

RESEARCH ARTICLE

Open Access



# An improved algorithm for superresolution reconstruction of ancient murals with a generative adversarial network based on asymmetric pyramid modules

Shang Ma<sup>1,2</sup>, Jianfang Cao<sup>1,2,3\*</sup> , Zhaoxia Li<sup>1</sup>, Zeyu Chen<sup>1,2</sup> and Xiaohui Hu<sup>1,2</sup>

## Abstract

Ancient Chinese murals are true portrayals of ancient Chinese life, but well-preserved murals are rare. Therefore, ancient mural preservation and repair are critical. To address the poor superresolution reconstruction of mural images with unclear textures and fuzzy details, we developed an improved generative adversarial network (GAN) algorithm based on asymmetric pyramid modules for ancient mural superresolution reconstruction. Asymmetric pyramid modules, which are composed of a series of dense compression units, were used to learn image features. To analyze the reconstructed image features, a perceptual loss function was integrated to optimize the model performance. The use of the improved algorithm for low-resolution mural images increased the image resolution while preserving their original feature details and textures, and the improvement effect was visually observed in terms of indices such as the peak signal-to-noise ratio and structural similarity. Compared with other superresolution-related algorithms, the proposed model increased the peak signal-to-noise ratio by 0.20–6.66 dB. The GAN-based mural superresolution reconstruction algorithm proposed in this study effectively improved the performance of reconstructed high-resolution mural images, which increases the significance of the reconstructed image for future research.

**Keywords:** Image processing, Superresolution reconstruction, Generative adversarial network, Deep learning

## Introduction

Ancient Chinese cultural relics have a long history and, to some extent, reflect the pursuits and longings of people during certain periods. However, intact murals are rare; therefore, mural restoration and protection should be prioritized.

Mural restoration work includes mural surface cleaning, pasting loose and falling objects onto the mural surface, and strengthening the mural surface with appropriate materials. Although these processes can delay the decay of a mural, they cannot fundamentally restore its original

appearance. With the development of computers, a number of algorithms can be used to “restore” mural images, including mural image superresolution reconstruction.

Since the beginning of the 21st century, as scholars and researchers have focused increasingly on deep learning algorithms, deep learning-based algorithms have emerged. Image superresolution reconstruction has also developed from traditional learning algorithms to deep learning algorithms. However, different deep learning algorithms have advantages and disadvantages in various scenarios, and therefore, there is no algorithm that is truly competent for all scenarios.

Currently, the main problems faced by mural image superresolution algorithms are as follows. (1) Mural image features are complex and clearly different from those of traditional images, which increases the difficulty

\*Correspondence: caojianfangcn@163.com

<sup>1</sup> Department of Computer Science & Technology, Xinzhou Normal University, No. 10 Heping West Street, 034000 Xinzhou, China  
Full list of author information is available at the end of the article

of feature extraction. (2) Deep learning-based super-resolution algorithms require satisfactory datasets for training, and there are currently no datasets that can satisfactorily meet the requirements. (3) When ordinary high-definition image datasets are used for model training, the reconstructed mural image exhibits poor feature restoration and texture detail performance.

Glasner [1] combined image self-similarity at the same scale and across scales and gradually enlarged images by searching and pasting. However, the reconstructed image was not ideal in terms of texture details. Since the beginning of the 21st century, superresolution methods based on deep learning have continuously emerged. By referencing the idea of image superresolution (SR) based on sparse coding, the triple convolution network superresolution convolutional neural network (SRCNN) [2] applied deep learning to image SR for the first time. This network performs image block extraction, nonlinear feature mapping and reconstruction through convolution operations. This method outperforms traditional algorithms to some extent. Because the SRCNN algorithm fails to fully utilize prior knowledge in the related field, some improved algorithms have emerged. The sparse coding-based network (SCN) [3] realizes image SR based on sparse coding through a neural network. The SCN uses the learned iterative shrinkage and thresholding algorithm (LISTA) to incorporate the sparse representation, mapping and reconstruction modules in the SR method based on coefficient representations and can enlarge an image to any scale with cascaded SCNs. To improve the reconstruction speed for high-resolution images, an efficient subpixel convolutional neural network (ESPCNN) [4] uses a convolutional neural network to improve image feature extraction by stacking multiple convolution kernels; however, the weight of the network is not ideal.

Further studies have revealed low-frequency similarities between low-resolution (LR) images and reconstructed high-resolution (HR) images. Based on this finding, Kim et al. [5] proposed a very deep CNN for SR (VDSR). The VDSR algorithm was the first to incorporate a residual structure into SR image reconstruction, which greatly reduces the computational burden of the network. Zhang et al. [6] proposed an algorithm for increasing the role of low-frequency information in the learning process, and their results show that incorporating the channel attention mechanism increases the average PSNR of the reconstructed images by 0.4 dB. The continuous updating of computer hardware has increased technical support for constructing deep learning-based networks [7–9]. After a development period, generative adversarial networks (GANs) have been applied in a variety of fields, and their excellent performance has attracted the attention of scholars and researchers. Ledig

et al. [10] proposed a superresolution GAN (SRGAN) for recovering high-frequency details from 4 upsamplings. This model includes a generative network and a discriminant network. The generative network is a residual network that generates an HR image from the LR image. The authenticity of the input image is then judged by the discriminant network. When the two networks reach a balance, the proposed network can be used for SR. However, GANs have some noticeable drawbacks, such as instability during training and gradient explosion and loss. Considering the drawbacks of GANs, Arjovsky et al. [11] proposed a Wasserstein GAN (WGAN). This model replaces the Jensen–Shannon divergence with the Wasserstein distance to assess the difference between the real and reconstructed images, overcoming the disadvantages of GANs. Although this algorithm can satisfactorily reconstruct superresolution images in ordinary scenes, mural images pose a great challenge for the model learning capacity, and the training time and network weight must also be considered. In addition, most of the currently available deep learned-based image superresolution reconstruction algorithms use the mean square error loss function as the optimization function to optimize the network structure; however, the images reconstructed by this method have some problems, such as loss of high-frequency information, fuzzy texture, and poor subjective evaluation. To overcome these problems, Han [12] used the perceptual loss function for model optimization. Yang et al. [13] completed the generator module reconstruction task by gradually extracting different image scales; in the discriminator module, they introduced microstep convolution and global average pooling. Su et al. [14] proposed a single remote sensing image superresolution method based on boundary equilibrium GANs, which improved the quality of the generated image and the network convergence speed. Mi et al. [15] used three convolution kernels to extract image features, which improved the overall model performance on images with complex features. For the superresolution reconstruction of ancient cultural mural images, Wang et al. [16] proposed dictionary learning through an MOD algorithm, which realized mural repair and impulse noise removal. Ma [17] proposed a method for superresolution reconstruction of single-color images based on sparse representation, which effectively improved the color authenticity of the reconstructed color mural image. Cao et al. [18] proposed a superresolution reconstruction algorithm for a stably enhanced generative confrontation network to solve the problems of low resolution and unclear texture details in ancient murals. Han et al. [19] proposed a method based on fuzzy enhancement for wavelet signal denoising, which increased the accuracy of the extracted feature information. These endeavors

