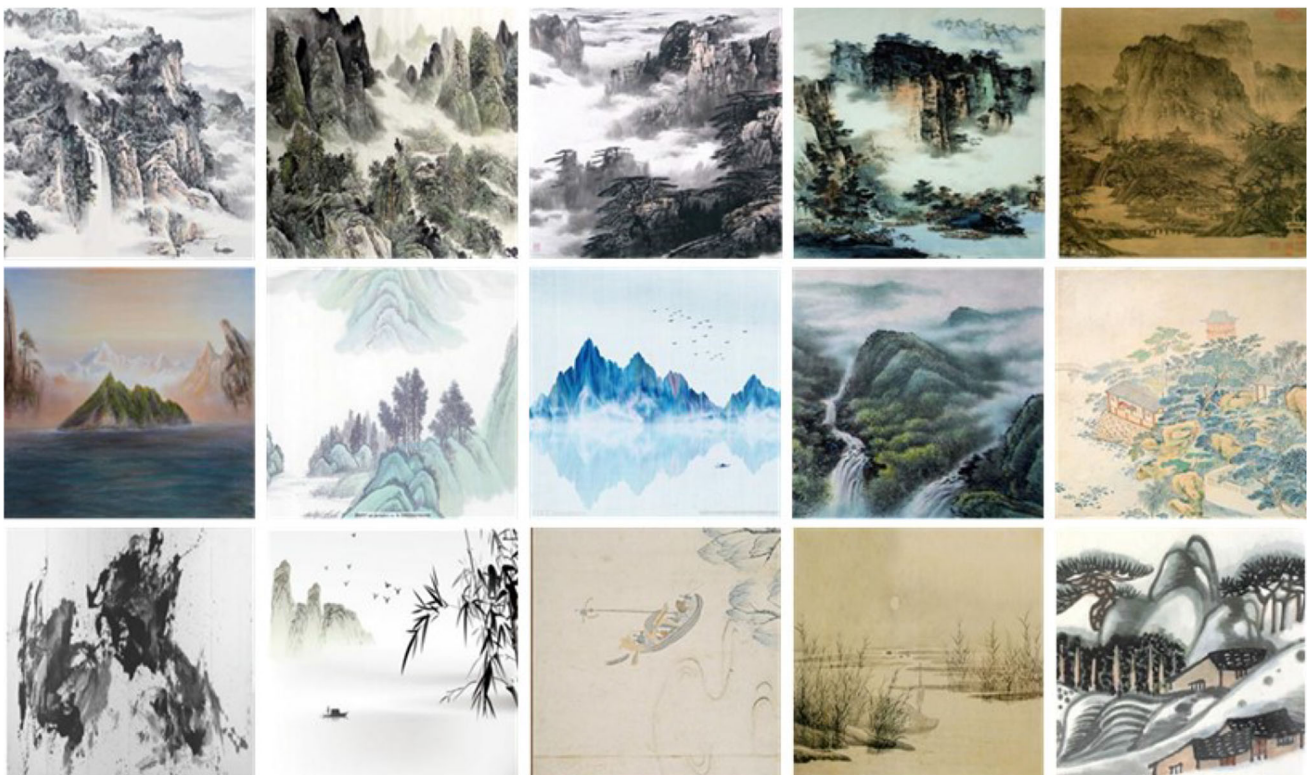
which corresponds to the average of the similarity between the features of any image in the real image set and the most similar samples that can be found in the generated set. This represents whether the images in the real set can be found similar in the generated set. Another method to find ISSIM of two sets is:

$$\text{ISSIM}_3 = \frac{1}{nm} \sum_i \sum_j \text{ISSIM}(X_j, Y_i) \quad (22)$$

ISSIM_3 corresponds to the average similarity of each sample in the generated image set and the real image set, which is named cross-similarity. The larger the values of these four metrics, the higher the similarity between the



(a) Source domain scenery photo examples



(b) Examples of landscape painting target areas

Fig. 5 Examples of source (a) and target (b) landscape images

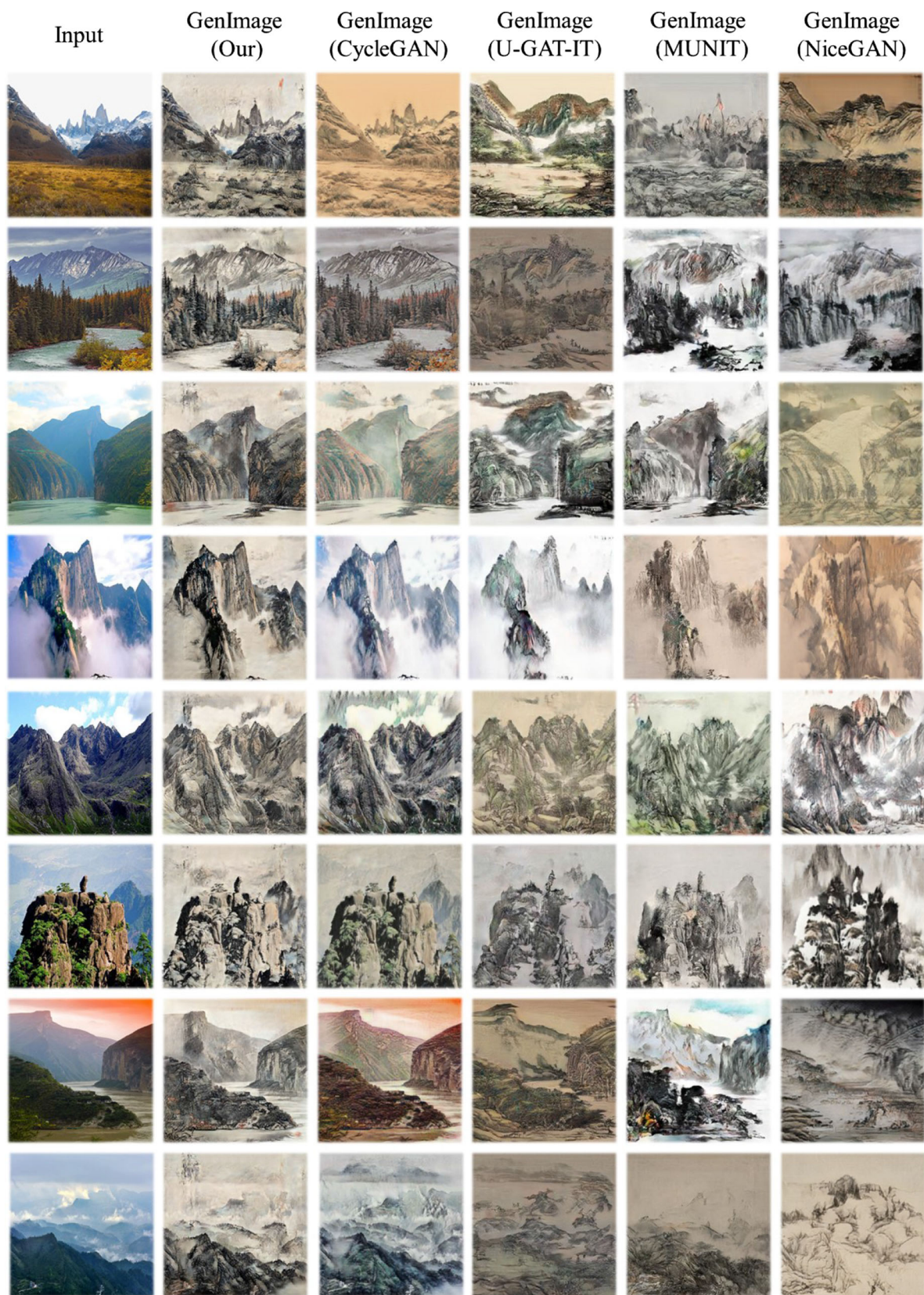


Fig. 6 Generalization simulations

Table 2 The IS (the higher, the better) and FID scores (the lower, the better) of different models

