Received: 12 February 2020 | Revised: 6 September 2020 | Accepted: 25 September 2020

IET Image Processing

DOI: 10.1049/ipr2.12059

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Multi-style Chinese art painting generation of flowers

Feifei Fu | Jiancheng Lv | Chenwei Tang | Mao Li

College of Computer Science, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, People's Republic of China

Correspondence
Jiancheng Lv, Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, People's Republic of China.
Email: lvjiancheng@scu.edu.cn

Funding information
National Natural Science Fund for Distinguished Young Scholar, Grant/Award Number: 61625204; the State Key Program of National Science Foundation of China, Grant/Award Number: 61836006

Abstract
With the proposal and development of Generative Adversarial Networks, the great achievements in the field of image generation are made. Meanwhile, many works related to the generation of painting art have also been derived. However, due to the difficulty of data collection and the fundamental challenge from freehand expressions, the generation of traditional Chinese painting is still far from being perfect. This paper specialises in Chinese art painting generation of flowers, which is important and classic, by deep learning method. First, an unpaired flowers paintings data set containing three classic Chinese painting style: line drawing, meticulous, and ink is constructed. Then, based on the collected dataset, a Flower-Generative Adversarial Network framework to generate multi-style Chinese art painting of flowers is proposed. The Flower-Generative Adversarial Network, consisting of attention-guided generators and discriminators, transfers the style among line drawing, meticulous, and ink by an adversarial training way. Moreover, in order to solve the problem of artefact and blur in image generation by existing methods, a new loss function called Multi-Scale Structural Similarity to force the structure preservation is introduced. Extensive experiments show that the proposed Flower-Generative Adversarial Network framework can produce better and multi-style Chinese art painting of flowers than existing methods.

1 | INTRODUCTION

Traditional Chinese paintings continue to innovate on the basis of inheritance. Among them, meticulous painting and ink painting are always the mainstream of Chinese art painting. The content of traditional Chinese painting is mainly natural scenery. Especially, the Chinese painting of flowers is an important and classic form of expression. Figure 1 shows three common traditional Chinese paintings about flowers, that is, real line drawing painting, meticulous painting, and ink painting. In Figure 1, we can see that the real line drawing requires clear and concise strokes. The meticulous paintings have high requirements for colour and texture, based on the same precise strokes as line drawing. The ink paintings pay more attention to the change of ink colour and the harmony between intensity and dryness. Generally speaking, the meticulous paintings realise the simulation of painting objects through lots of colours and precise strokes. The ink paintings achieve freehand expression by limited colour and natural smooth strokes. Without clear object segmentation line, the generation of ink paintings by using the depth of ink colour to construct the painting's structure is challenging.

From the methods based on traditional machine learning to the methods based on recently deep learning, image generation is always a hot area that scholars are constantly exploring. Since the early 2000s, Gatys et al. [1] show the strong power of Convolutional Neural Network (CNN) to extract the visual features. After that, the methods based on CNN have become a popular theme to solve image tasks. Recently, the Generative Adversarial Networks (GANs) [2–4], which can generate high-quality images through the adversarial training between generator and discriminator, has been recognised as one of the most popular methods for computer vision tasks, especially for the image generation and style transfer. The GauGAN [5] can generate realistic pictures through lines and colour blocks drawn by users at will. By leveraging Spatially Adaptive normalisation (SPADE) generator, this model can better preserve the semantic information. The BigGAN [6] can generate high-fidelity, high-quality images with natural boundaries. By leveraging orthogonal regularisation to the generator and making it obey the truncation

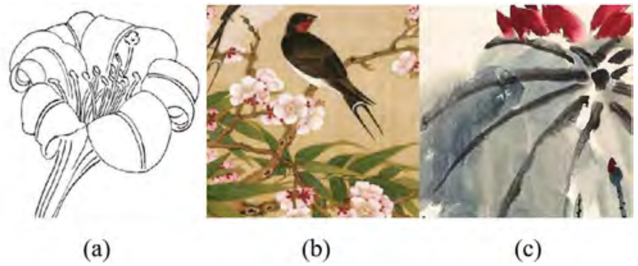


FIGURE 1 From left to right, (a) real line drawing, (b) meticulous painting, (c) ink painting, respectively

trick, the generation performance of GAN is greatly improved. In addition, there are also some classic style transfer models. For example, the Pix2Pix [7] implements the style transfer of paired images by using U-Net generator and patch discriminator, which provides a general framework for image-to-image translation problems. The CycleGAN [8] further proposes a cycle-consistency constraint to achieve the style transformation with unpaired data, thereby solving the problem that lacking of paired data in nature is difficult to train. The styleGAN [9] model proposes a new generator architecture to achieve automatic learning of image features and random changes of generated images, which can better control and understand the generated images by AdaIN [10] and truncation trick.

The proposal of a series of GANs has promoted the image-to-image translation work, and has gradually improved the generation results. The generation of multi-style Chinese art painting of flowers is also an image-to-image translation work. However, a few people have done the researches about the strong generation power of GANs. The one reason is that the training data is very difficult to collect. First, we must collect more traditional flower paintings involving obvious style characteristics from the internet. However, the techniques of Chinese art painting are complex, that is, there are many styles, which induce one image may contain several different styles. Second, the Chinese art paintings often have high resolutions and the size is large, which are not easy to train in the network. Third, the object that we train is mainly the flower, while most Chinese art paintings contain many other objects, such as animals, people and other landscapes, and we need to crop the images to capture flowers. Therefore, data collection and processing are difficult and complicated tasks. Another reason is the fundamental challenges from the freehand expressions. The classic style of traditional Chinese art painting includes line drawing, meticulous and ink, which have their own distinct characteristics. A good painting can express an emotion through content, texture, structure and colour as well as information characteristics. However, it requires that the painter has sufficient skills and the painter can be able to grasp the overall structural characteristics of painting and control the strength, which is challenging. The existing methods cannot solve our flower generation problem well.

To address these challenges and accurately generate meticulous and ink painting of flowers, we propose a Flower-GAN system to generate multi-style Chinese art painting of flowers. The Flower-GAN, consisting of attention-guided generator and discriminator, transfers the style among simple line drawing flowers, meticulous flowers, and ink flowers by an adversarial training way. In order to effectively learn the high-level semantic information and the representation of style features, our generator is designed as a vertical symmetrical structure, which contains nine resnet blocks. The patch discriminator discriminates the generated image on five scales. Moreover, in order to solve the problem of artefact and blur in image generation by existing methods, we further introduce a new loss function called Multi-Scale Structural Similarity (MSSSIM) to force the structure preservation. Combining it with the cycle consistency loss by adjusting the coefficient of two losses, we can control a reasonable generation of the structure and colour. Extensive experiments prove that our method produces more accurate and high-quality meticulous and ink paintings. We show the meticulous and ink paintings of the famous painters and our generated in Figure 2. Among them, in the first row, the second image is real meticulous painting, others are our generated. In the second row, the first and fourth images are real ink paintings, others are our generated.

The main contributions of our work are as follows:

1. We propose a Flower-GAN framework to generate multi-style Chinese art painting of flowers. This model provides an effective solution to the traditional Chinese art painting of flowers style generation task.
2. We introduce a new loss function called MSSSIM loss in the Flower-GAN system. The distance between the source image and the reconstructed image is compared on a multi-scale basis from three indicators, involving brightness, contrast and structure. The style consistency and content consistency can be accurately generated by combining it with cycle consistency loss, which solve the problem of bad results caused by previous methods.
3. We are the first to construct the unpaired flowers dataset containing three classic traditional painting styles: line drawing, meticulous and ink.

The rest of this paper is organised as follows: Section 2 presents the related work from three aspects while the details of our method are described in Section 3. Section 4 presents our used dataset and training details while several experiments are conducted and experimental results are analysed in Section 5. Finally, the conclusion and future work are discussed in Section 6.

2 | RELATED WORK

There are few researches on Chinese art painting of flowers generation. Here we introduce three aspects related to our work, that is, Chinese painting generation, image colourisation and image-to-image translation.

2.1 | Chinese painting generation

There have been many works researching on Chinese painting, especially on painting classification and generation [11–15].

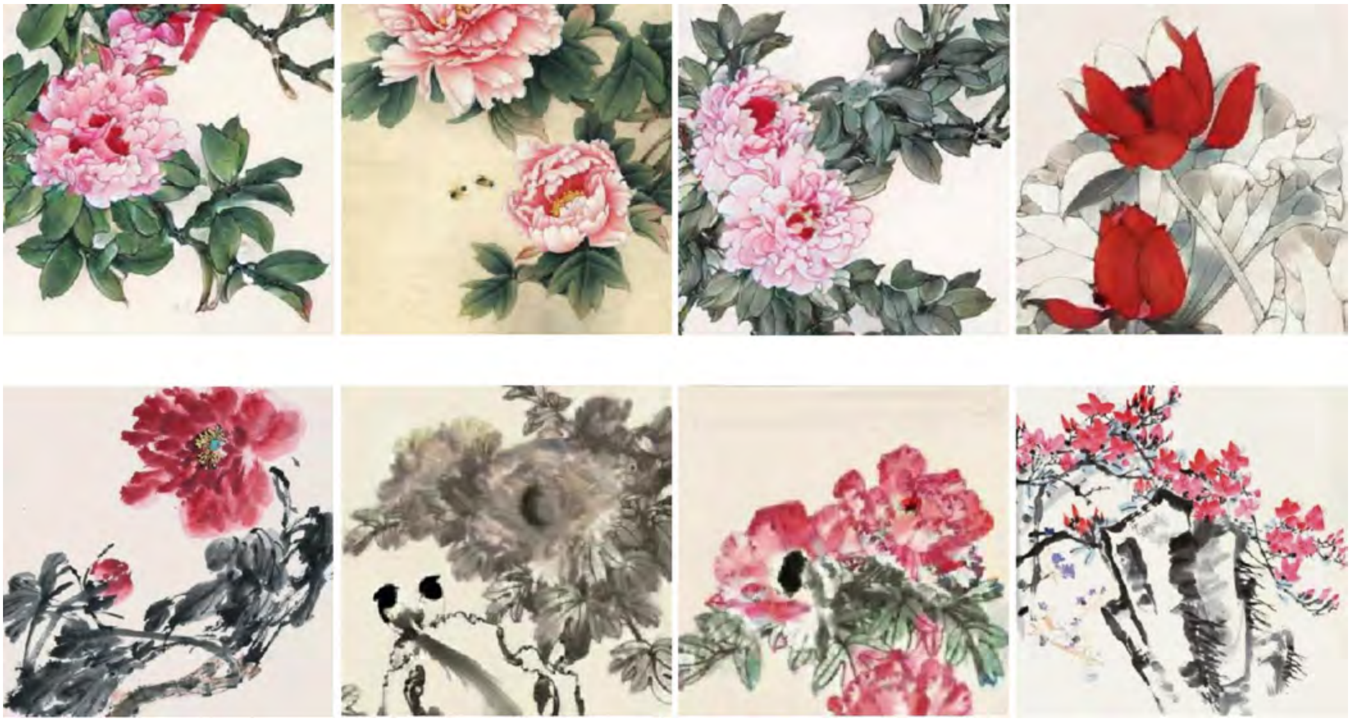


FIGURE 2 Illustration of meticulous (first row) and ink paintings (second row). Some of them are drawn by famous painting artists, and others are generated by the proposed approach. Guess which ones are generated by our method, and the answer will be announced in the introduction section