aimed to optimize the network models from different perspectives; however, none of them addressed the problem of the difficulty in training GANs. Furthermore, if the speed is increased at the expense of the network depth, the final result will be unsatisfactory. Targeting the problem of the sawtooth effect on the mural edges, Xu et al. [20] fused an attention network and a residual network to reconstruct murals, which increased the network training time while improving the texture details of murals, and consequently, their method is likely to lead to local optimization of the network.

Based on the aforementioned issues, we propose a ProGAN algorithm in this work. By gradually increasing the network depth, stacking multiple dense compression units and removing the model-enhancing aspect of the batch normalization layer during training, this model fully captures the feature information of mural images, indicating the improved generalization ability and robustness of the model. By replacing the cross-entropy loss with a least square loss, the training process of the model is further stabilized. The results of this study may provide a method for preserving ancient murals. The proposed model not only introduces new in-depth learning technologies but also provides its own contribution to the protection of digital cultural heritage.

## Methodology

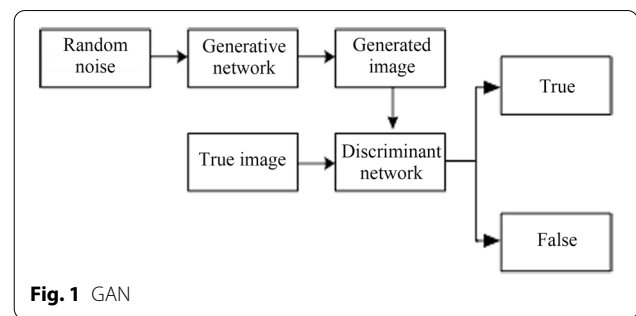
### Theoretical background

#### Generative adversarial network (GAN)

GANs have become a popular deep learning model in recent years, and they are one of the most promising methods for unsupervised learning on complex distributions. A GAN is composed of two basic networks: generators (G) and discriminators (D). The G network is an image generation network that is responsible for receiving random noise and generating an image based on the noise. The D network is responsible for determining whether the input image is authentic. The network has a probability interval between 0 and 1, with a higher probability value indicating that the image is more likely to be authentic. In the actual training process, the goal of the G network is to generate an image that is as close as possible to the real image to deceive the D network, while the goal of the D network is to distinguish the image generated by the G network from the real image. In this way, a mutual game process between G and D networks is formed. The basic structure of the GAN is shown in Fig. 1.

#### Progressive generative adversarial network

- (1) Overall design. To better reconstruct high-resolution images, we improved the structure of both the generative network and the discriminant network.



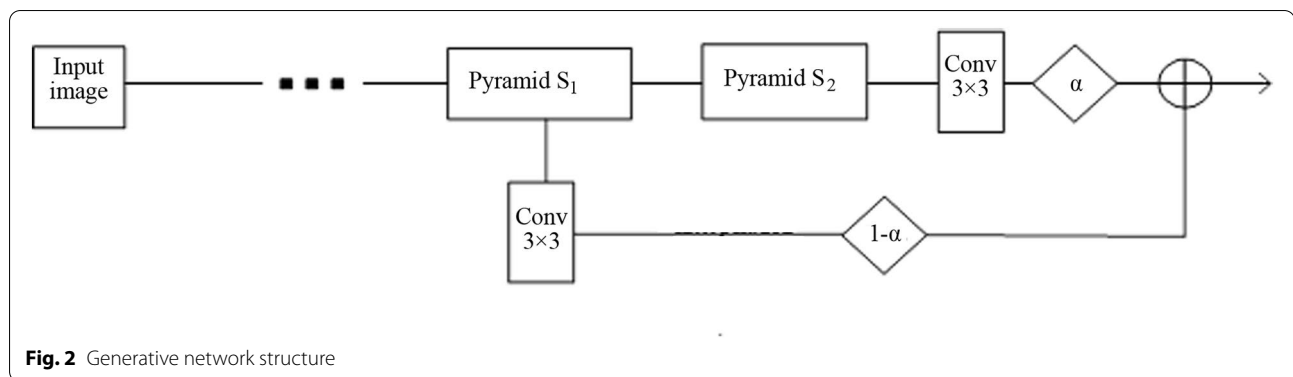
In the generative network design, upsampling low-resolution to high-resolution images is divided into several decomposition steps by introducing parameters  $\alpha$  that adjust the rhythm of each network layer during training. This design improves the smoothness and stability of the training process, thereby reducing the gradient explosion and gradient disappearance problems that GANs frequently experience during training. Furthermore, this modification can more fully extract the texture features and details of the input image, improving the quality of the image generated by the subsequent generator.

The discriminant network distinguishes whether the input image is the network-generated image or the real image. The details of the network model are described as follows.

- (2) Generative network design. This network is responsible for learning based on the characteristics of the input image. In the structure of the generative network, the sampling process on the image is divided into several pyramid modules. Each pyramid module uses a dense compression unit composed of different numbers of residual blocks. The outputs of the pyramid modules are then convoluted. To smooth the overall model training process, the parameter  $\alpha$  is added for control, and the output of the residual R in the preceding layer network is produced after bilinear interpolation.

The realization of the network is shown in Fig. 2.