Score	IS	IS ^{Gray}	IS ^{Diff}	FID _{All}	FID _{All} ^{Gray}	FID _{Cun}	FID _{Cun} ^{Gray}	FID _{All} ^{Diff}	FID _{Cun} ^{Diff}
U-GAT-IT	2.64	2.47	− 0.17	91.40	80.58	86.61	84.19	− 10.82	− 2.42
MUNIT	2.42	2.34	− 0.08	100.58	88.92	79.63	69.90	− 11.66	− 9.73
NiceGAN	2.57	2.46	− 0.11	82.32	68.85	66.22	55.49	− 13.47	− 10.73
CycleGAN	2.78	2.81	0.03	94.70	85.21	81.03	67.71	− 9.49	− 13.32
Our	2.23	2.44	0.21	115.17	94.97	80.50	61.14	− 20.2	− 19.36

The models with the best effect or the best improvement after graying have been marked in bold. The superscript “Gray” indicates the result of graying the image. No superscript indicates the result of the original color image, and “Diff” indicates the difference before and after graying. The subscripts All(Cun) subscripts indicate the FID scores of the generated set and the full training set(the “Cun method” set in the training set)

Table 3 The KID scores (the lower, the better), the checkpoint size, and the graphics memory occupancy for batch size being 1 of different models

Score × 100	KID _{All}	KID _{All} ^{Gray}	KID _{Cun}	KID _{Cun} ^{Gray}	Checkpoint size	Graphics memory occupancy
U-GAT-IT	2.66 ± 0.11	3.13 ± 0.10	3.56 ± 0.20	4.31 ± 0.19	2.49 GB	19.92 GB
MUNIT	5.67 ± 0.20	4.40 ± 0.15	4.13 ± 0.15	3.88 ± 0.17	0.46 GB	10.74 GB
NiceGAN	3.47 ± 0.24	2.51 ± 0.14	1.35 ± 0.07	1.51 ± 0.10	7.22 GB	19.88 GB
CycleGAN	3.49 ± 0.18	3.25 ± 0.17	2.12 ± 0.11	1.76 ± 0.08	0.11 GB	9.62 GB
Our	5.28 ± 0.22	3.28 ± 0.17	2.41 ± 0.10	1.27 ± 0.05	0.36 GB	8.52 GB

The models with the best effect have been marked in bold. The superscript “Gray” indicates the result of graying the image, while no superscript indicates the result of the original color image. The subscripts All(Cun) subscripts indicate the KID scores of the generated set and the full training set(the “Cun method” set in the training set)

Table 4 The metrics scores for artificially created abnormal image sets with seven typical abnormality which are simulated by selecting images from the grayed-out test set and generated set

	OF	MC	IS	FID	KID	ISSIM ₀	ISSIM ₁	ISSIM ₂	ISSIM ₃	ISSIM _c	ISSIM _s
Abn1	✓	1	1.59	26.19	0.84 ± 0.07	0.994	<u>0.919</u>	0.954	0.746	0.818	0.753
Abn2	✓	×	2.48	31.45	0.40 ± 0.05	0.997	<u>0.929</u>	0.867	0.659	0.827	0.653
Abn3	×	×	2.89	105.91	6.72 ± 0.18	0.784	<u>0.660</u>	0.636	0.476	0.630	0.581
Abn4	○	1	1.29	131.39	4.14 ± 0.13	0.967	0.842	0.770	0.701	0.862	0.788
Abn5	○	2	1.92	133.91	5.15 ± 0.15	0.966	0.813	0.752	0.646	0.905	0.671
Abn6	✓	1	1.32	109.48	2.42 ± 0.12	0.967	<u>1.000</u>	0.771	0.700	0.585	0.787
Abn7	✓	2	1.59	111.01	1.67 ± 0.07	0.982	<u>1.000</u>	0.769	0.674	0.629	0.706

The models with the best effect have been marked in bold, and the models that have been excluded because of overfitting or underfitting have been marked with underlines. The OF column indicates overfitting, ✓ is overfitting, × is underfitting and ○ is rightfitting. The MC column represents mode collapse, 1 represents mode collapse with a very concentrated sample distribution, 2 represents mode collapse with a broad sample distribution, and × represents no mode collapse occurred. The self-similarity of the test set for Abn1 is 0.751, for Abn2-Abn7 are 0.668, and the self-similarity of the artificially created generation set is listed in the following table

generated set and the real set. However, it is straightforward to deduce from the definition that the wider the distribution range of the images in two image sets X and Y, the smaller the value of ISSIM₃. That is, the better the diversity of the two sets, the smaller the value of ISSIM₃. Then considering diversity and similarity together, when the model is good, ISSIM₃ should take must take a value close to 1 thus ensuring similarity, but not too close to 1 thus

ensuring diversity. The self-similarity ISSIM_s of the real image set can be defined as:

$$ISSIM_s = \frac{1}{n^2} \sum_i \sum_i ISSIM(X_i, X_i) \tag{23}$$

ISSIM_s also represents the diversity of the real image set. The lower the self-similarity, the higher the diversity.

Table 5 The ISSIM scores of different models and the self-similarity of the generator set for the grayed image sets

Score	Compare with full training set					ISSIM _s	Compare with “Cun method” set				
	ISSIM ₀	ISSIM ₁	ISSIM ₂	ISSIM ₃	ISSIM _c		ISSIM ₀	ISSIM ₁	ISSIM ₂	ISSIM ₃	ISSIM _c
U-GAT-IT	0.963	<u>0.820</u>	0.789	0.572	0.804	0.625	0.945	0.795	0.791	0.609	0.882
MUNIT	0.940	0.778	0.747	0.549	0.848	0.674	0.955	0.778	0.762	0.604	0.875
NiceGAN	0.970	<u>0.820</u>	0.760	0.579	0.795	0.632	0.978	0.783	0.773	0.592	0.899
CycleGAN	0.951	0.797	0.767	0.545	0.882	0.585	0.967	0.786	0.788	0.606	0.897
Our	0.952	0.802	0.767	0.570	0.826	0.638	0.979	0.796	0.796	0.639	0.920

The models with the best effect have been marked in bold, and the models that have been excluded because of overfitting have been marked with underlines

By replacing X in the formula with Y , the self-similarity of the generated set can also be calculated.

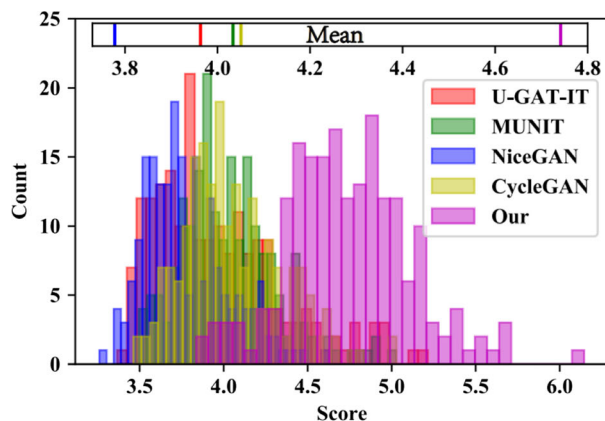
Considering the extreme case, where the distribution of the generated samples is exactly the same as the real image set, then $ISSIM_3 = ISSIM_s$, which takes the suitable value. Then, a scoring function is proposed here as:

$$F(x, s, p, a) = \begin{cases} (x/s)^a & x/s \leq 1 \\ (x/s)^{-p} & x/s > 1 \end{cases} \quad (24)$$