Earlier, Jiang et al. [16] propose to distinguish whether one image is Chinese painting or not using SVM, and make classification of meticulous and freehand brushwork. Sun et al. [17] propose to do the classification of painters using a sparse hybrid CNN according to the style of ink painting. Brush Stroke Texture Primitives (BSTP) [18] method is proposed to synthesise Chinese landscape painting from hand-made work. Dong et al. [19] propose a method to convert images into Chinese ink painting style by constructing its saliency map and non-physical ink diffusion. Recently, Tang et al. [20] propose a system to generate the animated drawing process of Chinese brush painting by focusing on stroke order. Easy drawing [21] proposes a method to generate artistic strokes in the style of flower painting by stylising rough sketch lines and makes good results. Although these works are attractive, the research on Chinese painting is still rough and far from being perfect because of the fundamental challenges of extracting the complex and abstract information of Chinese art painting. Some traditional methods cannot capture more important and complex semantic information because of the abstraction of Chinese art painting, and some methods just apply on small area paintings. For the rich Chinese paintings, we need to explore more in-depth researches on one specific aspect of Chinese painting.

2.2 | Image colourisation

With the development of deep learning, image colourisation has been greatly improved. In particular, black and white image colourisation [22–26] is the most widely used. Chybicki et al. [27] achieve a fully automatic method to colourise vintage cartoons using CNN model. The DeOldify project proposes a method for old photos colouring and restoration, also for old movie video colouring by using self-attention GAN [26] network. Besides, there are also some line-based colourisation methods [28, 29]. Style2Paints, as a powerful anime painting tool, use unsupervised GAN methods to dramatically improve the accuracy of colouring. CycleGAN [8] can also achieve the mutual transformation between pencil sketch and colour.

Our proposed work is similar to these works, while they are essentially different, mainly in the expression of dataset features and style features. In terms of data, the network input of black and white image colourisation is a grayscale image with just one channel, and there is a corresponding pair of colour images as ground truth. The data in our work is three channels, and there is no paired data used in unsupervised learning. In terms of style, such as cartoon, landscape and other colour images, the whole image is regarded as one. The colour is filled in the whole picture when generating, without obvious distinction between foreground and background, so it will not pay attention to the expression of specific style. Compared with traditional Chinese painting, the highlight lies in the description of main contents in the foreground, which shows a specific style. Comparing a classic work that is black and white image colourisation with our work, we enumerate the data of source domain and target domain used in both jobs, as shown in Figure 3. In Figure 3, we can clearly see the differences in data themselves and style control mentioned above. One work is image colourisation field,

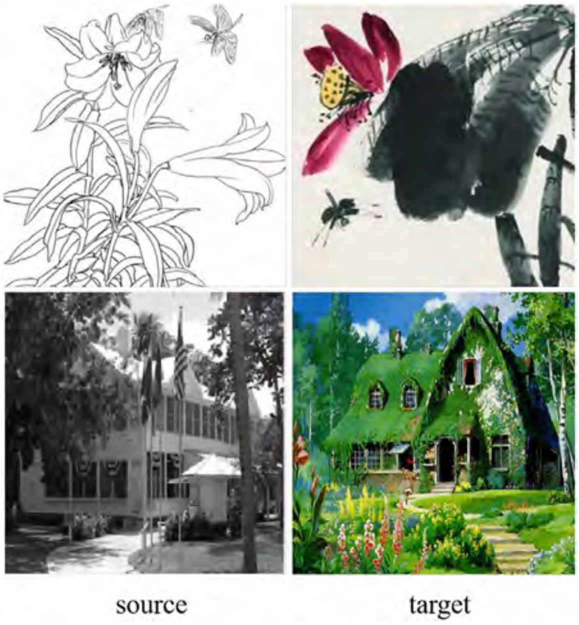


FIGURE 3 The source data and target data used in image colourisation and our work. Up row is our line drawing and ink painting data, and bottom row is grayscale image and its ground truth data for image colourisation

another is image translation field. The related experiments are shown in Section 5.2.

In contrast, our work is more inclined to sketch-to-image generation. The diversity (such as characters, landscapes, flowers and birds) and abstraction (such as line drawing, meticulous and ink) of traditional Chinese painting make Chinese painting itself own unique characteristics. For example, the ink paintings pay attention to the expression of ink shade and dryness. Structurally, Chinese painting pays attention to the local changes while focusing on the overall control. Therefore, it is difficult to simply colourise for the generation of Chinese art painting work.

2.3 | Image-to-image translation

There are many researches on image-to-image translation based on GANs. Here we introduce several classic methods on image-to-image translation.

CycleGAN [8] proposes the transformation between unpaired data with cycle consistency. It can be understood as bi-directional GAN [2] with two generators and two discriminators, which solves the problem that Pix2pix [7] must train data in pairs. The method can turn grey images into colour, images into semantic labels, stroke maps into photos, etc. However, the quality of the production is not very good. CycleGAN method cannot preserve the information well since cycle-consistency constraint makes the network more inclined to simple and easy generation.

UNIT [30] proposes the hypothesis of 'shared potential space'. First, using the encoder networks to map images to latent codes, different data share the same hidden space, which completes the high-level semantic sharing. Then the decoder networks are used to decode latent information to images. Combining VAE [31] and GAN [2], an unsupervised graph transformation network is constructed. For different domains, the image will generate different colours and specific details. This method can realise the transformation of different scenes, such as day and night transformation, season transformation and traffic sign transformation. Based on the extension of this method, MUNIT [32] is proposed. MUNIT further proposes the sharing of content space, with differences in style space to realise unsupervised transformation among multi-modal data.

AGGAN [33] is designed to use attention mechanism for image-to-image translation. The attention mechanism is added into the generator and discriminator, respectively, to focus on the attention area and control the generation of image semantics without changing the background. However, this model performs poorly for data of geometric structures. Building on ideas from these previous works, we develop an effective framework, named Flower-GAN, for multi-style Chinese flowers painting synthesis using attention mechanism. Compared with the previous methods, the proposed Flower-GAN can generate clearer images and mitigate the problem of artefact.

3 | FLOWER-GAN

The proposed task in this paper is to achieve the Chinese flower style transformation among three kinds of techniques of traditional Chinese painting, that is, line drawing, meticulous and ink. On one hand, this task can generate more abundant and interesting Chinese flower paintings that pass for authentic and provide a new thinking mode for modern painters. On the other hand, it makes up for the lacking of data for other related researches. In our preliminary experiments on line drawing-to-ink or line drawing-to-meticulous problem, we observe that the current unpaired image-to-image translation models fail to transform the style while retaining the content and geometric structure. After analysing the preliminary experimental results, we decide to adopt the attention mechanism to address this problem. Following we will describe the basic idea of our method in detail.

Our method is to learn the mutual mapping between the two domains, which can translate simple line drawing flowers into a meticulous or ink style, as well as achieve inverse transformation. Through several experiments, we find out that the style characteristics of the target domain are not well learned to the source domain. We study the data and analyse the model carefully to explore the reason. The one reason is abstractness of data freehand brushwork expression. It needs to maintain enough structure information and ensure proper texture generation after the style transformation. Specifically, the transformed ink paintings need exact composition, subtle colour, properly used ink (shade and dryness), and rich artistic conception. The transformed meticulous paintings need to be exquisite and delicate, rich in lines, and better with details, while the transformed line drawing needs to prevent images from the fragmentation, dullness and looseness and so on. Another reason is the study ability of model. For more abstract data, it is difficult for

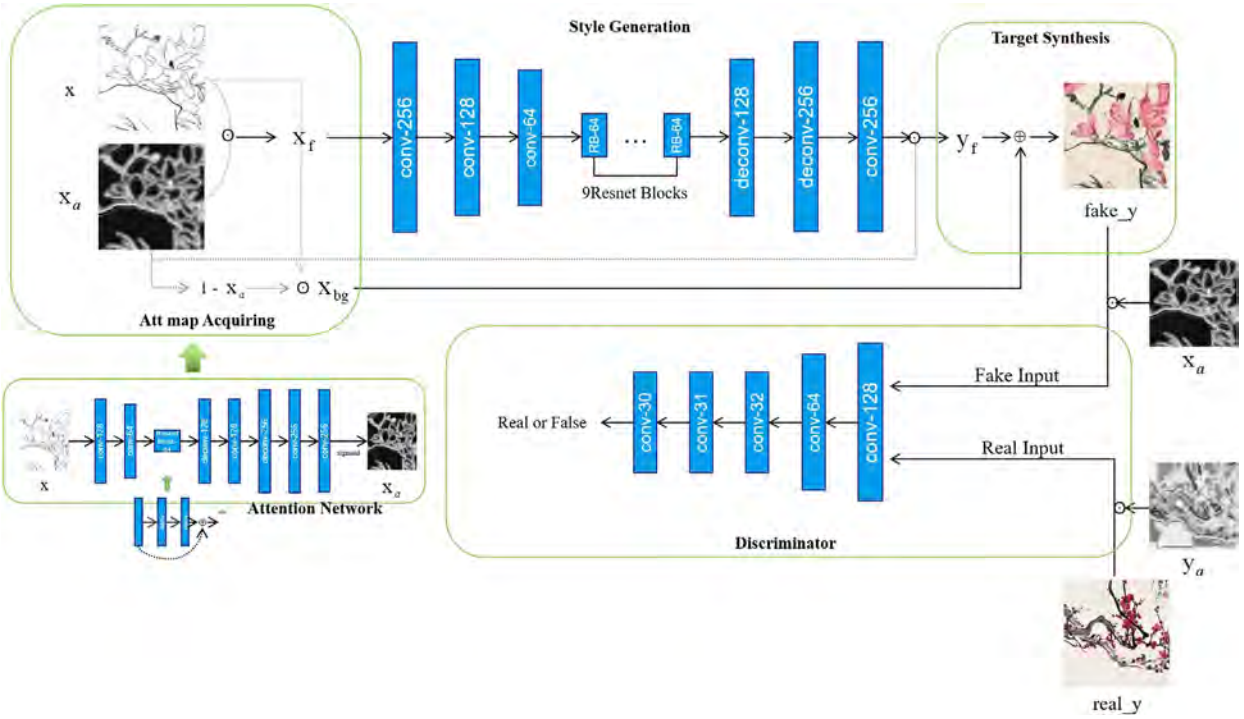


FIGURE 4 The model architecture of Flower-GAN system. Including the attention network A_X, the generator $G_{X\to Y}$ and corresponding discriminator D_Y networks in the forward mapping, the backward mapping is similar. The detailed information is described in Section 3.1

existing models to learn the structure and texture features of images well and often accompanied by artefacts. If the training is not good, the model may collapse. According to the above analysis, this task is full of challenges.