- (3) Discriminant network design. The network is composed of three asymmetric pyramid modules. In each module, a mean pooling module is integrated to downsample the input to reduce the spatial dimensions of the input image. Similar to the generative network, before the results are input into the asymmetric pyramid structure, the results first pass through image transformation layers with specific sizes because the generator outputs are different sizes. These image transformation layers are com-



posed of  $3 \times 3$  convolution layers. The final output of the asymmetric pyramid module is a feature matrix, and the discriminant network determines the true or false value of each pixel in the feature matrix. The advantage of this design is that the model can focus more on the image details. The discriminant network structure is shown in Fig. 3.

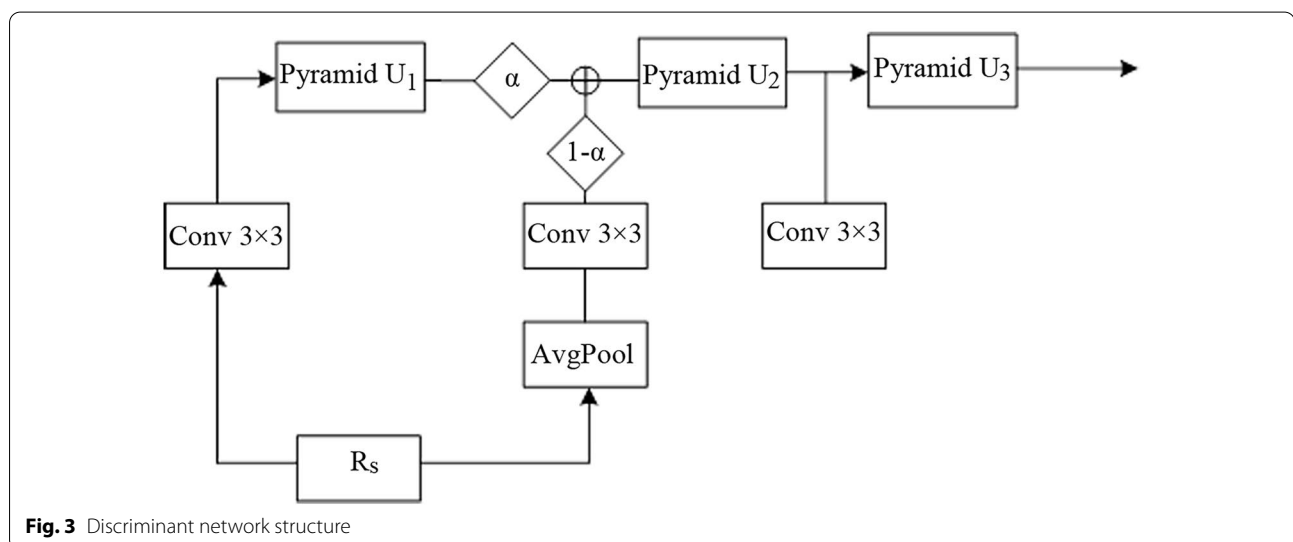
### Asymmetric pyramid network

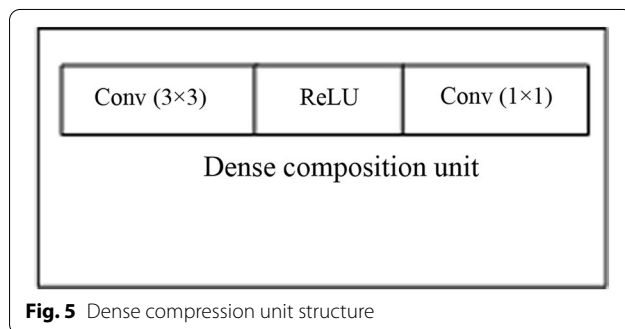
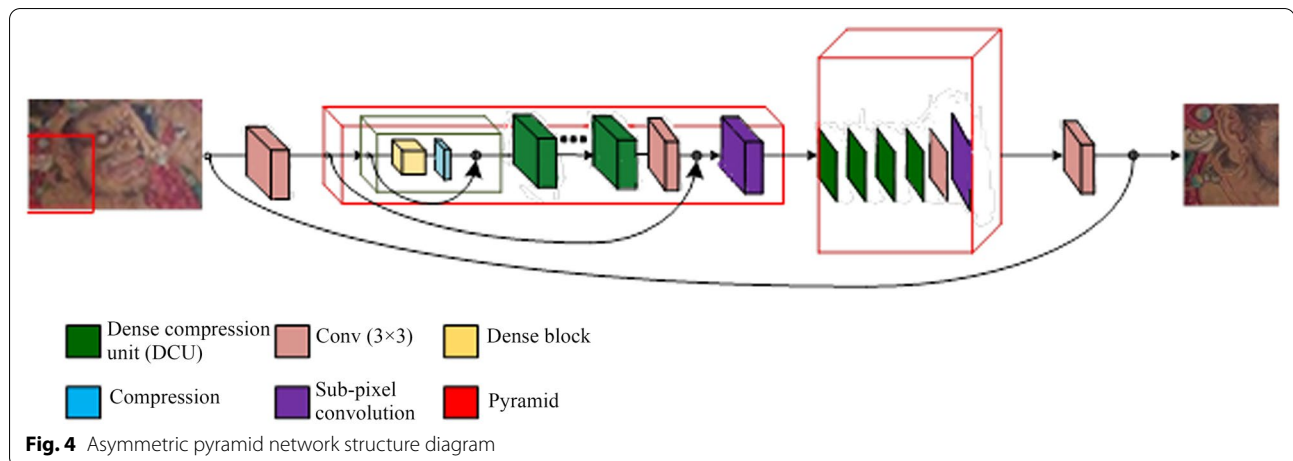
When extracting features, the traditional pyramid network [20] uses the Laplacian pyramid superresolution network to gradually reconstruct high-resolution images. In this study, we use an asymmetric pyramid network for feature extraction. The pyramid network decomposes the input image unit  $U$  into a series of simpler functions  $U_0, U_1 \dots U_s$ . The purpose of each function (or level) is to refine the extracted features, which are then subjected to 2 upsampling steps. Each pyramid layer consists of a cascaded dense compression unit and a subpixel convolution layer. Because more detailed feature extraction is

required in the lower layers, denser compression units are added to these layers; thus, the pyramid has an asymmetric structure, and the lower layers of the pyramid have higher computing capacities. This treatment not only reduces memory consumption but also expands the perception domain relative to the original image. Therefore, the reconstructed structure outperforms the traditional symmetrical structure in terms of quality and runtime. The diagram of the network structure is shown in Fig. 4.