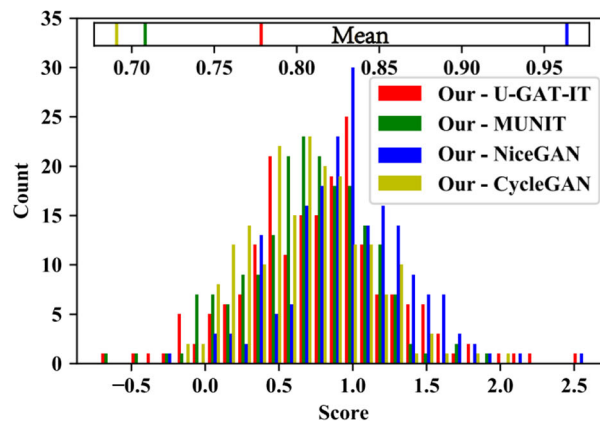
where s is the suitable value, p is the penalty factor, a is the adjustment factor, and p and a are both positive numbers. Then, when $s = ISSIM_s$, choosing the appropriate p and a , the score $F(ISSIM_3, s, p, a)$ related to $ISSIM_3$ will become a value between $[0,1]$. Similarly, if one does not want the generated set to be too close to the real set, one can use $F(ISSIM_1, s, p, a)$ to adjust the scores related to $ISSIM_1$, using the parameter p to penalize $ISSIM_1 > s$ for overfitting. With these adjustments, all scores become the bigger the better. Then, considering the above four metrics together, a composite metric can be defined as:

$$ISSIM_c = (ISSIM_0 F(x_1, s_1, p_1, a_1) F(x_2, s_2, p_2, a_2) F(x_3, s_3, p_3, a_3)) \quad (25)$$

where $x_i = ISSIM_i$, $s_3 = ISSIM_s$, and other parameters should be adjusted appropriately according to the real picture set and the actual situation. For example, if the self-similarity of the real set $ISSIM_s$ is large, then the overfitting thresholds s_1 and s_2 should be adjusted appropriately to large values; If you want the range of the distribution of the generated set and the real set to be as close as possible, you should increase p_3 and a_3 appropriately; if you care whether some modes in the real set are missing in the generated set, you can increase a_2 ; if you care whether the generated set incorrectly generates modes that do not exist in the real set, you can increase a_1 , and so on. The specific parameter values are discussed in the context of specific cases in the analysis of the results later on.



(a) NIMA scores of the five models



(b) Difference between our model and the NIMA scores of the other four models

Fig. 7 NIMA score histograms of 195 test images. The insets show the average of the NIMA scores or the difference of the NIMA scores

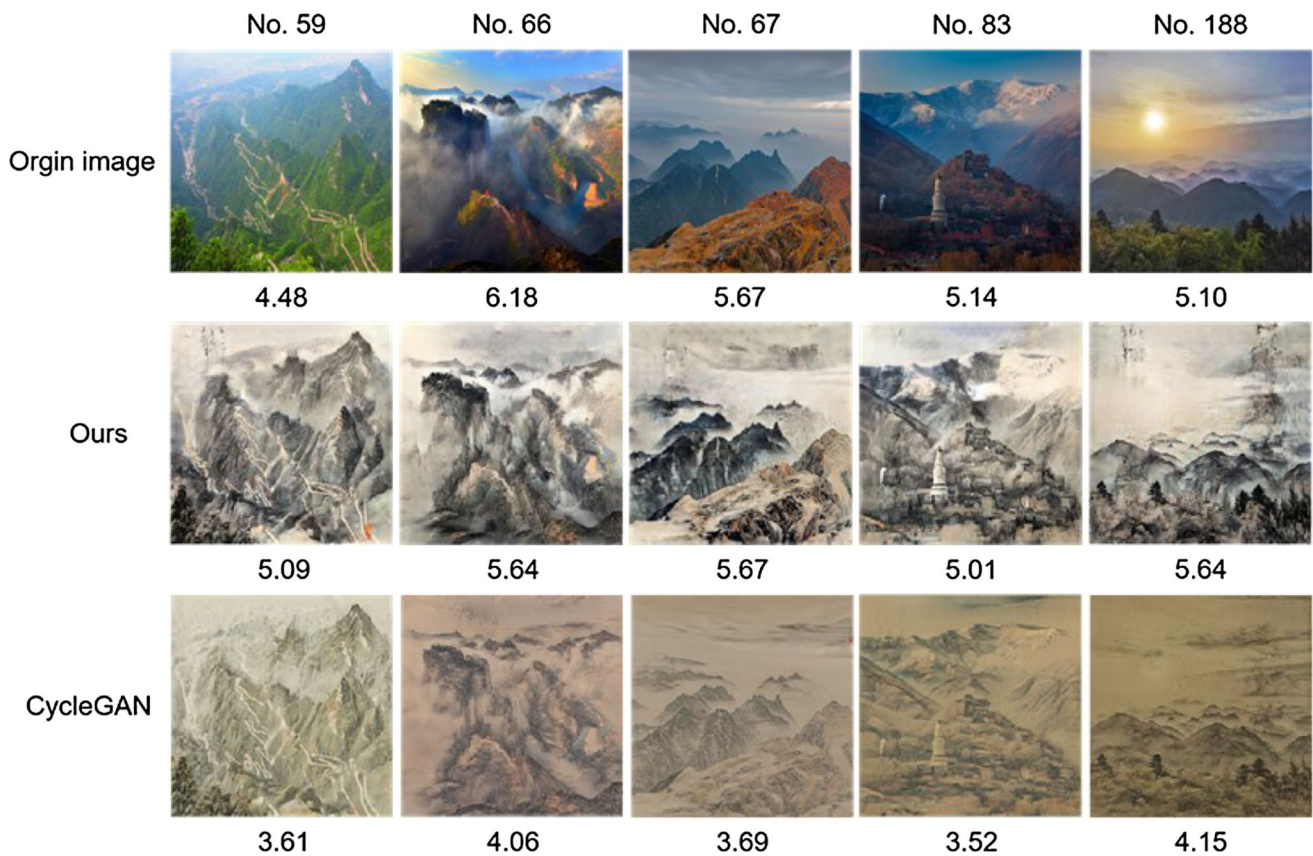


Fig. 8 The top five best performing images where our proposed method outperforms the Cycle-GAN method

5 Simulation results

5.1 Preparation of the simulation

5.1.1 Access to data

The Image-Downloader crawler program uses English and Chinese keywords to grab relevant pictures from the original URL of image search engines (Baidu, Bing, and Google) and download them in batches. A total of 2363 images were collected in the target domain (Chinese landscape paintings), and a total of 2562 images were collected in the source domain (scenery photos).

5.1.2 Data filtering

We use manual screening methods to eliminate irrelevant images. A total of 1956 scenery photos and 1884 Chinese landscape paintings were retained. We randomly assigned 1761 scenery photos as training samples in the source domain and 195 scenery photos as test samples; in the target domain, and also randomly assigned 1696 Chinese landscape paintings as training samples and 188 Chinese landscape paintings as test samples. Some example images

are shown in Fig. 5. This paper aims to better simulate the “Cun method” through neural networks to achieve better results with small samples and minor scales. However, relatively few Chinese landscape paintings were collected with distinctive “Cun method” features when resized to 256×256 -pixel. Of the 1696 training samples, only 695 had significant “Cun method” features, as shown in the first row of Fig. 5. The rest of the samples had general or weak “Cun method” features, as shown in the second and third rows of Fig. 5, respectively. The classification of whether or not the “Cun method” characteristics were significant was manually determined by the art school of the northwest university of professionals. The inclusion of samples with insignificant “Cun method” features improves the generalization ability of the neural network and prevents overfitting in small samples.

5.1.3 Simulation environment

The simulation used a workstation with an NVIDIA RTX 3090 graphics card. The operating system is Ubuntu 20.04 with PyTorch 1.7.1. The CUDA version is 11.2, and the CuDNN version is 8.1.