Working on these problems, we put forward to focus on the content of the image body. There are two kinds of thinking: (1) Based on the assumption of shared content space, the content feature and style feature can be extracted separately. (2) Using the attention mechanism to enable the network to focus on the important areas for generation. By extracting the attention map of the source image and inputting it to the generator network, the region of interest can be better mapped and the generation can be more correct.

Therefore, we propose a system called Flower-GAN based on CycleGAN [8] that specially designed to make Chinese art painting of flowers translation work, as shown in Figure 4. The generator G is responsible for generating a fake image with the target domain style to deceive the discriminator D, while the discriminator D is trained to determine true or false of the image, and finally achieves a Nash balance during the mutual confrontation process. Moreover, due to the proposed attention mechanism and loss function, our approach generates high-quality images.

3.1 | Network architecture

In this subsection, we describe the proposed network architecture in detail. As shown in Figure 4, in the forward direction, our model consists of one generator, one discriminator and one attention network. The backward direction also includes these three structures. Particularly, in this paper, X and Y denote the line drawing and ink or meticulous domains, respectively.

The Flower-GAN model learns a forward and backward mapping simultaneously between domains X and Y given unpaired training samples $x \in X$ and $y \in Y$. Except for the generators and discriminators, we add two attention networks $A_X : X \rightarrow X_a$ and $A_y : Y \rightarrow Y_a$ which learn the attention map X_a of X and attention map Y_a of Y, respectively. The attention mechanism is the main idea of our approach, which has been successfully applied to many models [26, 33]. With the attention map, the network achieves the image transformation that only focuses on the foreground area and unchanged background. The attention network architecture is similar to auto-encoder, which is used to extract the main structural features of the interested area, separate the foreground from the background and highlight the foreground structure. Input the original image x of the resolution 256 × 256 pixels, and output its attention map x_a of the same resolution, but the value of each pixel of x_a is [0,1] by the sigmoid activation function. The attention network has convergence, and the map edge features learned through the network perform better.

The generator network has a vertically symmetric structure, which consists of three parts: encoder, resnet blocks and decoder. Specifically, the encoder part consists of three convolutions followed by ReLU: the first layer having 7 × 7 kernel, the next two having 3 × 3 kernel. Then, we adopt nine resnet blocks where each block consists of two 3 × 3 convolution layers, in order to extract more abstract features. In the experiments, it is found that whether the quantity of blocks decreases

ALGORITHM 1 The whole procedure of Chinese art painting of flowers synthesis

Input: The line drawing image x

Output:

1Extract the foreground attention map x_a of source image x by using the attention network.

2The foreground attention map x_a is calculated by dot product with the source image x to obtain the foreground image x_f of the source image.

3The foreground image x_f of the source image is fed into the style generation network for training. The foreground image y_f of the target domain is obtained by the dot product calculation of the network output and foreground attention map x_a of the source image.

4The background image x_{bg} of the source image is obtained by calculating the dot product between the source image x and the background attention map $(1 - x_a)$ of source image.

5Combing the obtained foreground image y_f of the target domain in step 3 and the background image x_{bg} of the source domain to synthesise the final fake target domain image.

Output: The transformed style image *fake_y*.

or increases, the learning ability of network will be declined. The decoder part consists of two upsampling layers having 3 × 3 kernel followed by ReLU. Finally, one convolution layer with 7 × 7 kernel not followed by ReLU is used to restore the original image channel.

The discriminator network is a 70 × 70 patchGAN which is a popular and successful application in many image-to-image translation models [7, 8]. It consists of five convolution layers all using 4 × 4 kernel. The dimension of output feature map for each layer is 256→ 128 → 64 → 32 → 31 →30.

Our entire generation process consists of three main processes: attention map acquiring, style generation and target synthesis, as shown in Figure 4. Algorithm 1 summarises the whole procedure of Chinese painting synthesis and Algorithm 2 outlines the training procedure of the proposed Flower-GAN. In the first step, get the foreground and background of the image for subsequent network input. Input image x to the attention network to obtain the attention map x_a. In this case, the foreground image x_f and the background image x_{bg} are computed:

$$x_f = x \odot x_a, \tag{1a}$$

$$x_{bg} = x \odot (1 - x_a). \tag{1b}$$

In the second step, achieve the image style transformation work. We feed the interested area, that is, the foreground x_f into the encoder part, the dimension of feature map is reduced to 64 × 64. The extracted features from encoder part are then fed into resnet blocks to further extract important features. After that, through upsampling operation and last one convolution layer, the network outputs three channel transformed image. At last, compose the final translated image. We get the translated foreground image y_f with target domain style using the output of previous step by numerical operation:

$$y_f = G_{X \to Y}(x_f) \odot x_a. \tag{2}$$

ALGORITHM 2 Training procedure of the Flower-GAN framework. We train it in two stages using Adam [34] optimiser with an initial learning rate 0.0001 and linearly decay the rate after the half training iterations. The numbers of steps to apply to the Attention network, generator and discriminator network

```
Input:
    $X$: the line drawing image dataset,
    $Y$: the ink painting image or meticulous image dataset,
    $s, t$: the switch parameter and the small threshold constant,
    $\lambda_1, \lambda_2$: the hyper-parameters,
    $N, bs$: the training iterations and batch_size, respectively.
for $it = 1, \ldots, N$ epochs do
    Step One: training the Attention (Att) network and generator (G), discriminator (D).
    for $it = 1, \ldots, s$ do
        Sample batch of $bs$ examples $(x, y)$ from $X, Y$.
        $x_a \leftarrow Att(x; \theta_1)$ # generated attention map
        $x_f \leftarrow x \odot x_a$ # foreground of source image
        $y_f \leftarrow G(x_f; \theta_2) \odot x_a$ # generated foreground of target domain
        $x_{bg} \leftarrow x \odot (1 - x_a)$ # background of source image
        $fake\_y \leftarrow add(y_f, x_{bg})$ # generated fake target domain image
        $O_f \leftarrow D(fake\_y; \varphi_1)$ # judgement score of fake image
        $O_r \leftarrow D(y; \varphi_2)$ # judgement score of real image
end for
    Step Two: training generator(G), discriminator(D).
for $it = s, \ldots, N$ do
        $y_f \leftarrow G(x \odot x_a; \theta_2) \odot x_a$
        $x_{bg} \leftarrow x \odot (1 - x_a)$
        $fake\_y \leftarrow add(y_f, x_{bg})$
        $mask \leftarrow f(Att(x), t)$ # threshold processing of fake image
        $O_f \leftarrow D(fake\_y, mask; \varphi_1)$ # judgement score of masked fake image
        $mask\_ \leftarrow f(Att(y), t)$ # threshold processing of real image
        $O_r \leftarrow D(y, mask\_; \varphi_2)$ # judgement score of masked real image
end for
        Compute the Adversarial loss $L_{GAN}$ using Equation (4),
        Compute the Cycle consistency loss $L_{cyc}$ using Equation (5),
        Compute the MSSSIM loss $L_{ms}$ using Equation (6).
        $\theta_1, \theta_2, \varphi_1, \varphi_2 \leftarrow Adam(\nabla_{\theta_1, \theta_2, \varphi_1, \varphi_2}[L_{GAN} + \lambda_1 L_{cyc} + \lambda_2 L_{ms}])$
end for
```

Finally the foreground y_f is combined with the background x_{bg} to produce the final fake image *fake_y*:

$$fake\_y = y_f + x_{bg}. \tag{3}$$

3.2 | Loss Function

Our loss function consists of three components, that is, adversarial loss, cycle consistency loss and MSSSIM loss. Instead of using sigmoid cross entropy as our GAN objective, we use the least squares GAN objective for stable training.

The adversarial loss is employed to match the distribution of generated fake image to target real image. For the mapping $G_{X\to Y}$, its attention network A_X and its discriminator D_Y, the adversarial loss is defined as

$$\begin{aligned} L_{GAN}(G_{X\to Y}, A_X, D_Y) &= E_y[\log D_Y(y)] \\ &\quad + E_x[1 - \log D_Y(G_{X\to Y}(x))]. \end{aligned} \tag{4}$$

The cycle consistency loss is pivotal for unpaired data between two mappings, which is employed to ensure the generated fake image can be again mapped back to the source image by the backward mapping. The use of cycle consistency loss can effectively control the spatial distribution of the image generation, allow the network to retain contour features and avoid structural difference between the generated image and the source image. We use ℓ_1 distance for cycle consistency loss:

$$\begin{aligned} L_{\text{cyc}} &= \| G_{Y\to X}(G_{X\to Y}(x)) - x \|_1 \\ &\quad + \| G_{X\to Y}(G_{Y\to X}(y)) - y \|_1 . \end{aligned} \tag{5}$$

The MSSSIM loss is employed to compare two images on multiple scales from three indicators of Luminance L, Contrast C and Structure S. Compared to SSIM [35] function, MSSSIM is more convenient and effective to incorporate the details of image at different resolutions. Making good use of MSSSIM loss can achieve effective improvement in image quality. For the source image x and the reconstructed image x'', it is formulated as

$$L_{ms}(x, x'') = f(l_M(x, x''), c_M(x, x''), s_M(x, x'')). \tag{6}$$

It is a constituent function of these three indicators. The weight coefficients for each component constitutes their mapping relationship f. Under each indicator, the difference of two images is calculated and compared separately. In Equation (6), M represents multiple scales. M is equal to 1 means the size of source image, and is equal to 2 means the size of source image is reduced by half. For the detailed introduction of MSSSIM function, refer to [35, 36].

Based on above contents, the total objective loss L is formulated as

$$\begin{aligned} L &= L_{GAN} + \lambda_1 L_{\text{cyc}} + \lambda_2 L_{ms} \\ &= L_{GAN}(G_{X\to Y}, A_X, D_Y) + L_{GAN}(G_{Y\to X}, A_Y, D_X) \\ &\quad + \lambda_1 L_{\text{cyc}} + \lambda_2 (L_{ms}(x, x'') + L_{ms}(y, y'')), \end{aligned} \tag{7}$$

where λ_1, λ_2 are the weight parameters and y'' is the reconstructed image of y. The proportion of weight parameters has an important influence on the quality of experimental results. In order to compare the influence, we make several sets of comparative experiments. The experiments prove that the best effect is obtained when $\lambda_1 = 10, \lambda_2 = 1$. The combination of three losses results in accurate and higher quality images in our multi-style flowers dataset.

