Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Special Track on AI, the Arts and Creativity (Exhibits)

High-Resolution and Arbitrary-Sized Chinese Landscape Painting Creation Based on Generative Adversarial Networks

Peixiang Luo, **Jinchao Zhang**, **Jie Zhou**
WeChat AI
{peixiangluo, dayerzhang, withtomzhou}@tencent.com

Abstract

This paper outlines an automated creation system for Chinese landscape paintings based on generative adversarial networks. The system consists of three cascaded modules: generation, resizing, and super-resolution. The generation module first generates a square-shaped painting, then the resizing module predicts an appropriate aspect ratio for it and performs resizing, and finally the super-resolution module is used to increase the resolution and improve the quality. After training each module with the images we collected from the web, our system can create high-resolution landscape paintings in arbitrary sizes.

1 Introduction

Landscape painting is an important creative form in traditional Chinese art with a long history. During the last decade, the development of deep learning technology and the accumulation of online art data have promoted progress in the field of automatic Chinese landscape painting generation. However previous studies based on image translation or style transfer rely too heavily on image-conditioned inputs[Li *et al.*, 2018; Lin *et al.*, 2018; Zhou *et al.*, 2019], which constrains the content and quantity of generated paintings.

To generate Chinese landscape paintings without conditional image, [Xue, 2021] propose the first end-to-end Chinese landscape painting creation model named Sketch-And-Paint GAN (SAPGAN). Inspired by the process of human art creation, SAPGAN divides the generation process of paintings into two stages: sketch and paint. The former generates high-level structures, and the latter improves low-level details. However, SAPGAN can only generate low-quality landscape paintings of a fixed size, which are quite different from real landscape paintings.

To further improve the quality of Chinese landscape painting generation, we introduce a new dataset of 13970 high-quality traditional Chinese landscape paintings and build a modular landscape painting creation system. Specifically,

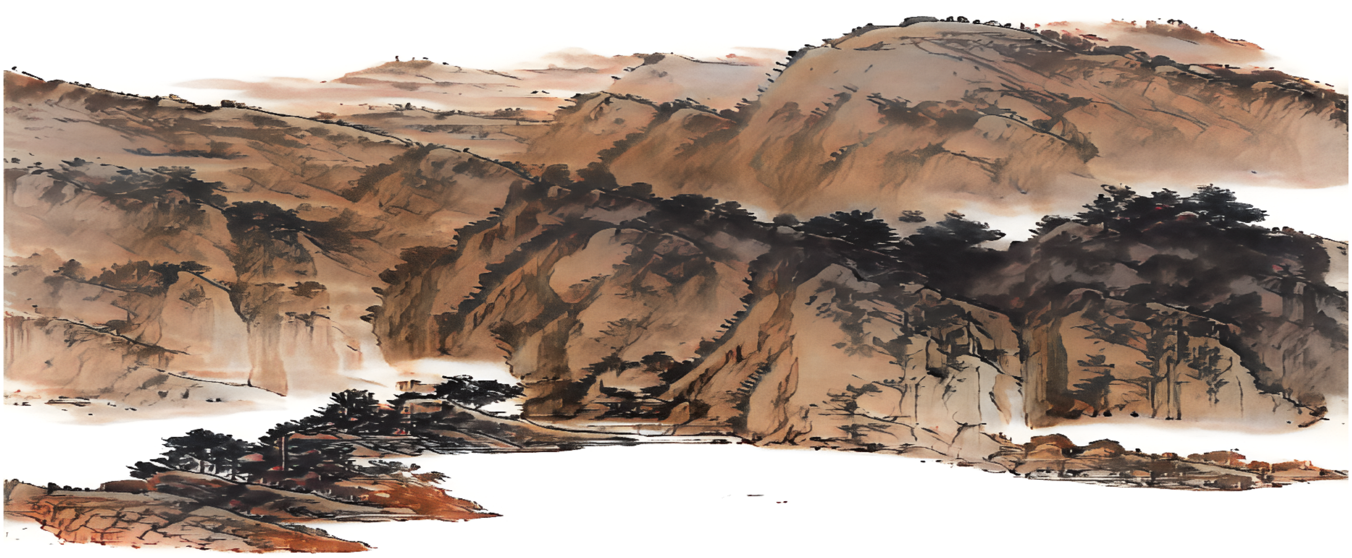


Figure 1: Our system can synthesize high-resolution landscape paintings like this one, which contains 2048 $\times$ 820 pixels

we train a conditional generative model based on stylegan2, which can generate corresponding paintings based on a given aspect ratio interval. We use a multi-layer fully connected network to predict the reasonable aspect ratio of the generated painting, then adjust the size of the painting according to this predicted aspect ratio, and finally we use a super-resolution model to improve the image quality. Compared with previous studies, our proposed system can generate higher-definition landscape paintings with different aspect ratios.