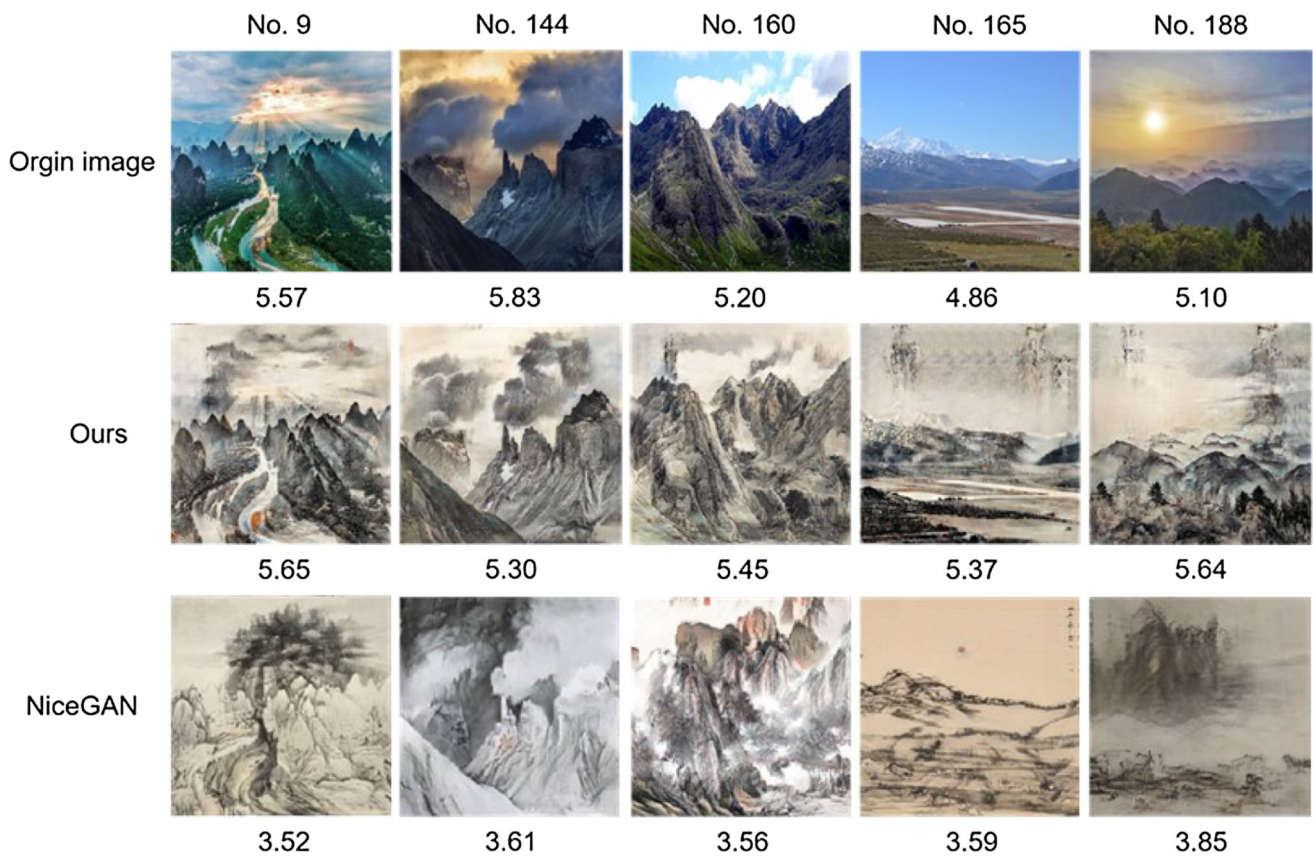


Fig. 9 The top five best performing images where our proposed method outperforms the NiceGAN method

5.1.4 Training process

To facilitate training and evaluation, all images are adjusted to 256×256 resolution. We trained 4,00,000 iters with a learning rate of 0.0002 in the first 200 epochs, and then the learning rate decays linearly to zero in the following 400 epochs (8,00,000 iters). We applied the Adam optimizer [33]. The batch size is set to 1.

5.2 Simulation results and analysis

This section shows the effectiveness of our designed CE-CycleGAN on realistic painting style transfer tasks and introduces the ablation study. All qualitative results shown in this section are evaluated on the test set of the prepared dataset. Figure 6 shows some of the simulation results of our method.

5.2.1 The IS, FID and KID metrics

The IS metric of the generated image set before and after graying is given in Table 2. It can be seen that the IS of the three other models decreases, and the score of CycleGAN, as well as our model, increases compared to the color images. It can be inferred that the three other models are

more sensitive to color information and pay relatively little attention to silhouette information. Before and after graying, the IS^{Diff} of our model is greater than that of the CycleGAN model, which indicates that contour enhancement can make the generated set performs better on sensitivity to image edges. Our model may be more advantageous in simulating image textures, which is the characteristic required to simulate the “Cun method”.

The FID scores of different models are shown in Table 2, which are calculated from the mean and covariance matrices of the 2048-dimensional feature vector set output by the two image collections from the Inception Net-V3. The FID_{All} is the FID score between the generated image set and the full training set, and the FID_{Cun} is the FID score between the generated image set and the “Cun method” set. It can be seen that the NiceGAN model generates the image set closest to the training set as well as the “Cun method” set. Considering that the “Cun method” is only a stroke method that is not sensitive to color information, the FID scores of the real image distribution and the generated image distribution after graying were also calculated. The FID scores significantly improved, and our model has the largest drop as shown by FID_{All}^{Diff} and

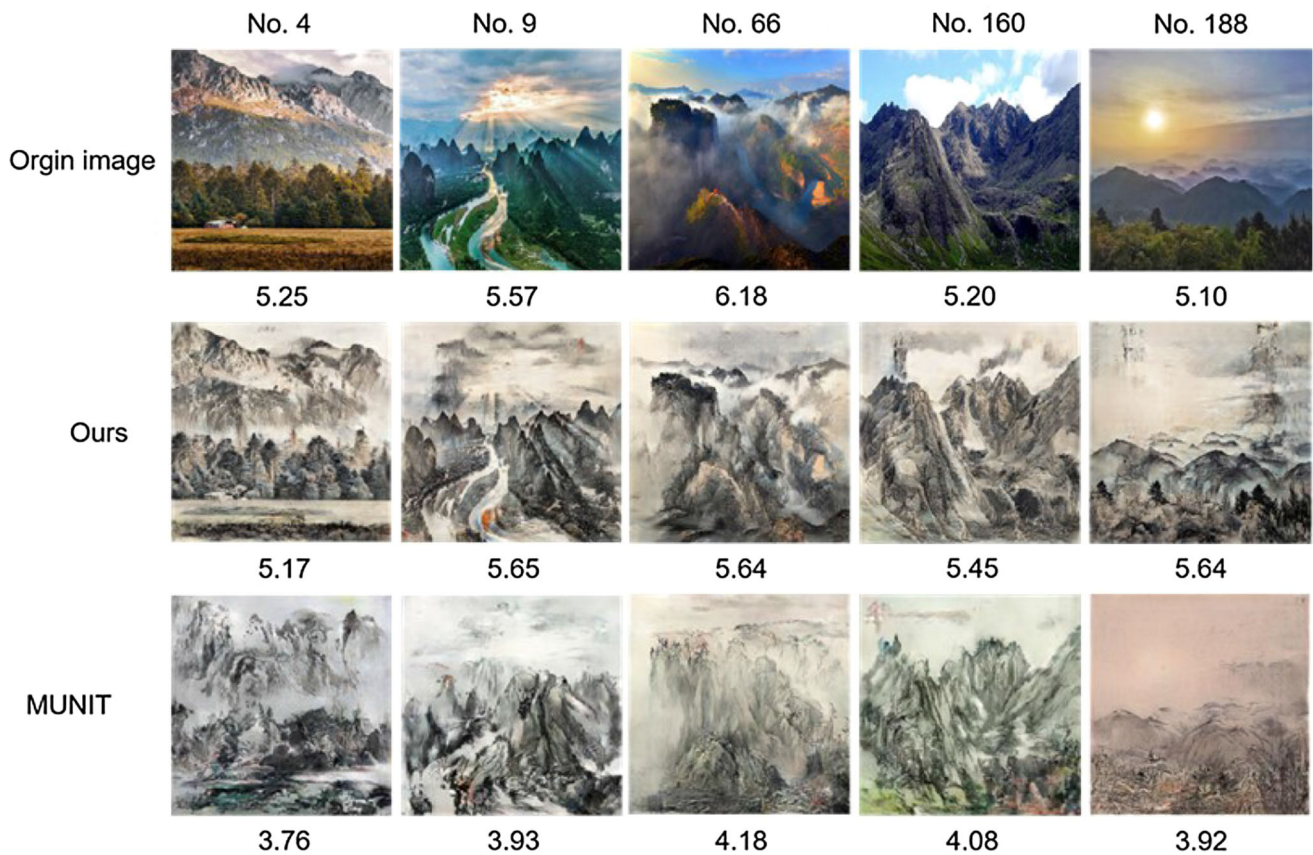


Fig. 10 The top five best performing images where our proposed method outperforms the MUNIT method

FID_{Cun}^{Diff} , which again illustrates that the generated set of our model is more sensitive to edge information.

Although the NiceGAN performed the best on the FID score, the human naked eye perception is not the same. As can be seen in Fig. 6, some of the images generated by NiceGAN are significantly worse for humans to feel. On the contrary, the image set generated by our model, which scored second in FID_{Cun}^{Gray} , gives a subjective feeling closer to the effect of the “Cun method”. So FID may not be the most appropriate metric for this paper, and then KID was calculated.