4 | DATASET

4.1 | Dataset collection

Chinese flowers painting dataset is a collection of three traditional Chinese painting style, that is, line drawing, meticulous and ink. The primary feature of the dataset is that the flower paintings involve obvious style characteristics. However, the number of images that are close to the traditional ink and brush styles is few on the internet, and even one image may contain several different styles, so it is not easy to collect images with only one unique style. Therefore, the collection of high quality dataset is a hard and time-consuming work. In this case, we collect these three datasets using line drawing, meticulous, ink and flowers four keywords through various channels, including the internet and WeChat Subscription. Then, we pick out the images that match style characteristics. Since the images are not required to be paired, we can ignore that when picking. Finally, we totally collect more than 600 images in each category.

4.2 | Dataset preprocess

In order to train the proposed network well, we need enough high-quality data. However, there are exactly two problems with the data we have collected. One problem is small dataset and another is images which cannot proceed directly for training. The images exist lots of noise for network training, such as large blank space, high resolution, a variety of objects and styles. If the data is directly fed into feature extraction network, there will be many unnecessary interference factors that are not conductive to feature extraction. In this way, even with a good network model, the final learning effect will be poor due to the quality of dataset. Therefore, we carry out three pre-processing operations on the dataset. First, the data from different sources comes in various formats and different resolutions, and generally Chinese paintings are large, so we convert all the flower images into one format. Then, the length and width of images with large resolution are resized to an appropriate size. And save the images in the form of Image.ANTIALIAS where the save quality is 75. Second, we enhance the data. We crop the image 6 times with the crop function by setting the length and width of six boxes following the pattern of up, down, left, right and middle. Although the traditional random cutting method is easy to deal with this problem, the effect is no more effective than our cropping in the box method. Third, considering that our task is to generate for flowers, however, in addition to the flowers, there are many other objects such as people and animals in the painting. For this purpose, we manually select high-quality images where the content of flowers has the largest density and appropriate area. Finally, we train and test the network performance with processed dataset. TABLE 1 shows the training set

TABLE 1 The number of images in the training set and testing set in each style

Style type	Training set	Testing set
Line drawing	1537	489
Ink	1560	442
Meticulous	1536	444

and testing set we used for line drawing, meticulous and ink three styles.

5 | EXPERIMENTS AND ANALYSIS

In this section, we describe our experiments on the Chinese painting of flowers dataset. First, we describe the implementation details in Section 5.1. Then, we show the qualitative results in Section 5.2 and evaluation metrics are presented in Section 5.3. In Section 5.4, we make a model analysis in detail and ablation experiments are conducted in Section 5.5. We further explore our model and carry out the extended experiments in Section 5.6. Finally, we make an user study in Section 5.7. Specifically, the following experiments demonstrate the effectiveness of our method and a detailed analysis of the method.

5.1 | Implementation details

We implement our Flower-GAN model with the use of TensorFlow as backend. During training, we do not need paired data but two domain data. For the dataset, we use line drawing images as the source domain, and the meticulous or ink images as the target domain. In our experiments, the training data and testing data are showed in TABLE 1. Each batch of training images is randomly matched from two unpaired datasets. The amount of training or testing input into the network can be aligned by adjusting the number of training or testing dataset according to our actual need, to ensure that each input image has an image of the corresponding domain. All training images (i.e. line drawing, meticulous and ink) are resized to 256 × 256 pixels.

To prevent the model collapse, we follow the idea of [33] to train the network in two stages. The detailed training procedure is shown in Algorithm 2. In the first stage, we train the attention network, generator and discriminator for 30 epochs. In the second stage, we stop the training of attention network and use the attended regions of image to train the discriminator by setting the threshold value equal to 0.1. It is worth mentioning that we use the sigmoid cross-entropy function and least-squares function as the adversarial loss, respectively, for comparison. The experiment proves that the least-squares loss is more effective in flower images style generation. For cycle consistency loss, we use ℓ_1 regularisation instead of ℓ_2 regularisation. We train our model for 100 epochs using Adam [34] optimiser with the batch size of 1. In the half of training epochs, we set the learning rate as 0.0001 and linearly decay the rate over the next epochs. For the loss hyper-parameters, λ_1 and λ_2 are set to 1 and 10, respectively.

5.2 | Comparison experiment

Ink painting generation. We firstly conduct experiments on ink painting of flowers generation. We compare our method with several classical models of unsupervised image-to-image translation including CycleGAN [8], UNIT [30], AGGAN [33], as shown in Figure 5. In Figure 5, we can clearly see that compared with other methods, our proposed method achieves the best result, and produces the accurate and high-quality images. In the following, we will carefully explain the shortcomings of each method in the experiment.

CycleGAN [8] fails to capture the style features well. As shown in Figure 5, on the whole, the results of second column do not look like a complete art work, which are very vague and exist lots of noise. On the local, CycleGAN has learned some structural characteristics of the object, but the details are not handled properly, which causes that the images lack a lot of important details, such as petals and leaves. Moreover, the edge features are not studied well. The edges of object seem to adhere to the background, giving people a false impression. As the most important visual component of style transformation, that is, colour features, which are rarely captured, such as the expression of flower pistil features and the expression of bird wings. The generated results are still far from satisfactory.

Based on shared hidden space assumption, UNIT [30] encourages style transformation. However, the results are still not ideal. Due to lacking part of structural information and semantic information not learned well, we can see that all objects in the images appear to have the same colour distribution, including petals, leaves, branches and birds. The style generation does not grasp the strength and skills of using ink, which causes that the representation of ink intensity and dryness are disharmonious. And the composition of images is not exact enough to form a heavy feeling.

Compared with the two previous methods, AGGAN [33] has certain improvement. However, there are also clear deficiencies. From the results of fourth column in Figure 5, we can see that some style features are incorporated into the content space. There are some clear noise and visual artefacts. Although AGGAN has learned the expression of few abstract brushwork features, the network learning is not stable, which causes that most images cannot deal with the representation of colour features in structural details. The important reason is that this model cannot learn the accurate structural details well.

Considering the similarity between our work and image colourisation, we also use the colourisation method recently provided in DeepCartoonColorizer (DCC) [27] to compare with our method. Following the steps of processing data in DCC, first, we convert the input images to Lab colour space. Then, we extract the L channel and a, b channel of Lab colour space, separate them into two parts as the input data X and ground truth Y. At this time, we begin the network training. When testing, we decolourise the testing data to get the

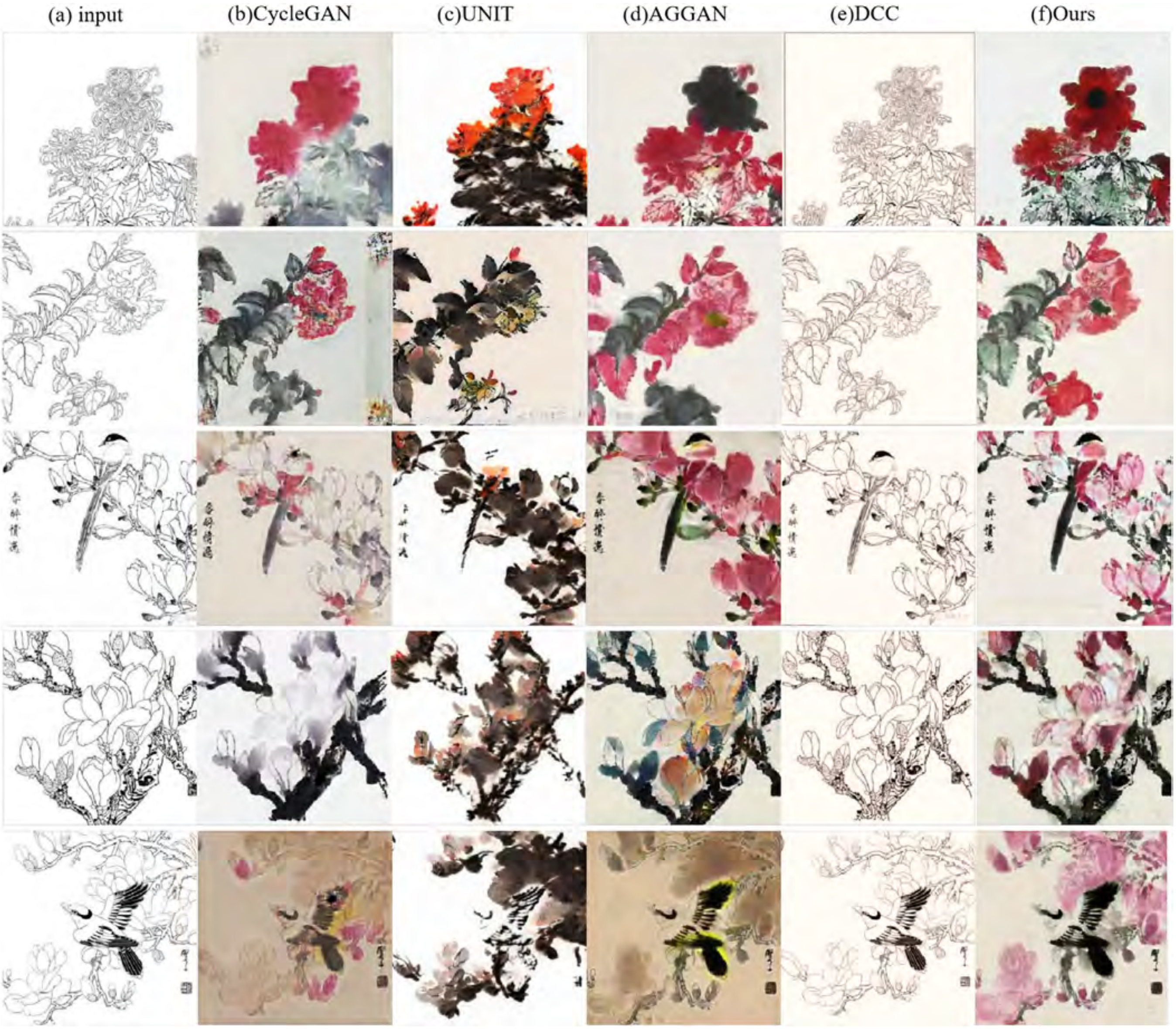


FIGURE 5 Comparison results of ink painting with CycleGAN [8], UNIT [30], AGGAN [33], DCC [27] and our Flower-GAN. We can see that our method generates more accurate images and mitigates the problem of artefact

grayscale images as input and recolourise them with the trained network. We carry out the experiment on the ink painting dataset and use the line drawing dataset to test. The experimental results are shown in the (e) column of Figure 5. From the results, we can see that the DCC method fails to colourise our data. The DCC model cannot capture the important colour features to achieve colourisation. As the image colourisation method, DCC emphasises the importance of pixel position. The image is filled with the corresponding position colour obtained by convolutional network. This method determines the quality of colourised results by the ground truth. However, there is no ground truth to correspond to our data. Combined with the views mentioned in Section 2.2, it proves that it is unsuitable to solve our Chinese painting of flowers generation by simply colourisation using the CNN model.