2 Materials and Methods

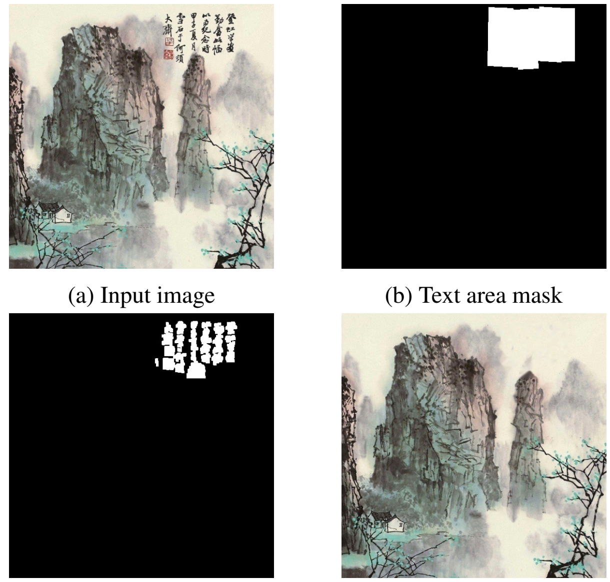
2.1 Data Preparation

We crawled 13970 traditional Chinese landscape paintings from online art websites, search engines and electronic museums. Artists in ancient China often wrote poems on their works to express their mood when painting. About 40% of the landscape paintings we collected contain ancient poems. If we use landscape paintings with ancient poems to train the generative model, the model will generate textures similar to ancient poems on the image, but it is not semantically readable. Therefore, we need to remove ancient poems from the original image.

The ancient poem removal method we adopt first determine text pixels in a coarse-to-fine manner, and then use the known pixels to repair text pixels. This method can achieve considerable results and does not require human-labeled training data. Specifically, it consists of the following three steps:

1. Text area detection: ancient poems are composed of Chinese characters, therefore we use a pre-trained text detection model [Liao *et al.*, 2020] to detect the poem area on the landscape painting.
2. Text segmentation: we choose the OSTU threshold selection method [Otsu, 1979] to determine specific text pixels in the text area. Since the color of ancient poems is quite different from the background, the OSTU algorithm based on the maximum variance method can effectively separate the text pixels from the background.
3. Text pixel repair: we recalculate the value of the text pixel using the known pixels around the text pixel, here we adopt the fast marching method [Telea, 2004] for image inpainting .

Figure 2 shows an example that uses the above steps to remove ancient poems.



(a) Input image (b) Text area mask

(c) Text pixel mask (d) Output image

Figure 2: An example of ancient poem removal

2.2 Generation Module

The aspect ratios between landscape paintings vary widely, depending on what the painting presents and the artist's preferences. Landscape paintings with different aspect ratios will bring people completely different feelings.

However the paintings need to be processed into the same size during GAN-based model training. [Xue, 2021] normalizes all training images to 512x512 resolution by cropping or resizing, which results in the model only generating images with an aspect ratio of 1:1 during testing. The fixed aspect ratio limits the model to generate more aesthetic and continuous landscape paintings.

To avoid this problem, we incorporate aspect ratio information during model training so that the model can generate images with the specified aspect ratio at test time. We classify the training images based on their aspect ratio. Specifically, we classify images with aspect ratios in the range [0, 0.5) as class 1, images with aspect ratios in the range [0.5, 1) as class 2, and so on. In this way, we divide all training images into five categories. Table 1 shows the number of corresponding training images under different categories.

We then trained a conditional generative adversarial network [Mirza and Osindero, 2014] based on the classified landscape paintings. The training images are resized to 512x512 resolution for normalization, but their original aspect ratio bins are used as conditional input to the discriminator and generator. Considering that the training images under each category are at the thousand level, which is relatively limited, the generative model we choose is stylegan2 [Karras *et al.*, 2020b] with adaptive discriminator augmentation mechanism [Karras *et al.*, 2020a], which can effectively prevent the discriminator from overfitting and generate excellent results with limited data.