The KID scores of different models are shown in Table 3, with the variance being also presented to ensure reliable, due to its high variance can make it unreliable to provide values only [34]. As we can see, when color information is considered, the image set generated by U-GAT-IT is closest to the full training set, and the image set generated by NiceGAN is closest to the “Cun method” set. When the color information is not considered, the image set generated by NiceGAN is closest to the full training set, and the image set generated by our model is closest to the “Cun method” set. Overfitting tends to occur

when the sample size is small, and the model has too many parameters. Combining the number of parameters reflected by the checkpoint size in Table 4 and the fact that some of the outputs in the fourth and sixth columns in Fig. 6 show a visible distortion in the contours relative to the inputs, it can be determined that U-GAT-IT and NiceGAN show signs of overfitting. Excluding these two models, the KID scores indicate that the generated set of the CycleGAN model is clearly closer to the full set. After performing contour enhancement, the generated set of the model is closer to the “Cun method” set while ignoring the color information, although the distance to the full set becomes larger. This analysis shows that introducing the edge-enhanced translation branch can make the images generated by the neural network closer to the “Cun method” effect.

5.2.2 The ISSIM metrics

To explore the validity of various metrics, a series of artificially created abnormal sets were tested, and the results of seven typical anomalies are shown in Table 4. Abn1 represents the case of low diversity in the training set, which is made by selecting some similar images from

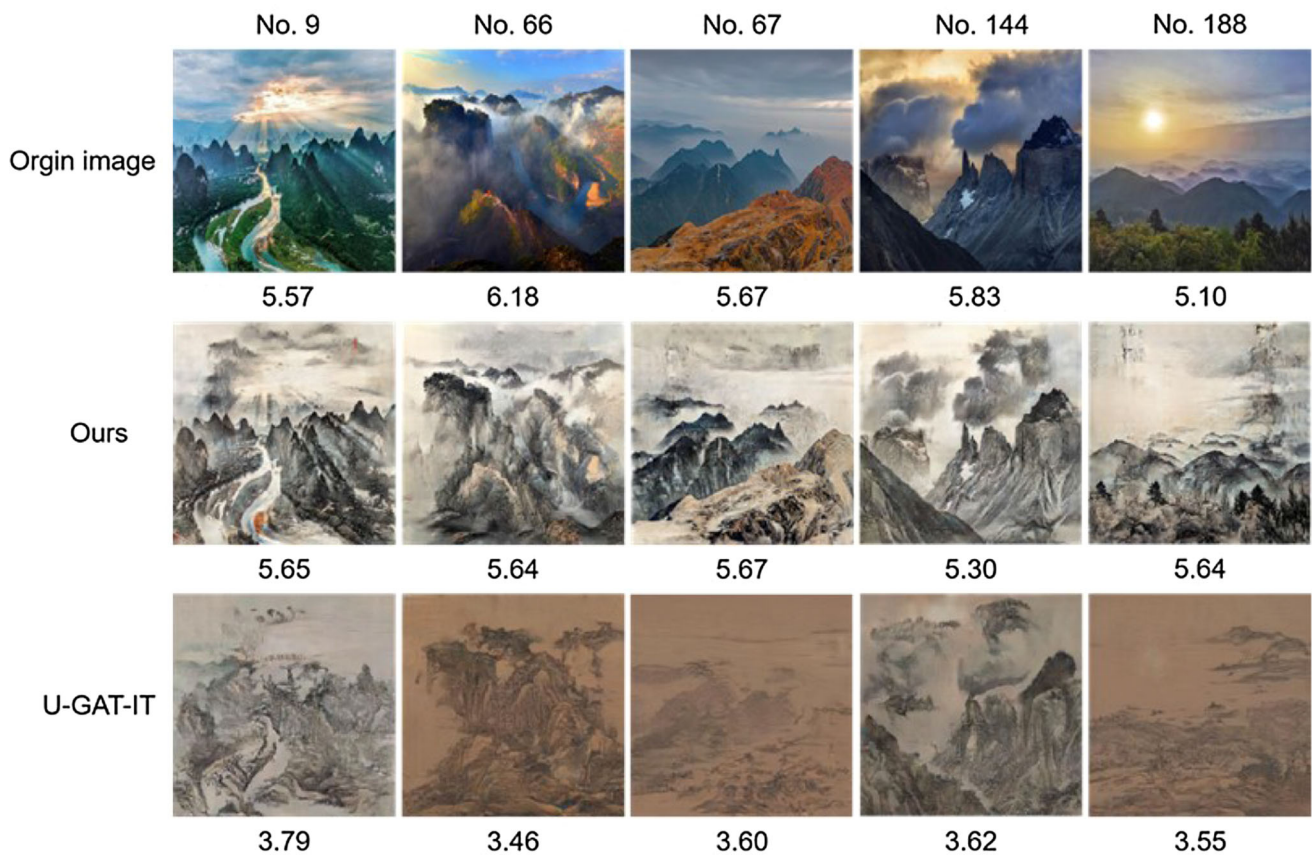


Fig. 11 The top five best performing images where our proposed method outperforms the U-GAT-IT method

the training set and the generation set to form new sets. The self-similarity of Abn1's training set is 0.751, and assume that the generated set is a good fit with the training set. Abn2-Abn7 use the grayed "Cun method" set as the "training set". Some images from the "Cun method" set were copied to the "training set" to simulate overfitting in Abn1-Abn2 and Abn6-Abn7. Abn6 and Abn7 are extreme cases where all the images in the "generated set" are from the "Cun method" set. In Abn1 and Abn4-Abn7, the same images are copied several times to simulate mode collapse. But in mode collapse case 1, the distribution of the copied images is more concentrated, while in mode collapse case 2, the distribution of the copied images is broader. In Abn3, randomly selected scenery photos were used as the "generation set" to simulate the underfitting.

From the test results, the IS scores are more sensitive to mode collapse. Once the mode collapse occurs, the IS score will become worse. For the FID and KID metrics, overfitting leads to better scores, and mode collapse leads to worse scores. So when the scores of these two metrics are good, one needs to be alert to whether overfitting has occurred. In the case of severe underfitting (Abn3), FID still achieves relatively good scores, while KID is relatively more sensitive. From this point of view, the KID metric is

more reasonable than the FID metric. In contrast, the ISSIM metrics proposed in this paper, because it is directly related to the similarity, its value can visually reflect the overfitting and underfitting. Overfitting directly leads to higher scores for $ISSIM_1$, and underfitting directly leads to lower scores. Determining whether overfitting or underfitting occurs requires determining a threshold value, which should be related to the self-similarity of the training set. Note that for different models, the determination of the threshold needs to be combined with the actual situation of the model, and it is best to carry out a certain test and combine the effect of human visual perception. Through testing, this paper sets the threshold for overfitting as $\min(0.9, (ISSIM_1)^{1/3})$ and the threshold for underfitting as $(ISSIM_1)^{0.8}$. Then, the right fitting range for Abn1 is $[0.795, 0.9]$ and for Abn2-Abn7 is $[0.724, 0.874]$. The outliers that $ISSIM_1$ does not fall within these ranges have been underlined in Table 4. It can also be seen from Table 4 that mode collapse leads to an increase in the self-similarity of the generated set, especially for mode collapse case 1.