In comparison, our method produces clear and realistic stylised images regardless of the foreground or the background of images, as shown in the (f) column of Figure 5. It allows the images not only to preserve the content and structural features, but also the stylistic features to be well expressed, generating better detail and texture. The attention mechanism prompts the network to focus on the generation of flowers. The application of MSSSIM loss promotes the network to pay more attention to the representation of structural information and texture features. The effective combination of loss function makes the model more specific in style representation, and makes the generation of detailed features better. All experiments prove that our method can better learn the style characteristics of the target domain and generate accurately.

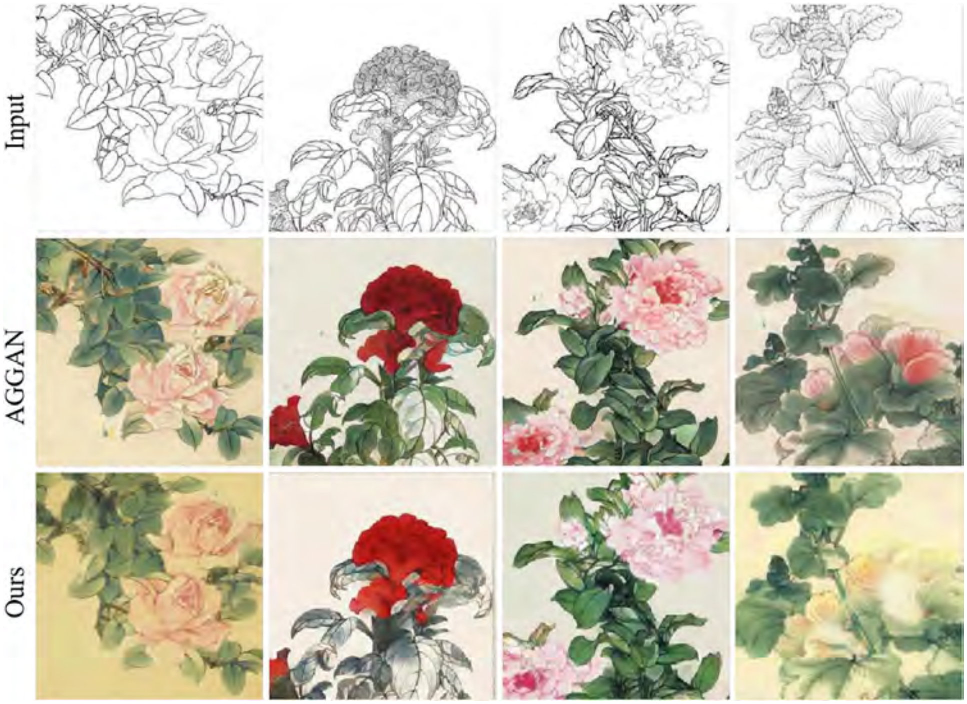


FIGURE 6 The generated meticulous painting of flowers from line drawing painting of flowers by AGGAN [33] and our approach

Meticulous painting generation. Based on the trained model, we conduct the experiments from line drawing painting to meticulous painting, and compared with AGGAN [33] which performs better in the ink style transformation experiments, as shown in Figure 6. The meticulous style pays more attention to the representation of strokes. From Figure 6, it is clear that AGGAN temps to produce lots of noises. Comparing with AGGAN, we get the fine exquisite meticulous brushwork flowers, which have real colour, abundant texture and precise strokes as line drawing. We can see that the generated images are almost close to the real, not only successfully deceive the discriminator, but also escape human visual inspection.

Line drawing generation. In addition, translated line drawing paintings resulted from ink painting images are also demonstrated in Figure 7. In the regions which need more texture and details, the generated results by CycleGAN and AGGAN always produce blur or distortion, AGGAN still ignores the expression of some texture features. In the regions where the feature of geometric structure changes rapidly, CycleGAN and AGGAN have weak learning ability, fail to accurately acquire such structural features for good transformation and form visual artefacts. However, our method can get rid of these problems by introducing the MSSSIM function, which performs better in details and structures.

Moreover, to better demonstrate the effectiveness of our Flower-GAN model, we show the close-up views of an image generated by the proposed method, as shown in Figure 8. The colours of ink paintings pay attention to the expression of the intensity and dryness of ink, namely the harmony of freehand brushwork. The experimental results indicate that our model performs well in capturing the exact colour features. In addition, we use the harmony as a measure in the user study. The details are shown in Section 5.7.

5.3 | Experiment on generated images quality

The peak signal-to-noise ratio (PSNR) is the most common all-reference image quality evaluation metric, which is based on the error between the current image and the corresponding pixel points of the reference image, that is, the image quality evaluation based on the error sensitivity. PSNR is often used in image colourisation tasks. Since there is no ground truth to compare with the generated images, the PSNR evaluation metric is not applicable for our work. Inception score (IS) [37] and Frechet Inception Distance (FID) [38] are two of the most popular evaluation metrics for the generative model to measure the quality and diversity of generated images. IS computes KL-Divergence between conditional and marginal label distributions over generated samples. A high score of IS indicates that the model generates the images with higher quality and diversity. FID computes Wasserstein-2 distance between generated samples and real samples in the feature space. A lower score of FID indicates that the generated images have higher quality and diversity. However, IS only considers the generated samples without considering the real data, that is, IS cannot reflect the distance between the real data and the sample. Therefore, the reliability of FID as an evaluation metric is higher than that of IS. In our

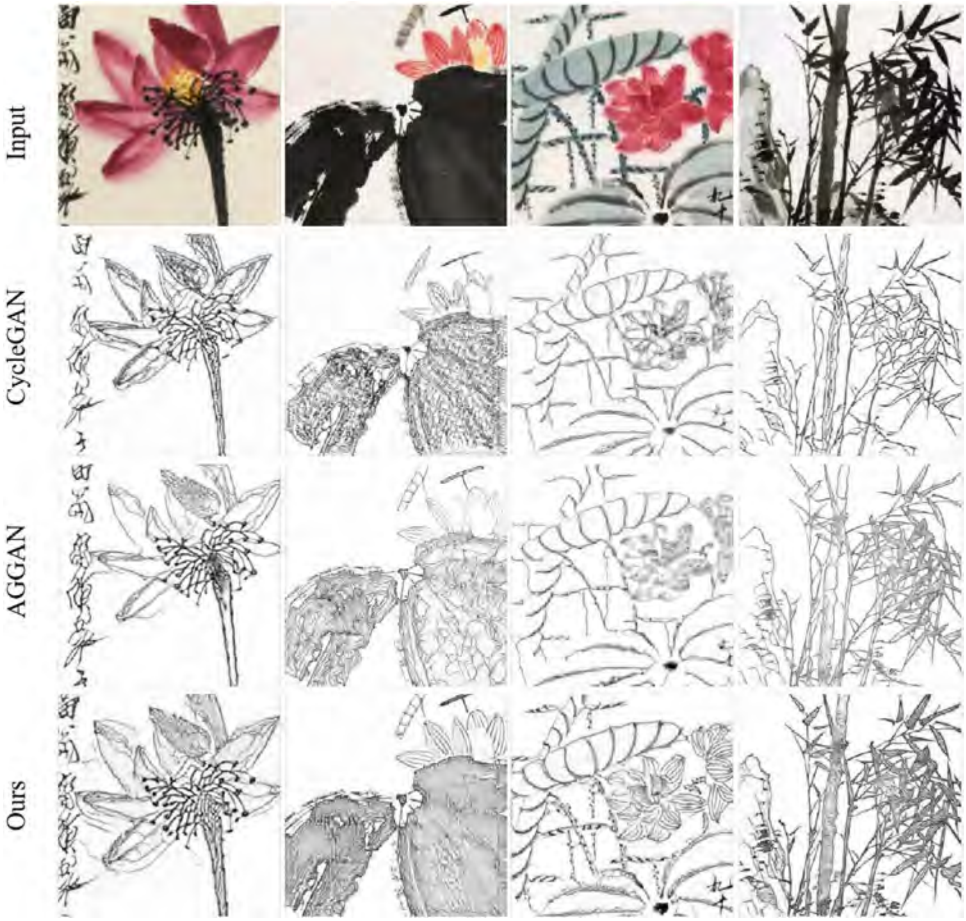


FIGURE 7 The generated line drawing paintings from ink paintings through the backward mapping

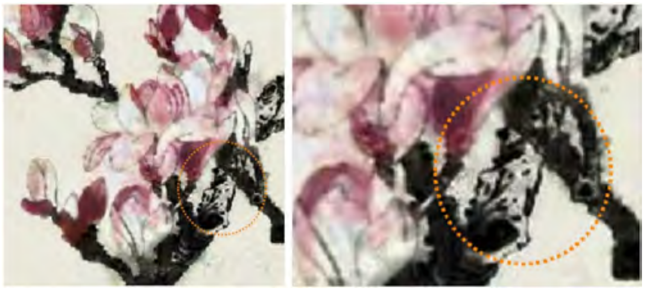


FIGURE 8 The generated image of our method better reflects the shade and dryness of the ink

TABLE 2 Comparison of CycleGAN, UNIT, AGGAN and ours using FID and IS metrics

Methods	FID↓	IS↑
CycleGAN [8]	1.384	(3.175,0.342)
UNIT [30]	1.984	(3.309,0.648)
AGGAN [33]	1.285	(3.756,0.495)
Ours	1.272	(3.355,0.312)

work, we use FID as the main basis for judgement and IS as a reference.

For quantitative evaluation, we compare our Flower-GAN with other baselines including CycleGAN [8], UNIT [30], and AGGAN [33] using the FID and IS. We evaluate the FID between generated images and real images and evaluate the IS on the full test generated images. Note that the choice of matching between the generated images and the real images is random when evaluating FID. The comparison results are presented in TABLE 2. From the results, we can see that our method has a lowest FID value, indicating our generating distribution is closer to the real distribution than other methods. Comparing the reference metric IS, it is clear that our method gets a lower value than AGGAN method but higher than other two methods. In a comprehensive comparison, the comparison results show that the generated samples by our approach are of higher quality and more diversity.

5.4 | Parameter sensitivity

In this work, there are several hyper-parameters to be fine-tuned. Specially, we do the parameter sensitivity experiments on the switch parameter epoch number s in Algorithm 2, the training batch size bs, and the weight parameter λ_2 in Equation (7).