2.3 Resizing Module

The painting can be generated by specifying the aspect ratio bin after the generation module is trained, but the resolution

Aspect ratio bin	Number of samples
[0, 0.5)	799
[0.5, 1)	3823
[1, 1.5)	2320
[1.5, 2)	4148
[2, 2.5)	2880

Table 1: Distribution of training data divided by aspect ratio bin

of the generated image is fixed. In order to generate paintings with different aspect ratios, we use the resizing module to predict a reasonable aspect ratio of the image and resize the image according to the prediction. The aspect ratio prediction model used here is a multi-layer fully connected network of pyramid structure. The input of the model is the image embedding extracted by the pre-trained resnet50 network [He *et al.*, 2016], and the output of the model is the aspect ratio of the image.

We trained the aspect ratio prediction model on the dataset shown in Table 1. We use the trained model to predict the aspect ratio of the landscape painting output by the generation module, and then resize the image according to the predicted aspect ratio. When resizing, we first fix the long side of the image and then use a linear interpolation algorithm to adjust the short side to fit the predicted aspect ratio.

2.4 Super-Resolution Module

About 60% of our training images are crawled from search engines, and they have compression distortion problems such as common ringing and overshoot artifacts, which lead to low quality of the generated results.

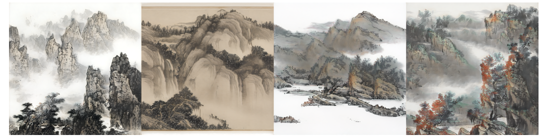
To further improve the resolution and quality of the generated paintings, we train a super-resolution model that can effectively remove noise and increase resolution. We select 1150 high-definition and high-resolution images from the collected data for model training. These images are crawled from online museums and have 2K pixels on at least one of the axes (vertical or horizontal). Here we adopt ESRGAN[Wang *et al.*, 2018] as the model architecture and follow the setting of [Wang *et al.*, 2021] during training. Specifically, we use a high-order degradation process to model practical degradations , and utilize *sinc* filters to model common ringing and overshoot artifacts. The super-resolution model trained with synthetic data is able to enhance details while removing annoying artifacts for the generated landscape paintings.

3 Results

In Figure 3, we show the paintings of different sizes generated by the system. We can see that the results generated by our proposed system can achieve high-resolution while maintaining excellent continuity and integrity. Moreover, our system can generate paintings with different aspect ratios.

4 Conclusion

We propose a multi-stage creation system for ancient Chinese landscape paintings. The system first uses a conditional generative model based on the stylegan2 architecture to generate paintings with specified aspect ratio bins. Then the system adjusts the size of the painting through the resizing module to obtain paintings with different aspect ratios. Finally, the quality and resolution of the paintings are improved by the super-resolution module based on ESRGAN. Our proposed system can generate high-resolution and arbitrary-sized landscape paintings.



(a) Samples with an aspect ratio equal to 1. The size of each sample above is 512×512.

(b) Samples with an aspect ratio greater than 1, sizes from top-to-bottom: 2024×1536,1124×512 and 2048×512.

(c) Samples with an aspect ratio less than 1, sizes from left-to-right: 1636×2576 and 908×1867.

Figure 3: Chinese landscape paintings of different sizes generated by our system. A larger visualization can be found at https://zndls.com/YU1zrLpy

References

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Karras *et al.*, 2020a] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems*, 33:12104–12114, 2020.

[Karras *et al.*, 2020b] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[Li *et al.*, 2018] Bo Li, Caiming Xiong, Tianfu Wu, Yu Zhou, Lun Zhang, and Rufeng Chu. Neural abstract style transfer for chinese traditional painting. In *Asian Conference on Computer Vision*, pages 212–227. Springer, 2018.

[Liao *et al.*, 2020] Minghui Liao, Zhaoyi Wan, Cong Yao, Kai Chen, and Xiang Bai. Real-time scene text detection with differentiable binarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11474–11481, 2020.

[Lin *et al.*, 2018] Daoyu Lin, Yang Wang, Guangluan Xu, Jun Li, and Kun Fu. Transform a simple sketch to a chinese painting by a multiscale deep neural network. *Algorithms*, 11(1):4, 2018.

[Mirza and Osindero, 2014] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[Otsu, 1979] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[Telea, 2004] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34, 2004.

[Wang *et al.*, 2018] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[Wang *et al.*, 2021] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021.

[Xue, 2021] Alice Xue. End-to-end chinese landscape painting creation using generative adversarial networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3863–3871, 2021.

[Zhou *et al.*, 2019] Le Zhou, Qiu-Feng Wang, Kaizhu Huang, and Cheng-Hung Lo. An interactive and generative approach for chinese shanshui painting document. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 819–824. IEEE, 2019.