The above discussion of the test results shows that the ISSIM metrics do provide intuitive similarity and a richer

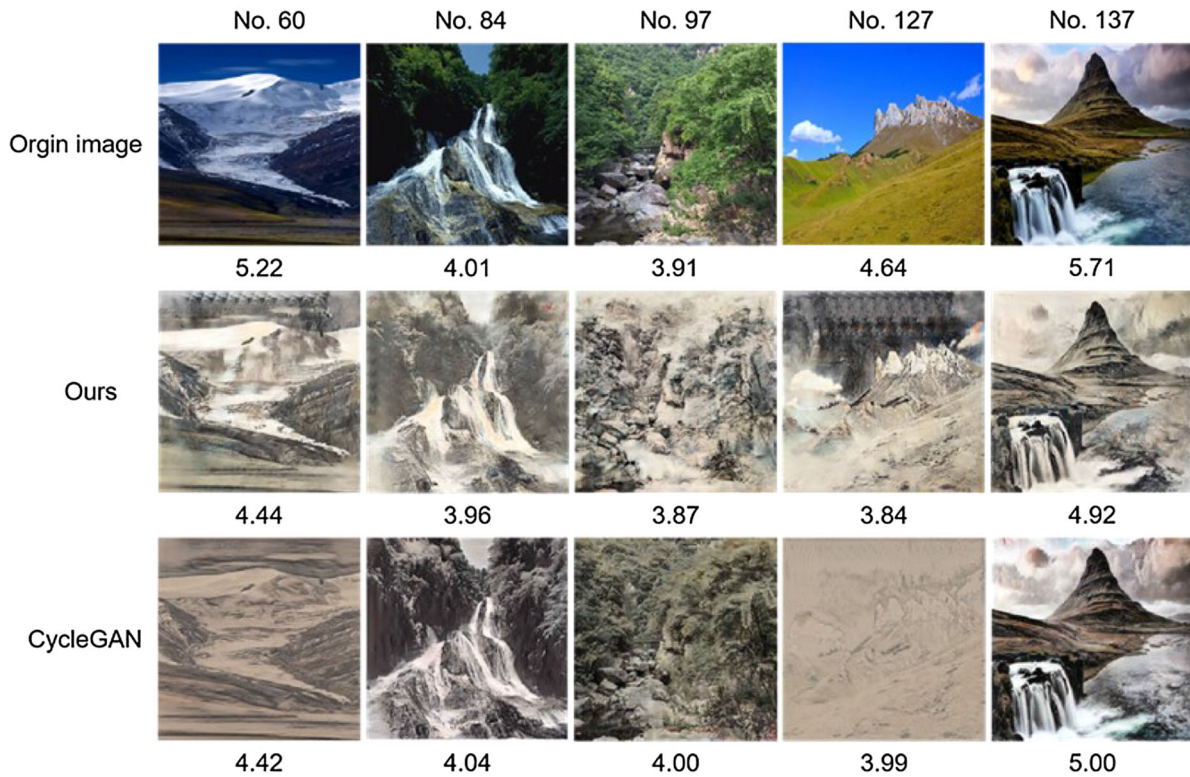


Fig. 12 The top five worst performing images of our proposed method compared to the CycleGAN method

set of information than other metrics. In order to combine this information and give a single reference value, parameters can be set to calculate the composite metric $ISSIM_c$. s_1 and s_2 can be set to the threshold value mentioned above as $\min(0.9, (ISSIM_1)^{1/3})$. For the value of $ISSIM_1$ - $ISSIM_3$ exceeding the threshold, the penalty intensity gradually weakens. p_1 - p_3 are set to 12, 9, and 6, respectively. These values are relatively large, making it impossible to obtain a high score for the overfitting cases. In this paper, we focus more on whether the images in a generated set are similar to the images in the training set, so the importance of $ISSIM_1$ - $ISSIM_3$ decreases. Then a_1 to a_3 are set to 1.5, 0.9, and 0.5, respectively. The calculated composite metric $ISSIM_c$ based on these parameter settings has been listed in Table 4, and it can be seen that the scores for both overfitting and underfitting are relatively low. For cases where the fit is good but the mode collapses (similar to Abn4-Abn5), the case with mode collapse with a broad sample distribution (similar to Abn5) scores higher.

The $ISSIM$ results of the grayed image sets are shown in Table 5. The self-similarity of the test set for the full training set is 0.527, for the “Cun method” set is 0.668. Compared to the full training set, the diversity of the “Cun method” set is reduced, and therefore the self-similarity increases from 0.527 to 0.668. In accordance with the

previous discussion, the right fitting range of $ISSIM_1$ for the full training set is [0.599, 0.808], and for the “Cun method” set is [0.724, 0.874]. $ISSIM_1$ for U-GAT-IT and NiceGAN are larger than 0.808, so these two models are suspected of overfitting. Combined with the specific results shown in Fig. 6, it can be determined that overfitting occurred in both models may be due to their excessive parameters compared to the number of training samples. Therefore, these two models should be excluded (underlined in Table 4) when considering which model works better on the tasks in this paper. The self-similarity of the generation set of all models is larger than 0.527, which suggests that all models are suspect of model collapse, especially for MUNIT with the largest generation set self-similarity.

From the composite metric $ISSIM_c$, the generated set of CycleGAN has the best similarity with the full training set, while the generated set of our model has the best similarity with the “Cun method” set. It can be seen from Table 5 that after contour enhancement, the $ISSIM$ metrics have been improved, and the cross-similarity has been improved from 0.606 to 0.639. That is, the contour enhancement makes the modes collapse toward the “Cun method” set. For the full training set, the effects of the models in order from good to bad are CycleGAN > Our > MUNIT > U-GAT-IT > NiceGAN; for the “Cun method” set are

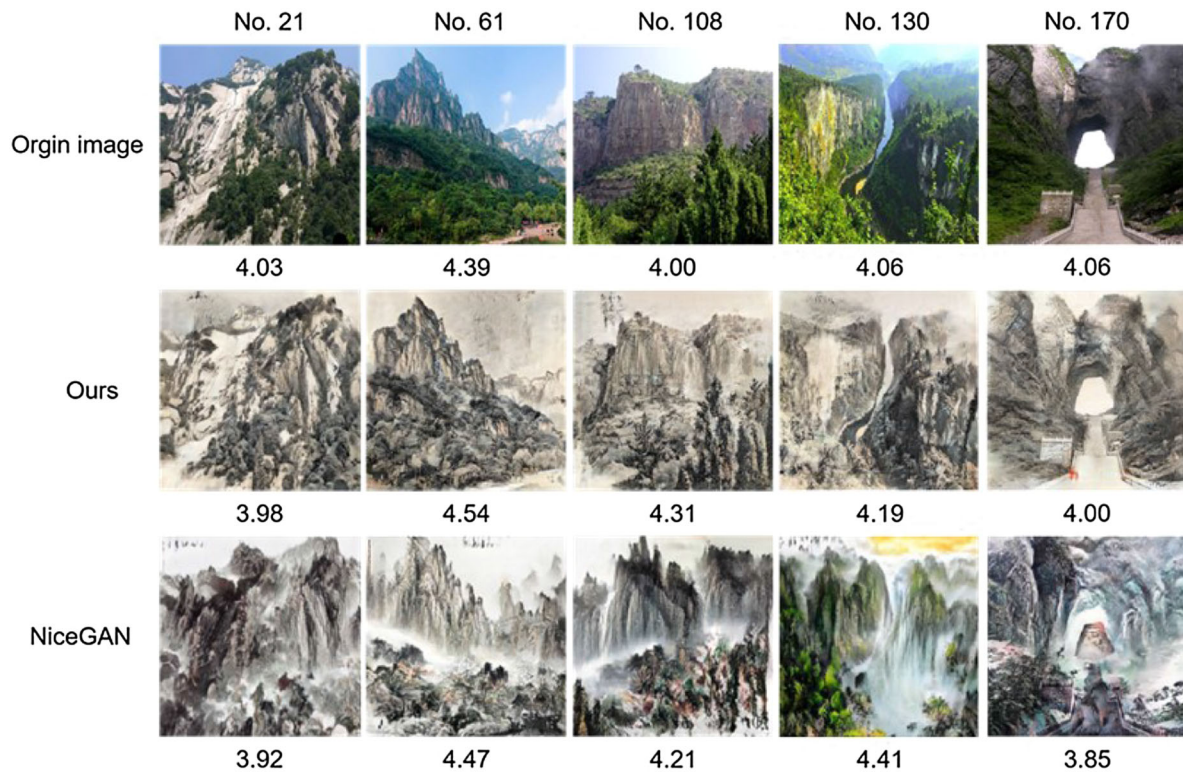


Fig. 13 The top five worst performing images of our proposed method compared to the NiceGAN method

Our > CycleGAN > MUNIT if two over-fitted models are excluded. The score of $ISSIM_c$ is consistent with human visual perception, which indicates that the previously set parameters are reasonable.

So, the similarity with the “Cun method” set is the best in our model, with a 97.89% similarity in the feature means, a 63.94% similarity in the overall compared features, a 79.62% similarity by finding the most similar images, and the composite similarity metric being 0.92. The $ISSIM$ proposed in this paper is as simple to calculate as the FID score, also takes into account the judgment of diversity, and can effectively describe the similarity between image sets.