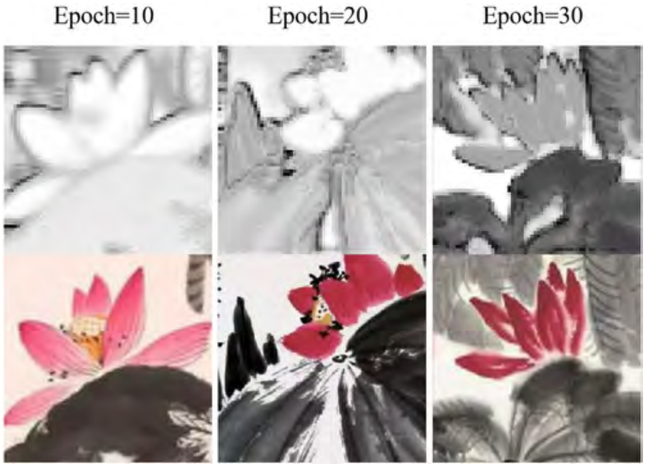


FIGURE 9 The changes of attention map as epoch increases

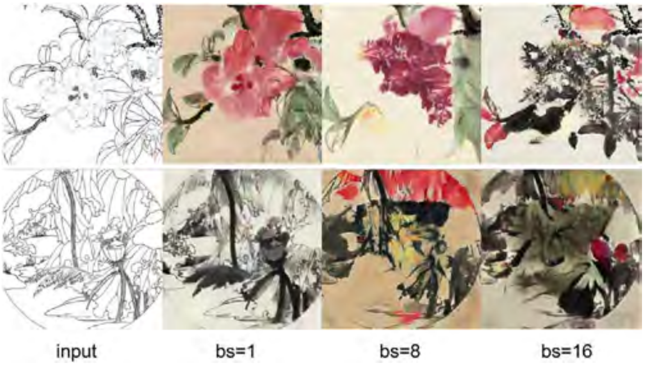


FIGURE 10 The effect of batch size (*bs*) factor. It is clear that the result is best when batchsize is equal to 1

As suggested in [33], the hyper-parameter λ_1 is set as 10 in all experiments. In the following, the parameter sensitivity experiments are described in detail.

Parameter sensitivity on epoch number. We show the performance of attention network on the Ink painting dataset by controlling the switch parameter s, that is, the epoch size. At the same time, we set the batch size bs and the weight parameter λ_2 to 1, respectively. As shown in Figure 9, we can observe that the performance of attention network to extract image features is enhanced gradually with the increase of epochs. When the network is trained to about 30 epochs, it can accurately extract the structure of image foreground region. The experimental result indicates that when the switch parameter s is equal to 30, we can stop the training of attention network in our entire training procedure and we only use the extracted attended regions from the attention network to participate in the following style transformation experiment.

Parameter sensitivity on batch size. We show the influence of batch size on style transformation experiments over the parameter bs within the range {1,8,16} while setting the parameter s to 30 and λ_2 to 1. The source images we used are line drawing paintings and the target images are ink paintings. The experimental results are shown in Figure 10. In Figure 10, we can observe that the result is preferable when the parameter bs equals to 1. With the gradual of bs, the performance of Flower-GAN model decreases. The structural details are getting blurred and even lost. The images lack lots of important semantic information and become very messy and distorted. Therefore, we set the parameter bs as 1 in all experiments by analysing the influence of different batch size on experimental results.

Parameter sensitivity on trade-off weights. We show the effect of MSSSIM function on line drawing-to-ink style translation experiments over the weight parameter λ_2 within the range {0.1,0.5,1,5} while fixing the cycle consistency parameter λ_1 of 10. The quantitative results are demonstrated in Figure 11. First of all, we can observe that there is no image result with the parameter of 5. The reason is that the generator is difficult to learn when λ_2 is equal to 5. Our model fails to capture the representation of image features. From the results, it is clear that when the parameter equals to 0.1, the composition colour of image is very bright and dazzling, as shown in the second column of Figure 11. When the parameter is gradually increased, the image colour is more natural and harmonious. In addition, many images have noise and artefacts when the parameters are small. When the parameter equals to 1, the structure information is better learned. The images are more consistent. The stitching of flowers and leaves are not deliberate. Flower-GAN model is relatively better generation in terms of details and overall control. The experimental results indicate that proper MSSSIM parameter λ_2 would help the model capture the exact colour features and force the structure preservation.

5.5 | Ablation experiments

We perform the ablation experiments to study the role of each part in Flower-GAN. We train all experiments with batch size of 1, using Adam [34] optimiser with a variable learning rate whose initial value is 0.0001. Figure 12 shows the examples of ablations in our full loss function, in which all results are trained on ink painting style transformation dataset. The following results show that each component plays an important role in Flower-GAN. First, the initialisation phase, that is, the first stage of network training, which helps the attention network converges to a reasonable range, thereby avoiding model collapse. Without initialisation, since the attention network has not been trained in the initial stage, it will guide the discriminator to make a wrong judgement when training with the GAN network at the same time. This situation is very easy to cause network learning instability, and the model cannot effectively obtain the characteristics of the attended region so that it fails to generate. Second, in terms of image generation, ℓ_2 regularisation often produces a blurry effect. For the shape, colour and other characteristics of the image, ℓ_2 will choose a balance mode to get a generalised image, while ℓ_1 regularisation will avoid this problem and help deal with substantial style difference between the stylised image and source image. Lastly, the MSSSIM function guides the

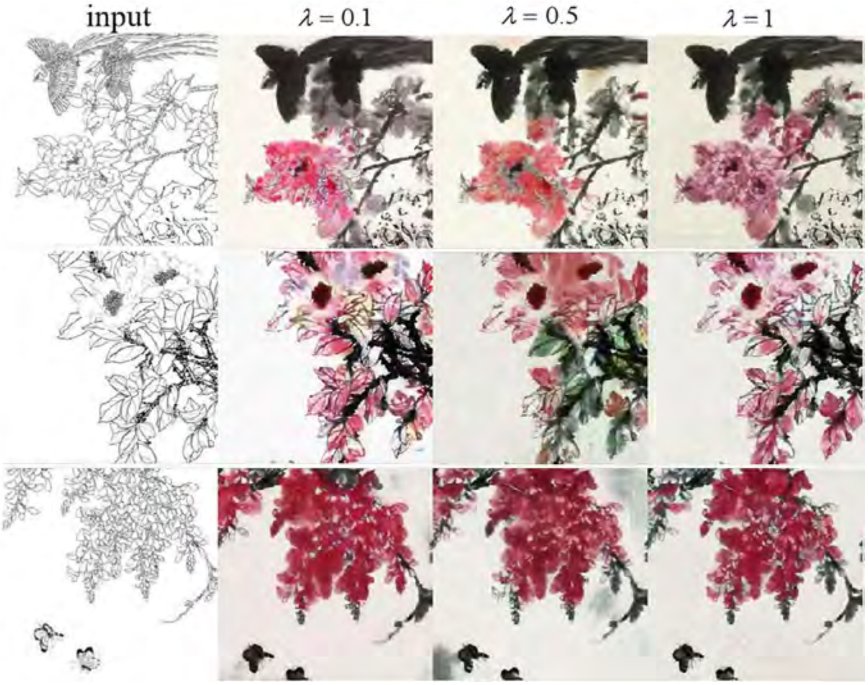


FIGURE 11 The generated results when changing the weight coefficient of MSSSIM loss

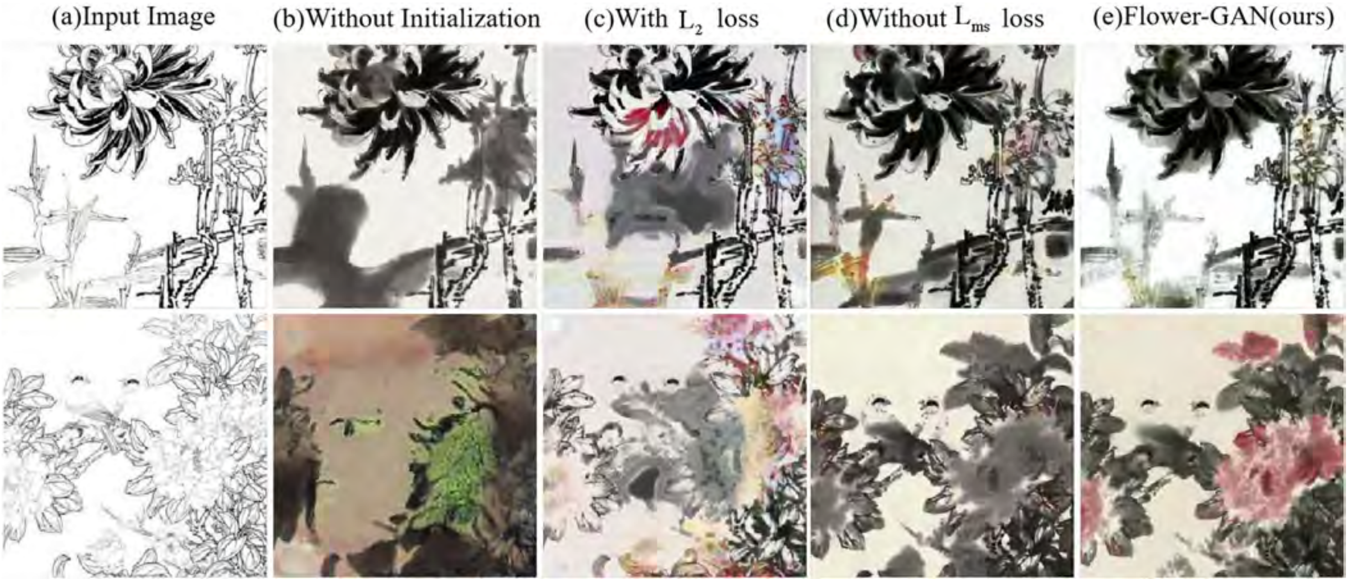


FIGURE 12 Results of removing/changing components in the loss function of Flower-GAN

network to pay attention to the preservation of colour and structure information, so that the images can be generated better in detail.

Through a series of experiments on the weight parameter of MSSSIM function, we can observe that MSSSIM may be more sluggish in brightness and colour; however, it can hold high frequency information better. Although ℓ_1 function may preferably maintain colour luminance characteristics, thus we use them in combination. This added loss makes our network perform very well in two ways: First of all, it focuses on the local and overall structural similarity of the content. On one hand, it pushes the attention regions can learn the local features better. On the other hand, it makes the overall style characteristics close to the distribution of the style features of the target domain as much as possible, including the background. Second, it induces colour generation to conform to the natural law characteristics. It forces the generation of the attention regions to pay attention to the morphological colour change of the object in the training data, so that the generated images have exact composition and true colour.