### Dense compression unit

The core component of each layer of the pyramid is a dense composition unit. Based on recent studies [21–24], this unit is composed of a modified dense connection block and a  $1 \times 1$  convolution layer. Originally, the dense layers are designed to begin with the batch normalization layer. In contrast, in this study, we remove the batch normalization layer in the dense compression unit, as well as the rectified linear unit function (ReLU) in the first layer. In contrast to the dense network, the network in





this study employs a  $1 \times 1$  convolution layer as the final layer of each dense composition unit to indicate the end of the dense connection of each unit. This modification effectively reorganizes the information obtained by the network layer and increases the number of DCUs; furthermore, it greatly reduces the parameters of the model and provides relevant data in subsequent comparative experiments without affecting the peak signal-to-noise ratio (PSNR). The structure diagram of the dense composition unit used in this study is shown in Fig. 5.

### Loss function design

- (1) Least squares loss. The purpose of the loss function is to improve the expressiveness of the model during the training process. In this study, we use a least squares loss function to adjust the parameters of the model and ensure that the model fit the datasets as well as possible. Each dataset contains  $n$  points (composed of data pairs), i.e.,  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , where  $x_1$  is an independent variable and  $y_1$  is a dependent variable. Assume that the model function has

the form  $f(x, \beta)$ , where  $m$  adjustable parameters are stored in the vector  $\beta$ . The goal of the least squares loss function is to find the optimal parameter value by minimizing the sum of the squares of the residuals (i.e., by continuously optimizing the model).

In the model, the least squares loss function of the discriminator can be expressed as follows:

$$L_{D_s}^i = [D(\hat{r}_i^s)]^2 + [D(r_i^s) - 1]^2 \quad (1)$$

- (2) Perceived loss. To ensure that the image produced by the generator is as similar to the real image as possible in terms of features and structure, a perceptual loss is introduced into the VGG network to form the loss function of the generator. After a high-resolution image is reconstructed by the generator, the VGG loss calculates the error loss between the reconstructed and real images. This loss function and the least squares loss are combined to form the loss function of the generative network for model optimization. This treatment enables the generated image to be closer to the real image. The expression of the loss function of the generator is as follows:

$$L_{G_s}^i = [D(\hat{r}_i^s) - 1]^2 + \sum_{k \in \{2,4\}} \|\Phi_k(\hat{y}_i) - \Phi_k(y_i)\|^2 \quad (2)$$

where  $\hat{r}$  represents the predicted residual of the generated image and  $\Phi_k$  represents the  $k$ th pooling layer of VGG16 [25].

### Algorithm flow descriptions

The main idea underlying the proposed ProGAN is a progressive training process, which is manifested in the design of the network structure. The detailed algorithm flow is as follows:

Input: an LR image.

Output: a superresolution reconstruction network model.

Step 1. Input an LR image into the generative network  $G$  to obtain a high-resolution image  $HR_0$  after reconstruction.

Step 2. Input an authentic image (HR) into the generative network  $G$  and calculate the least squares loss

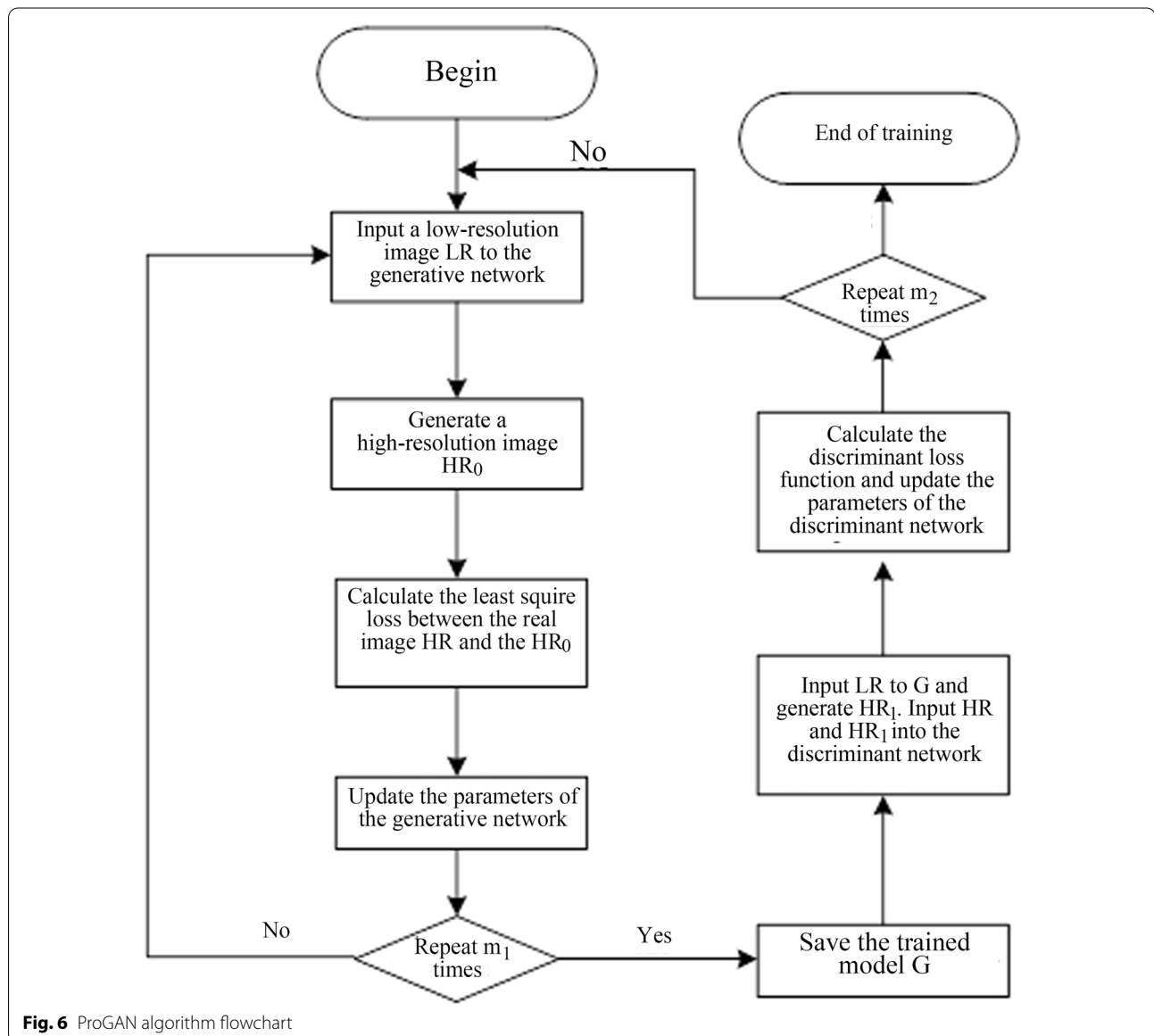
between  $HR$  and  $HR_0$  to update the parameters of the generative network.