5.2.3 The NIMA metric

In addition to the proximity of the generated images to the “Cun method,” this paper also focuses on the aesthetic quality of the generated images. Evaluating image technology quality and aesthetic effects have been a long-standing problem in image processing and computer vision [35]. The technical quality assessment measures the image damage at a pixel level, such as noise, blurriness, and artificial compression. In contrast, the aesthetic effect assessment captures the semantic features of emotion and beauty in the image. In general, image quality assessment

can be divided into two types [36]: Peak Signal to Noise Ratio (PSNR) with full reference (FR), and standard structure similarity (SSIM) [32] with no-reference (NR) [32]. This paper also focuses on the artistic effect of transformed images, so this technical quality evaluation method is not applicable. Fortunately, NIMA [20] can predict the human evaluation opinions on images from direct perception and attractiveness, which has advantages similar to human subjective scoring, so we choose it as the image quality evaluation metric. NIMA generates a score histogram for any image. The image is scored 1–10 points, and the images of the same subject are compared directly. This design is consistent in form with the histogram generated by a human scoring system, and the evaluation effect is closer to the result of human evaluation.

Figure 7 shows the comparison of NIMA score histograms between our method, CycleGAN method, NiceGAN, MUNIT, and U-GAT-IT [16] on 195 test images. It can be seen from Fig. 7 that our method is superior to the CycleGAN method, the NiceGAN method, the MUNIT method, and the U-GAT-IT method. The average score of the images generated by our method is 4.74 points, the CycleGAN method is 4.05 points, the NiceGAN method is 3.78 points, the MUNIT method is 4.03 points, and the U-GAT-IT method is 3.96 points. From the evaluation metrics of NIMA, the performance of our method improves

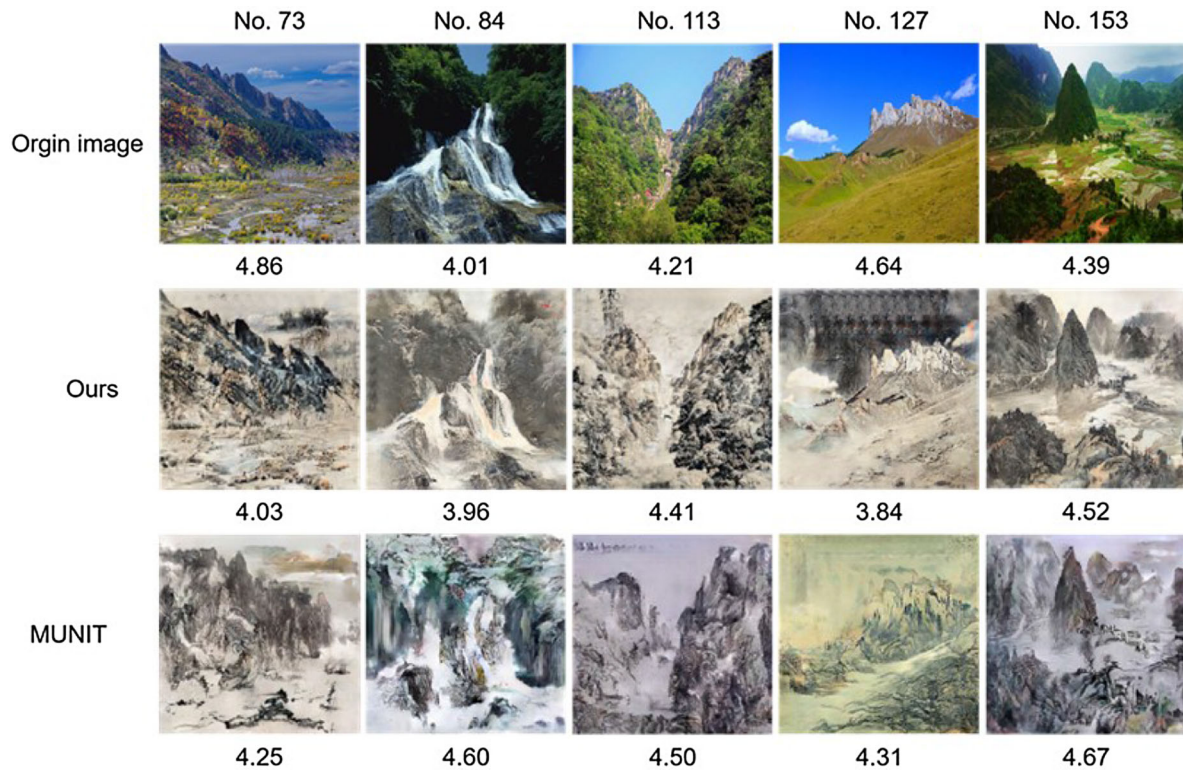


Fig. 14 The top five worst performing images of our proposed method compared to the MUNIT method

about 17% on average relative to the performance of the CycleGAN method, about 25% on average relative to the NiceGAN method, about 18% on average relative to the MUNIT method, and about 20% on average relative to the U-GAT-IT method.

To further analyze the effectiveness of our method, we selected the top five images that outperform the CycleGAN method, NiceGAN method, MUNIT method, and the U-GAT-IT method, as shown from Figs. 8, 9, 10 and 11, respectively. Meanwhile, the worst five images compared with the CycleGAN, NiceGAN, MUNIT, and U-GAT-IT methods are shown from Figs. 12, 13, 14 and 15. We can find that the original photos in Figs. 8, 9, 10 and 11 are relatively rich in content, and their NIMA scores are higher, as shown in Table 6. In contrast, the original photos, shown in Figs. 12, 13, 14 and 15, have relatively monotonous contents, and their NIMA scores are generally lower, as shown in Table 7. These results are consistent with our expectations. Our approach consists of highlighting the boundaries of the rocks and trees in photos. When the image is informative, these boundaries stand out. When the images are monotonous, the boundary information is not rich. Therefore, the method in this paper is mainly applicable to the case of photos with rich details.

6 Conclusion

This paper proposed the CE-CycleGAN framework to transfer scenery photos to a realistic landscape painting style with the “Cun Model.” An edge detection operator is introduced for the distinctive features of the edges of landscape paintings. A gradient-guided path is designed after obtaining the gradient of the image, which enhances the edge transformation from the photograph to the painting. The KID metric and ISSIM metric produce somewhat similar results to the “Cun method”. For our experiments, the ISSIM proposed in this paper is a more powerful metric than FID and KID, which intuitively give the similarity between the generated set and the real set. The simulation results also showed that the subjective evaluation effect of this method is satisfactory, and it is superior to the other four methods under the NIMA metric. The results of our work show that introducing the NIMA metric and ISSIM metric into the loss function would be a feasible improvement direction.

Although the CE-CycleGAN method in this paper has made progress in edge prominence, the generated result is still awaiting further optimization in the future compared with the realistic paintings method used by the artist. Chinese landscape painting focuses on realism, unlike Western painting, so the presence of the sky or clouds



Fig. 15 The top five worst performing images of our proposed method compared to the U-GAT-IT method

Table 6 Our method is compared with the Cycle-GAN method and U-GAT-IT method, respectively, and the NIMA scores of the top five best-performing images

	Image No.59	Image No.66	Image No.67	Image No.83	Image No.188
Ours	5.09	5.64	5.67	5.01	5.64
CycleGAN	3.61	4.06	3.69	3.52	4.15
	Image No.9	Image No.144	Image No.160	Image No.165	Image No.188
Ours	5.65	5.30	5.45	5.37	5.64
NiceGAN	3.52	3.61	3.56	3.59	3.85
	Image No.4	Image No.9	Image No.66	Image No.160	Image No.188
Ours	5.17	5.65	5.64	5.45	5.64
MUNIT	3.76	3.93	4.18	4.08	3.92
	Image No.9	Image No.66	Image No.67	Image No.144	Image No.188
Ours	5.65	5.64	5.67	5.30	5.64
U-GAT-IT	3.79	3.46	3.60	3.62	3.55