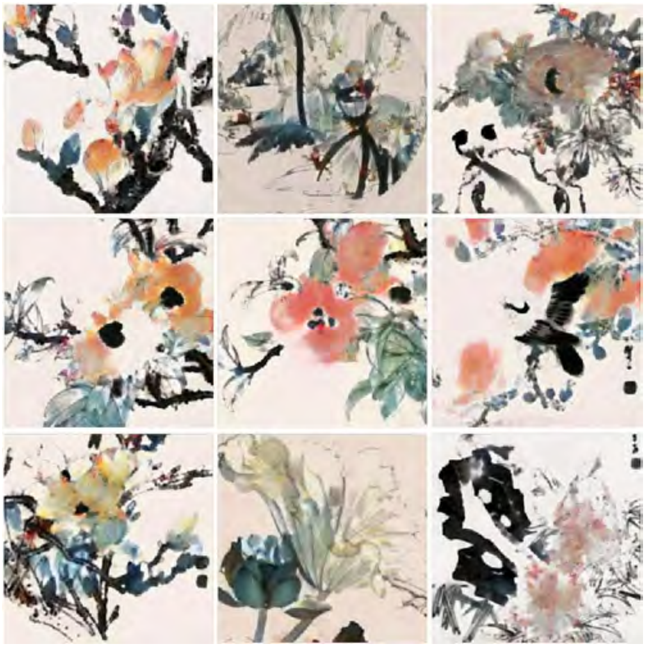


FIGURE 13 The generated ink style images through morphological manipulation of training style dataset

5.6 | Extended experiments

With the experiments on ink painting of flowers style transformation, we find that although we have generated accurate and high quality images, the results are not very abstract, and the artistic conception is not as rich as we expected. Through observation, we think that part of the reason is probably the problem of dataset. Most of the images used in the training set are not very abstract, and the resolution of the processed images are not large enough, which cause that we can only produce a similar effect with the dataset.

In order to make the generated images more full of artistic conception, we perform morphological processing operation on the ink painting of flowers dataset. The operation could remove high frequency content, blur the edges and eliminate noise in the image. We show the generated results in Figure 13. From the result, we can see more colour distribution in the images than in previous experimental results. The brightness is increased, but the picture is blurred. In addition, we show the test results to students and teachers in our lab, and ask them to comment whether the resulting images are more abstract and artistic than previous results. In the end, most students maintain that the results are not as good as before. They think the images are blurred or distorted, and the generated colour is not exact. A few students are the opposite. They believe that it is this variety of colours and the semantic incompleteness that make the image more abstract and artistic.

Moreover, to prove the scalability of our method, we conduct experiments on the landscape dataset. We first collect landscapes pictures in nature and extract their edges using canny edge detection method, as training set A. Then, we crawl the pictures of ink landscape paintings on the network and pre-process them as mentioned in Section 4.2, as training set B. Finally, a total of 2168 unpaired images for network training, 360 for testing. The input images are resized to 256 × 256 pixels. The experimental results are shown in Figure 14. We observe the generated images lack a lot of details and the network does not capture more semantic features. Although the generated images show the basic appearance, there are still problems by the way of extracting contour to transform. This explains the inconsistency between contour features and the stroke features. It is also a limitation of our method to be solved in future.

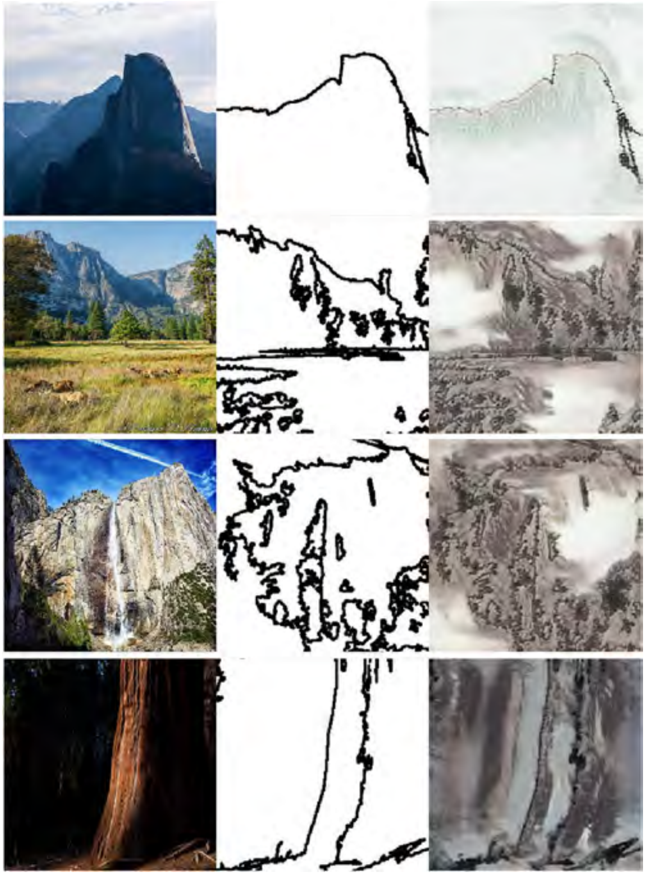
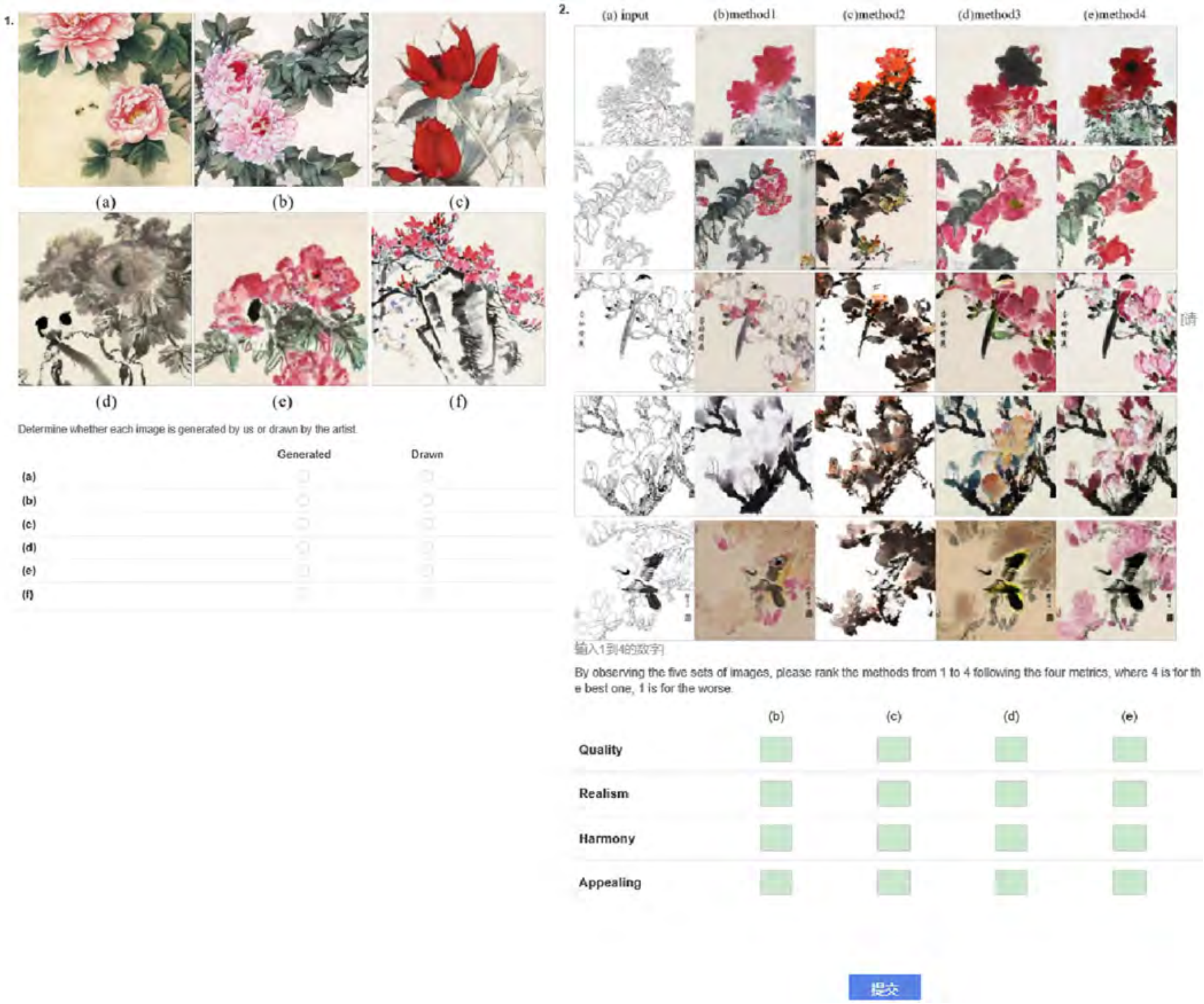


FIGURE 14 The generated landscape images with ink style. The first column is the pictures of nature, second column is the input after extracting the edges, and last row is our generated

5.7 | User study

To reduce the subjectivity of qualitative analysis, we also conduct a user study to compare our approach with other baselines including CycleGAN [8], UNIT [30] and AGGAN [33]. In this study, the users are presented with a web page containing two questions to answer. The form of questions is shown in Figure 15. For the first question, we randomly select six images from our tested results or artists, in which there are half of meticulous images and half of ink painting images. And among them, (a) and (f) are real images from artists, and the other four are generated by our approach. The users need



Determine whether each image is generated by us or drawn by the artist

	Generated	Drawn
(a)	○	○
(b)	○	○
(c)	○	○
(d)	○	○
(e)	○	○
(f)	○	○

输入1到4的数字

By observing the five sets of images, please rank the methods from 1 to 4 following the four metrics, where 4 is for th e best one, 1 is for the worse.

	(b)	(c)	(d)	(e)
Quality	🟩	🟩	🟩	🟩
Realism	🟩	🟩	🟩	🟩
Harmony	🟩	🟩	🟩	🟩
Appealing	🟩	🟩	🟩	🟩

FIGURE 15 The form of our user study web page

to answer the source of each image which is our generated or artist drawn. For the second question, we provide users with four metrics to evaluate the generated images in a variety of ways, that is, quality, realism, harmony and appealing. We ask users to rank the given five sets of images according to their visual perception from 1 to 4 following the four metrics, where 1 is for the worse one and 4 is for the best one. These images are generated by our method and other baselines, respectively. In the study, we use the method number instead of the real method name to reduce users' attention to the method name itself. Specifically, method 1, method 2, method 3 and method 4 correspond to CycelGAN, UNIT, AGGAN and ours, respectively.

The result of question 1 is shown in Figure 16. We can observe that our generated results have successfully fooled the visual perception of users. For the listed four generated images, more than half of the users think they are drawn by artists. In particular, for the image (b), the number of users who think it is from the artist exceeds 80% . The result indicates that our method produces realistic results comparable to real images. The result of question 2 is shown in Figure 17. We can clearly observe that method 4, that is, our method, achieves the maximum value in all metrics including quality, realism, harmony and appealing. From the result, we can also observe that method 2, that is, UNIT [30], performs better than the other two methods in terms of quality and appealing from the visual perception. Method 3, that is, AGGAN [33], has the smallest proportion of all methods for the metric realism. The generated results by method 1, that is, CycleGAN [8], are the least attractive. Based on the average statistic value of the four metrics, we can see that the results generated by our method are more popular among users. The survey results indicate that our approach generates both more realistic and more harmonious high-quality images than others.