Step 3. Repeat Steps 1 and 2  $m_1$  times and store the results in the posttraining generative model  $G$ .

Step 4. Input the LR image into the generative network  $G$  to generate a high-resolution image,  $HR_1$ . Input  $HR$  and  $HR_1$  into the discriminant network to calculate the loss and then update the parameters of the discriminant network.

Step 5. Calculate the perceived loss between  $HR$  and  $HR_1$ .

Step 6. Iterate Step 2 to Step 6  $m_2$  times to continuously update the generative and discriminant networks.



**Fig. 6** ProGAN algorithm flowchart



The ProGAN algorithm flowchart is shown in Fig. 6.

### Experiment

In this study, the hardware parameters were as follows: (1) CPU, Intel Core i7-7700 K, (2) memory, 16 GB, and (3) graphics card, NVIDIA GeForce GTX 1080Ti. The software included CUDA (version 9.1) and the Ubuntu operating system. Python 3.6 and the PyTorch framework were used to write the test code. The compiling software was PyCharm Community Edition 2021.1  $\times$  64.

The quantity and quality of the murals in the dataset were not sufficient for training the network. Therefore, the dataset used for training was the publicly available high-resolution training dataset DIV2K. The training images included 800 high-resolution pictures, and the test dataset included 100 high-resolution images. The high resolution of the images enabled the network model to better learn the details to be retained during the process of image reconstruction from low resolution to high resolution, and the subsequent reconstruction of the mural images also performed well. In this experiment, the initial learning rate was set to 0.0001, the input batch\_size was 24, and 80 epochs were trained.

The peak signal-to-noise ratio (PSNR) is a common objective index for measuring the image quality of lossy transformations (such as image compression and image restoration). For superresolution reconstruction, the PSNR was defined by the maximum pixel value  $L$  between images and the mean square error MSE between images, where  $n$  was the number of bits in each sampling value. The PSNR is inversely proportional to the logarithm of the mean square error of the real high-resolution image and the generated image. The PSNR is calculated as follows:

$$PSNR = 10 \cdot \log_{10} \left[ \frac{(2^n - 1)^2}{MSE} \right] \quad (3)$$

The structural similarity index method (SSIM) was used to measure the structural similarity between images based on the brightness, contrast and structure. Assume that the pixels of the high-resolution image are represented by  $I$  and the pixels in the reconstructed images are represented by  $\hat{I}$ ; then, the SSIM formula can be expressed as a weighted product of the comparisons of the brightness, contrast and structure. The SSIM is calculated as follows:

$$SSIM(I, \hat{I}) = [C_l(I, \hat{I})]^\alpha [C_c(I, \hat{I})]^\beta [C_s(I, \hat{I})]^\gamma \quad (4)$$

where  $C_b$ ,  $C_c$  and  $C_s$  represent the comparison values of the two images in terms of brightness, contrast and structure, and  $\alpha$ ,  $\beta$  and  $\gamma$  represent the three weight values.

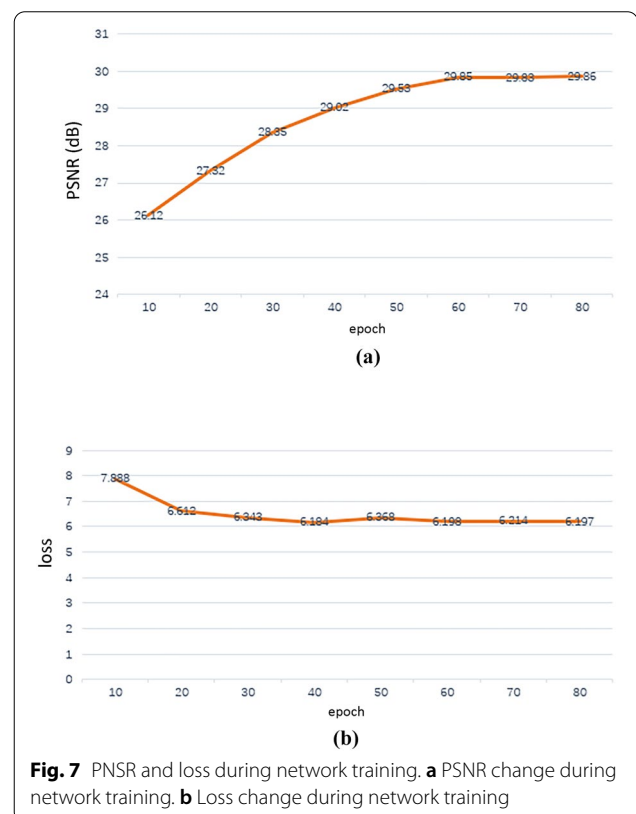
The final evaluation index is the random scores of the population. Five volunteers with normal vision were randomly selected to provide different scores for the reconstructed images. The scores were divided into several different grades: 1.0, 2.0, 3.0, 4.0 and 5.0. A higher score indicated that the reconstructed image better matched the real image.

## Results and discussion

### Model training loss and analysis

The loss function is a useful tool for network structure assessment. During the training with the ProGAN algorithm in this study, gradient optimization was realized by continuously updating each of the parameters. After training for 80 epochs, the parameters achieved an optimal state, which resulted in the maximization of the loss function. The variations in the loss and PSNR values according to epoch training are shown in Fig. 7.