The image with higher NIMA scores have been marked as bold

turning into mountains in the generated images is irrelevant. However, it is essential to note that Chinese landscape painting focuses on white space, which is not reflected in the generated images and could be improved by adding selective attention mechanisms in the future. The use of the Sobole operator to highlight contours in this paper does bring the generated image closer to the “Cun

method”. We infer that the effectiveness comes from the fact that the “Cun method” itself has significant contours. Therefore, the CE-CycleGAN method in this paper will also be effective for large samples and large sizes. It also reveals that a better way to simulate the “Cun method” is to enhance the contours by different methods, for example, introducing other edge extraction and contour extraction

Table 7 Our method is compared with Cycle-GAN method and U-GAT-IT method respectively, and the NIMA scores of the top five worst performing images

	Image No.60	Image No.84	Image No.97	Image No.127	Image No.137
Ours	4.44	3.96	3.87	3.84	4.92
CycleGAN	4.42	4.04	4.00	3.99	5.00
	Image No.21	Image No.61	Image No.108	Image No.130	Image No.170
Ours	3.98	4.54	4.31	4.19	4.00
NiceGAN	3.92	4.47	4.21	4.41	3.85
	Image No.73	Image No.84	Image No.113	Image No.127	Image No.153
Ours	4.03	3.96	4.41	3.84	4.52
MUNIT	4.25	4.60	4.50	4.31	4.67
	Image No.2	Image No.22	Image No.90	Image No.119	Image No.185
Ours	4.52	4.44	4.06	4.21	4.41
U-GAT-IT	4.93	4.80	4.78	4.41	4.66

The image with higher NIMA scores have been marked as bold

operators, as well as further adjustment of the line thickness and trend of the extracted contours to be closer to the texture strokes of “Cun method”, or even using different contour enhancement effects to simulate different “Cun method” effects.

Funding This work was supported in part by the National Natural Science Foundation of China under Grant 62006191, in part by the Key RD Program of Shaanxi under Grant 2021ZDLGY15-03, 2021ZDLGY15-04, in part by Changjiang Scholars and Innovative Research Team in University under Grant IRT-17R87, in part by the Xi’an Key Laboratory of Intelligent Perception and Cultural Inheritance under grant 2019219614SYS011CG033 and in part by the Shaanxi Provincial Department of Education Special Scientific Research Project 20JK0940.

Declarations

Conflict of interest Xianlin Peng, Shenglin Peng, Qiyao Hu, Jinye Peng, Jiabin Wang, Xinyu Liu, and Jianping Fan declare that they have no conflict of interest

References

- Huang SW, Way DL, Shih ZC (2003) Physical-based model of ink diffusion in Chinese ink paintings. *J World Soc Comput Graph* 10:520
- Huang L, Hou Z, Zhao Y, Zhang D (2019) Research progress on and prospects for virtual brush modeling in digital calligraphy and painting. *Front Inf Technol Electron Eng* 20:1307
- Li XX, Li Y (2006) Simulation of Chinese ink-wash painting based on landscapes and trees. In: Fourcaud T, Zhang XP (eds) 2006 Second international symposium on plant growth modeling and applications, vol 328. IEEE, Los Alamitos
- Chen TD, Yu CH (2009) Hairy brush model interactive simulation in Chinese ink painting style. In: The 2009 international symposium on information processing (ISIP), vol 184. Citeseer
- Chen T (2009) Non-photorealistic rendering of ink painting style diffusion. In: Lin TY, Hu XH, Xia JL, Hong TP, Shi ZZ, Han JC, Tsumoto S, Shen ZJ (eds) 2009 IEEE international conference on granular computing, vol 78. IEEE, New York
- Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial Networks 27:2672
- Zheng Z, Yang X, Yu Z, Zheng L, Yang Y, Kautz J (2019) Joint discriminative and generative learning for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, vol 2138
- Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H (2016) Generative adversarial text to image synthesis. In: International conference on machine learning, vol 1060. PMLR
- Zhu J, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, vol 2223
- Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434)
- Brock A, Donahue J, Simonyan K (2018) Large scale GAN training for high fidelity natural image synthesis. arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434)
- Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, vol 4401
- Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, Metaxas D (2017) StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision, vol 5907
- Antipov G, Baccouche M, Dugelay JL (2017) Face aging with conditional generative adversarial networks. In: 2017 IEEE international conference on image processing (ICIP), vol 2089. IEEE
- Choi Y, Choi M, Kim M, Ha JW, Choo J (2018) StarGAN: unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, vol 8789
- Kim J, Kim M, Kang H, Lee K (2019) U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance

- normalization for image-to-image translation. arXiv preprint [arXiv:1907.10830](https://arxiv.org/abs/1907.10830)
17. Huang X, Liu M, Belongie S, Kautz J (2018) Multimodal unsupervised image-to-image translation. In: Proceedings of the European conference on computer vision (ECCV), vol 172
 18. Chen R, Huang W, Huang B, Sun F, Fang B (2020) Reusing discriminators for encoding towards unsupervised image-to-image translation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, vol 8168
 19. Chen M (2018) The modern meaning of traditional Cun method. *Fine Arts* 1:58–61
 20. Talebi H, Milanfar P (2018) NIMA: neural image assessment. *IEEE Trans Image Process* 27:3998
 21. Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, vol 1125
 22. Zhang F, Gao H, Lai Y (2020) Detail-preserving CycleGAN-AdaIN framework for image-to-ink painting translation. *IEEE Access* 8:132002
 23. Kittler J (1983) On the accuracy of the Sobel edge detector. *Image Vis Comput* 1:37
 24. Chen A, Xing H, Wang F (2020) A facial expression recognition method using deep convolutional neural networks based on edge computing. *IEEE Access* 8:49741
 25. Karatsiolis S, Christos S (2020) Modular domain-to-domain translation network. *Neural Comput Appl* 32:6779
 26. Barratt S, Sharma R (2018) A note on the inception score. arXiv preprint [arXiv:1801.01973](https://arxiv.org/abs/1801.01973)
 27. Dowson DC, Landau BV (1982) The Fréchet distance between multivariate normal distributions. *J Multivar Anal* 12:450
 28. Bińkowski M, Sutherland DJ, Arbel M, Gretton A (2018) Demystifying MMD GANs. arXiv preprint [arXiv:1801.01401](https://arxiv.org/abs/1801.01401)
 29. Hung SK, Gan JQ (2021) Facial image augmentation from sparse line features using small training data. In: International workshop on artificial neural networks, vol 547. Springer
 30. Shmelkov K, Schmid C, Alahari K (2018) How good is my GAN?. In: Proceedings of the European conference on computer vision (ECCV), vol 213
 31. Devries T, Romero A, Pineda L, Taylor GW, Drozdal M (2019) On the evaluation of conditional GANs. arXiv preprint [arXiv:1907.08175](https://arxiv.org/abs/1907.08175)
 32. Zhou W, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13:600
 33. Devan P, Khare N (2020) An efficient XGBoost–DNN-based classification model for network intrusion detection system. *Neural Comput Appl* 32:12499–12514
 34. Ravuri S, Vinyals O (2019) Seeing is not necessarily believing: limitations of biggans for data augmentation. In: ICLR workshop on international conference on learning representations
 35. He N, Xie K, Li T, Ye Y (2017) Overview of image quality assessment. *J Beijing Inst Graph Commun* 25:47
 36. Wu J, Xia Z, Zhang H, Li H (2018) Blind quality assessment for screen content images by combining local and global features. *Digit Signal Process* 91:31

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Xianlin Peng¹ · Shenglin Peng²  · Qiyao Hu² · Jinye Peng² · Jiaxin Wang² · Xinyu Liu³ · Jianping Fan⁴

✉ Shenglin Peng
pengshenglin@nwu.edu.cn

✉ Qiyao Hu
huqiyao@nwu.edu.cn

✉ Jinye Peng
pjy@nwu.edu.cn

Xianlin Peng
pxl@nwu.edu.cn

Jiaxin Wang
wangjiaxin@stumail.nwu.edu.cn

Xinyu Liu
liuxinyu@xpu.edu.cn

Jianping Fan
jfan@uncc.edu

¹ Art School, Northwest University, Xi'an 710127, China

² College of Information Science and Technology, Northwest University, Xi'an 710127, China

³ Network and Information Management Office, Xi'an Polytechnic University, Xi'an 710048, China

⁴ Department of Computer Science, University of North Carolina at Charlotte, Charlotte 28223, USA