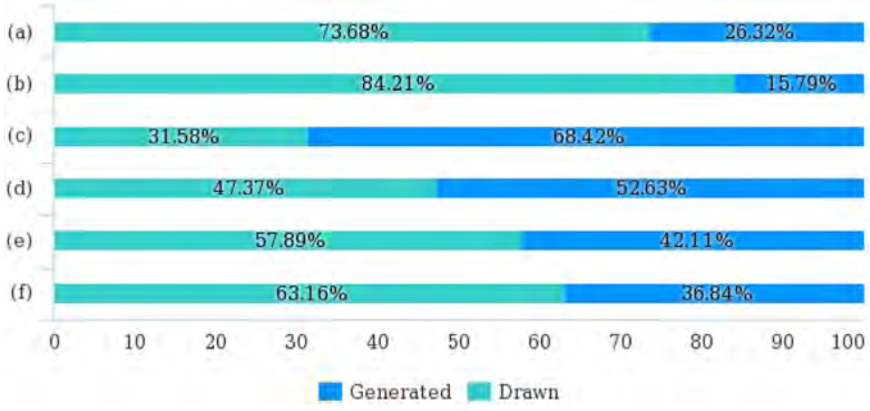


FIGURE 16 The results of question 1 that the users judge whether the images are from our generated or drawn by artists

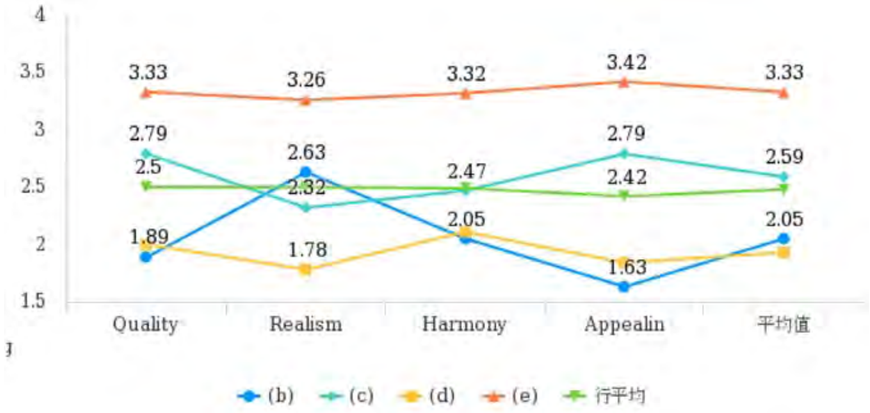


FIGURE 17 The results of question 2 that our method is compared with other baselines including CycleGAN [8], UNIT [30], AGGAN [33] methods in terms of quality, realism, harmony and appealing four metrics

6 | CONCLUSION AND FUTURE WORK

We specialise in multi-style Chinese art painting generation of flowers, which is important and classic, by deep learning method. By virtue of the powerful GAN framework, this method adds techniques such as cycle consistency, attention mechanism, and multi-scale structural similarity to accurately transform the style of traditional Chinese painting datasets. As far as we know, this is the first approach dedicated to the generation of multi-style Chinese art painting of flowers. At present, there are a few researches on the generation of traditional Chinese painting.

In this study, we first construct unpaired flower datasets containing three classic traditional painting style: line drawing, meticulous and ink. Then, to address the fundamental challenge from freehand expression, we propose a Flower-GAN system to generate multi-style Chinese art painting of flowers based on the collected datasets. Our method is used to image-to-image translation task between line drawing images and meticulous or ink style images by an adversarial training way. Moreover, in order to solve the problem of inaccurate, blurred or distorted images generated by existing methods, we further introduce a new loss function called MSSSIM to force the structure preservation. It can effectively improve the quality of generated images, such as colour information, when used in combination with cycle consistency loss. Experimental results show that our method is able to learn a mapping relation that achieves transformation among the multi-style of Chinese art painting of flowers.

We believe that our work is interesting and significant, which is an extension of deep learning research in the field of art and is the inheritance and sharing of traditional Chinese culture. We still have lots of work to do in future. First, collect more high-quality data sets to further improve the quality of generated results. Second, carry out more researches on traditional ink paintings to produce attractive art works. We will continue our research to generate images that are closer to the traditional style and more artistic conception, not just for flowers, but also for landscapes and characters.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Fund for Distinguished Young Scholar under Grant 61625204 and the State Key Program of National Science Foundation of China under Grant 61836006.

The roadmap is provided at https://github.com/feifelicia/Flower-GAN.

REFERENCES

1. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2016, pp. 2414–2423. IEEE, Piscataway, NJ (2016)

2. Goodfellow, I., et al.: Generative adversarial nets. Annual Conference on Advances in Neural Information Processing Systems, December 2014, pp. 2672–2680. Curran Associates, Red Hook, NY (2014)
3. Mao, X., et al.: Least squares generative adversarial networks. Proceedings of the IEEE International Conference on Computer Vision, October 2017, pp. 2794–2802. IEEE, Piscataway, NJ (2017)
4. Zhao, J., et al.: Energy-based generative adversarial network. arXiv preprint arXiv:1609.03126 (2016)
5. Park, T., et al.: Semantic image synthesis with spatially adaptive normalization. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019. IEEE, Piscataway, NJ (2019)
6. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis, Preprint, arXiv:180911096 (2018)
7. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, July 2017, pp. 1125–1134. IEEE, Piscataway, NJ (2017)
8. Zhu, J.Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision, October 2017, pp. 2223–2232. IEEE, Piscataway, NJ (2017)
9. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4401–4410. IEEE, Piscataway, NJ (2019)
10. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. Proceedings of the IEEE International Conference on Computer Vision, October 2017, pp. 1501–1510. IEEE, Piscataway, NJ (2017)
11. Jiang, S., Gao, W., Wang, W.: Classifying traditional Chinese painting images. Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, December 3003, pp. 1816–1820. IEEE, Piscataway, NJ (2003)
12. Liu, C., Jiang, H.: Classification of traditional Chinese paintings based on supervised learning methods. 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), August 2014, pp. 641–644. IEEE, Piscataway, NJ (2014)
13. Zhang, D., Pham, B., Li, Y.: Modelling traditional Chinese paintings for content-based image classification and retrieval. Proceedings of the 10th International Multimedia Modelling Conference 2004, January 2004, pp. 258–264. IEEE, Piscataway, NJ (2004)
14. Jiang, S., Huang, T.: Categorizing traditional Chinese painting images. PCM'04: Proceedings of the 5th Pacific Rim conference on Advances in Multimedia Information Processing - Volume Part I, November 2004, pp. 1–8. Springer, Berlin (2004)
15. Sheng, J., Jiang, J.: Recognition of Chinese artists via windowed and entropy balanced fusion in classification of their authored ink and wash paintings (IWPS). Pattern Recognit. 47(2), 612–622 (2014)
16. Jiang, S., et al.: An effective method to detect and categorize digitized traditional Chinese paintings. Pattern Recognit. Lett. 27(7), 734–746 (2006)
17. Sun, M., et al.: Brushstroke based sparse hybrid convolutional neural networks for author classification of Chinese ink-wash paintings. 2015 IEEE International Conference on Image Processing (ICIP), September 2015, pp. 626–630. IEEE, Piscataway, NJ (2015)
18. Yu, J., Guo, G.M., Peng, Q.S.: Image-based synthesis of Chinese landscape painting. J. Comput. Sci. Technol. 18, 22–28 (2003)
19. Dong, L., Lu, S., Jin, X.: Real-time image-based Chinese ink painting rendering. Multimed. Tools Appl. 69, 605–6200 (2014)
20. Tang, F., et al.: Animated construction of Chinese brush paintings. IEEE Trans. Visual. Comput. Graphics 24(12), 3019–3031 (2017)
21. Yang, L., et al.: Easy drawing: Generation of artistic Chinese flower painting by stroke-based stylization. IEEE Access 7, 35449–35456 (2019)
22. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. Proceedings of the IEEE International Conference on Computer Vision, December, 2015, pp. 415–423. IEEE, Piscataway, NJ (2015)
23. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision - ECCV 2016. ECCV 2016, pp. 649–666. Springer, Berlin (2016)
24. Larsson, G., Maire, M., Shakhnarovich, G.: Learning representations for automatic colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision - ECCV 2016, pp. 577–593, Springer, Berlin (2016)
25. Cao, Y., et al.: Unsupervised diverse colorization via generative adversarial networks. In: Ceci, M., Hollmén, J., Todorovski, L., Vens, C., Džeroski, S. (eds.) Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 151–166. Springer, Berlin (2017)
26. Zhang, H., et al.: Self-attention generative adversarial networks. Proceedings of the 36th International Conference on Machine Learning, pp. 7354–7363. (2019)
27. Chybicki, M., et al.: Deep cartoon colorizer: An automatic approach for colorization of vintage cartoons. Eng. Appl. Artif. Intell. 81, 37–46 (2019)
28. Zou, C., et al.: Lucss: Language-based user-customized colourization of scene sketches. Preprint, arXiv:180810544 (2018)
29. Wang, T.C., et al.: Video-to-video synthesis. Preprint, arXiv:180806601 (2018)
30. Liu, M.Y., et al.: Unsupervised image-to-image translation networks. Annual Conference on Advances in Neural Information Processing Systems, December 2017, pp. 700–708, Curran Associates, Red Hook, NY (2017)
31. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. Preprint, arXiv:13126114 (2013)
32. Huang, X., et al.: Multimodal unsupervised image-to-image translation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision - ECCV 2018, pp. 172–189. Springer, Berlin (2018)
33. Mejjati, Y.A., et al.: Unsupervised attention-guided image-to-image translation. Conference on Advances in Neural Information Processing Systems, pp. 3693–3703. Curran Associates, Red Hook, NY (2018)
34. Kingma, D., Ba, J.: Adam: A Method for Stochastic Optimization. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings (2015)
35. Wang, Z., et al.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
36. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, November 2003, pp. 1398–1402. IEEE, Piscataway, NJ (2003)
37. Barratt, S., Sharma, R.: A note on the inception score. Preprint, arXiv:1801.01973 [stat.ML] (2018)
38. Heusel, M., et al.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. Preprint, arXiv:1706.08500 [cs.LG] (2017)

How to cite this article: Fu F, Lv J, Tang C, Li M. Multi-style Chinese art painting generation of flowers. *IET Image Process*. 2020;1–17. https://doi.org/10.1049/ipr2.12059