According to the variations in the loss and PSNR values, the loss function stabilized near 6.2, and the PSNR value no longer showed an increasing tendency after



**Fig. 7** PSNR and loss during network training. **a** PSNR change during network training. **b** Loss change during network training

**Table 1** Comparison of the composition effects of different DCU structures

Structure	PSNR	Number of parameters (M)	Runtime (s)
Dense block $\times 1$	28.31	8.22	0.19
DCUs $\times 4$	28.32	1.79	0.10

increasing to approximately 29.85. At this moment, the network achieved a satisfactory convergence effect.

#### Influence of the dense compression unit composition on the network performance

In this study, the dense compression unit (DCU), which is the core component of the pyramid structure, was designed by referencing the residual network [26]. In our study, a  $1 \times 1$  convolution compression layer was used as the final layer of each dense compression unit. This treatment increased the number of dense compression units but did not affect the PSNR performance, and the number of model parameters was successfully reduced. The comparison results are summarized in Table 1.

The results provided in Table 1 were based on comparisons of the Set14 dataset. During operation, we tested the reconstruction effect of the two structures at  $4\times$  super-resolution. Although the number of DCUs was increased in our study, this increase did not affect the PSNR value of the reconstructed images. Furthermore, the number

of parameters was greatly reduced. The adoption of such a structure during training resulted in a network model that consumes less time but is more lightweight with better performance.

#### Superresolution reconstruction effect

To test the expressiveness of the algorithm for mural image reconstruction, several representative murals with different styles were selected for  $4\times$  superresolution processing. The results are shown in Fig. 8. The reconstructed high-resolution images retained some of the texture details of the original images, and the clarity was also satisfactory. The PSNR and SSIM values of the reconstructed image are summarized in Table 2. Based on the numerical values, the reconstructed image had satisfactory visual expressiveness.

Furthermore, from the mural images of different styles, we randomly selected a total of 40 images (5 for each style), and then noise was added. After superresolution reconstruction with the proposed model, the measured average PSNR and SSIM values were 21.22 and 0.37, respectively. The reconstructed images are shown in Fig. 9, where images with more added noise (Fig. 9a and c) and those with less added noise (Fig. 9b and d) were selected for effect displays. The PSNR and SSIM values are summarized in Table 3.

As shown in Fig. 9; Table 3, for images with more added noise, the noise was also proportionally amplified after

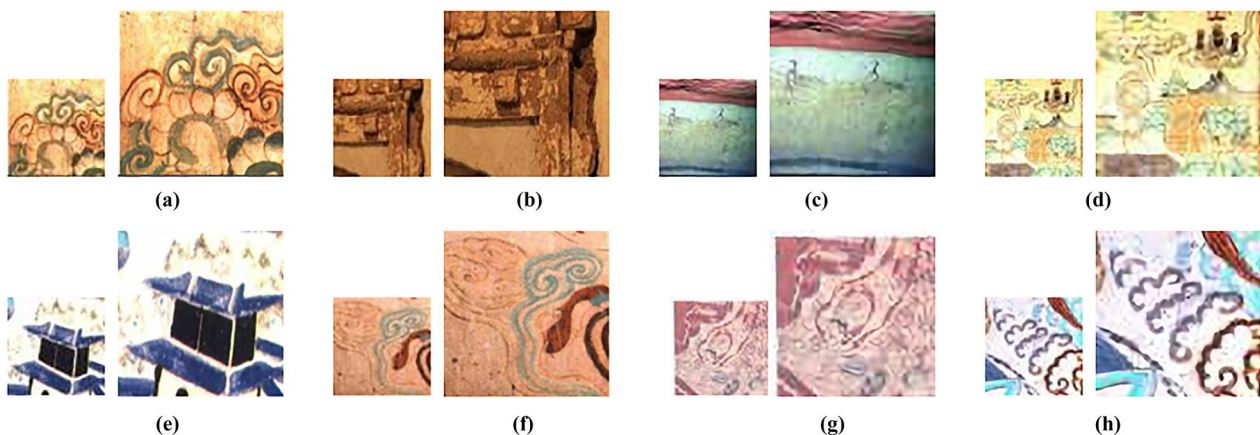
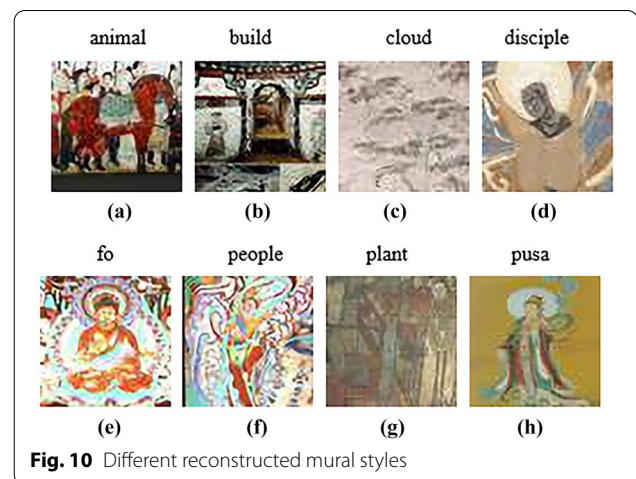
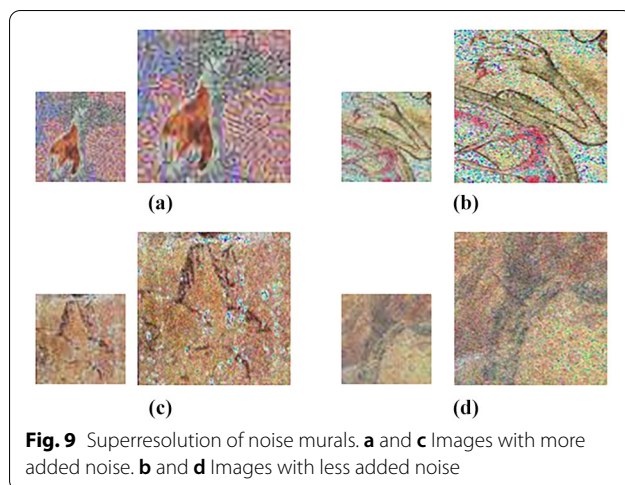
**Fig. 8** Reconstruction of different styles of murals**Table 2** PSNR and SSIM of the mural after reconstruction

Image	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
PSNR	28.02	29.17	27.88	25.97	29.87	30.94	25.68	26.75
SSIM	0.76	0.75	0.71	0.66	0.84	0.77	0.58	0.78





**Table 3** PSNR and SSIM of the noise image after reconstruction

Image	(a)	(b)	(c)	(d)
PSNR	17.31	24.75	22.03	26.94
SSIM	0.23	0.47	0.43	0.48

(a) and (c), images that with more added noise; (b) and (d), those with less added noise

superresolution handling, as a consequence of which the images remained indistinct (also attested to by the PSNR value). In contrast, the images with less added noise displayed more distinct images as well as noise after superresolution handling.

### Comparative experiment of the reconstructed mural images of different types

To test the applicability of the proposed model for various mural styles, we selected eight types of mural images: animal, building, cloud, disciple, fo, people, plant and pusa with 851, 647, 732, 640, 484, 819, 588 and 610 images, respectively, for a total of 5371 images.

First, the ProGAN superresolution reconstruction model was used for  $4\times$  superresolution reconstruction of the above 5371 images. The reconstructed images are shown in Fig. 10. The average PSNR was 27.15 dB, and the average SSIM was 0.68.

Then, for comparison purposes, we selected three classical superresolution methods, i.e., bicubic interpolation (BI) superresolution reconstruction, SRGAN superresolution reconstruction, and enhanced deep residual networks, for single image superresolution (EDSR) superresolution reconstruction, as well as two other superresolution methods already applied to mural images in the literature [17] and [20]. To better

reflect the performance of the algorithms on mural images of different styles, the 5371 images were divided based on their style. Six different algorithms were used for  $4\times$  superresolution reconstruction, and then the reconstruction effects were compared (Fig. 11).

As shown in Fig. 11, the performance effect of the traditional superresolution reconstruction method BI was not satisfactory when dealing with certain types of images, i.e., murals. The reconstructed images exhibited obvious blurs and unclear textures, as well as considerable feature loss. In addition, the reconstructed images contained artifacts and noise that did not exist in the original images. In the literature [17], the mural image superresolution reconstruction method is based on a traditional algorithm (not one based on deep learning). This method can achieve a satisfactory effect when applied to mural images with simple textures. In regard to the mural images with complex textures involved in this study, however, it created artifacts in the reconstructed image, similar to the BI algorithm, resulting in an unsatisfactory overall effect. The method used in the literature [20] improved to some extent the texture blurring effect and margin zigzagging that CNNs encounter when dealing with mural images, which can be reflected by an improved image structure after reconstruction compared with that after BI reconstruction. However, when applied to images of the “fo”, “disciple”, “people” and “pusa” classes, it failed to handle the detailed features, ultimately leading to unsatisfactory effects. Compared with the traditional BI algorithm, the deep learning-based SRGAN algorithm had a noticeably better performance due to the GAN method applied for superresolution reconstruction. The reconstructed image displayed a satisfactory reduction in texture detail while retaining the color and features of the original images. Compared with the BI algorithm,



the image reconstructed by this algorithm did not have obvious colors that were too bright or dark, and the original image contrast was essentially maintained. However, this algorithm showed varying performance for the details of images with different styles.

After reconstructing images in the disciple, fo, people and pusa categories, this algorithm increased the number of artifacts in the faces of the reconstructed images, while for the other image categories, such as nature and building, this method did not satisfactorily preserve the detailed features. EDSR is a superresolution reconstruction network model that relies on an exploration period. This method showed satisfactory performance in terms of retaining texture details and colors. Compared with the original HD image and the images reconstructed by the two previous methods, the images reconstructed by this algorithm were the most similar to the real image

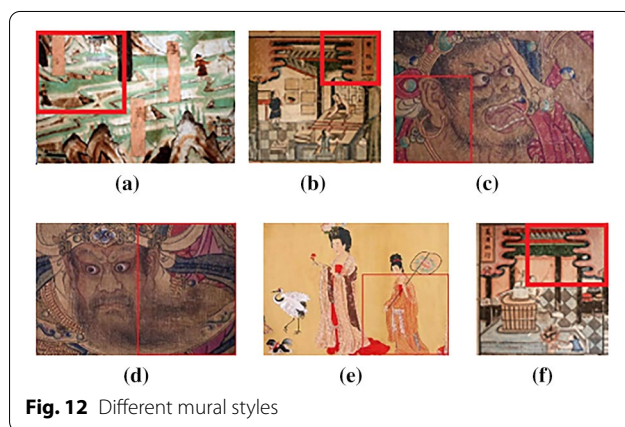
and had a satisfactory performance for different styles of mural images. The network model proposed in this study further improved the expressiveness of the reconstructed images. It not only retained the color of the original background of the image but also considered the texture details and features of the image. Compared with the other algorithms, the proposed algorithm exhibited clear advantages. The following table summarizes the PSNR and SSIM values for the different style images after reconstruction.

As shown in Table 4, the ProGAN algorithm noticeably outperformed the other algorithms in reconstructing different styles of mural images. Compared with the BI algorithm, which had the worst performance among the considered algorithms, the ProGAN algorithm increased the average PSNR value by 6.66 dB. Although the EDSR algorithm showed a satisfactory reconstruction effect,



**Table 4** PSNR and SSIM after reconstruction of various murals with different algorithms

Image type	Index	BI	Literature [17]	Literature [20]	SRGAN	EDSR	ProGAN
Animals	PSNR/SSIM	20.92/0.44	27.36/0.63	27.81/0.64	28.05/0.66	28.44/0.68	28.56/0.68
Builds	PSNR/SSIM	19.89/0.44	24.21/0.57	25.15/0.62	25.41/0.63	25.62/0.64	25.85/0.65
Cloud	PSNR/SSIM	19.84/0.48	24.40/0.62	25.40/0.66	25.62/0.67	25.87/0.69	26.18/0.70
Disciple	PSNR/SSIM	20.21/0.46	25.88/0.62	26.70/0.62	26.81/0.66	27.15/0.68	27.35/0.69
Fo	PSNR/SSIM	20.06/0.45	24.05/0.56	24.88/0.58	25.06/0.61	25.23/0.62	25.49/0.63
People	PSNR/SSIM	20.23/0.43	26.92/0.64	27.50/0.63	27.75/0.67	28.12/0.69	28.28/0.70
Plant	PSNR/SSIM	21.20/0.50	26.75/0.66	27.44/0.65	27.60/0.70	28.05/0.72	28.21/0.72
Pusa	PSNR/SSIM	20.49/0.44	24.99/0.58	25.83/0.61	25.91/0.62	26.11/0.64	26.27/0.65



the proposed algorithm increased the PSNR value by 0.2 dB. In summary, the ProGAN algorithm was more suitable for mural superresolution reconstruction.

#### Comparative experiment of the reconstructed murals

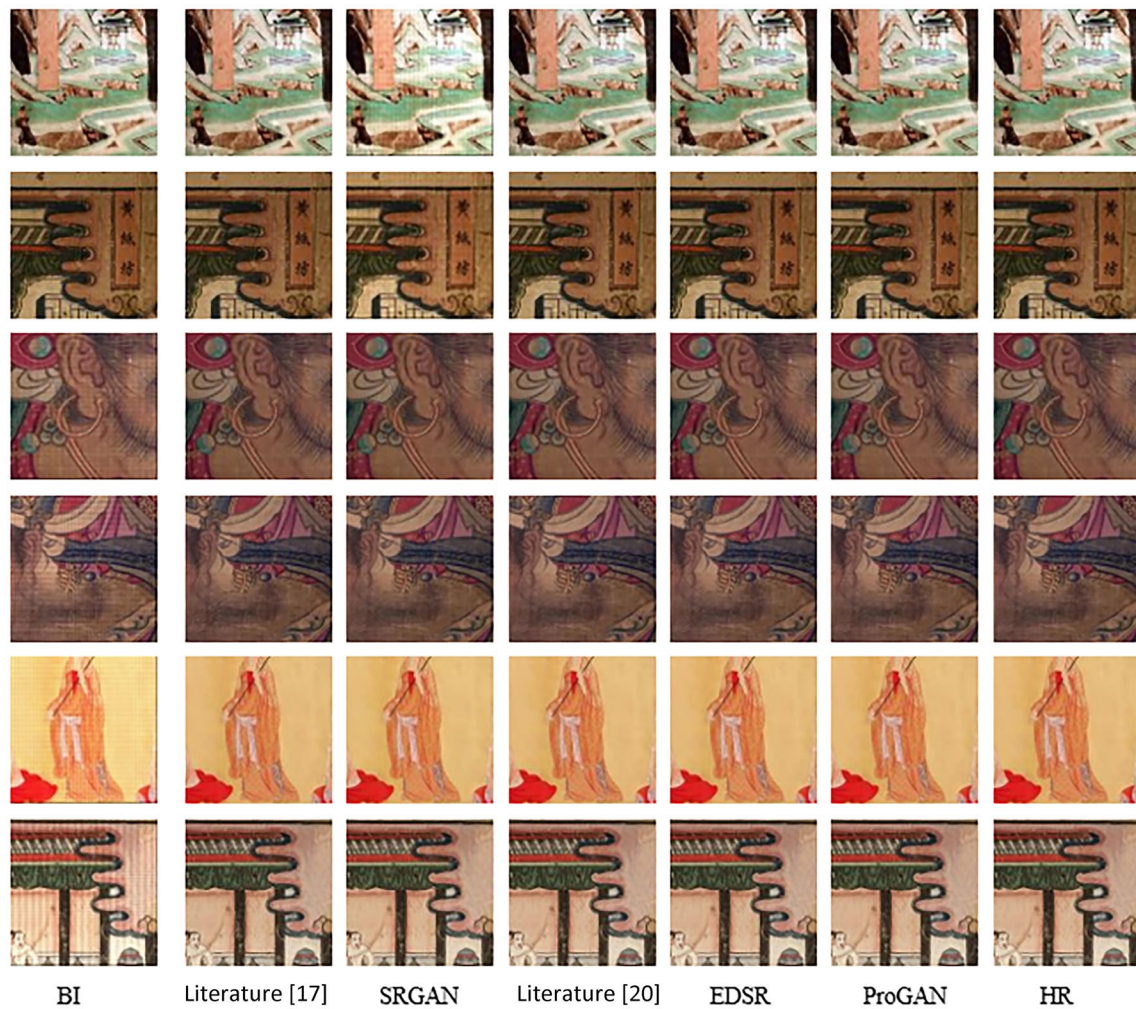
Because reconstructing mural images has high requirements for local details, for the comparative experimental analysis, we selected 6 high-resolution mural images (Fig. 12) and reconstructed the image details with super-resolution. Then, the reconstructed images were compared. The results are shown in Fig. 13.

As shown in Fig. 13, the traditional BI algorithm showed poor performance in superresolution reconstruction. The superresolution image reconstructed by this algorithm had clear distortions in color and brightness. These distortions did not occur when the method in the literature [17] was used for images with simple texture (Fig. 12b and f), and a satisfactory effect was achieved. Despite the achieved improvement, the method in the literature [17] did not achieve a satisfactory effect when applied to images with complex textures

(e.g., the maid's belt in Fig. 12d). Compared with the BI algorithm, the SRGAN algorithm showed a better performance. The reconstructed image did not have clear texture issues; however, the performance in the details of the reconstructed images was not satisfactory, especially for images with more texture details, such as the images in Fig. 12c and d. Although the method used in the literature [20] performed quite well in resolving the margin zigzagging problem and the reconstructed image exhibited clear contour detail, there was still much room to improve the color. The EDSR algorithm performed well in reconstructing the details of the images and therefore outperformed the first two algorithms for images with many details (Fig. 12c and d) or images with dull backgrounds (Fig. 12b, e and f). However, the reconstructed images did not perform well in terms of brightness, and the performance of the image features needed to be improved.

In summary, the BI algorithm, as an early classical superresolution method, did not satisfactorily reconstruct image features and texture details; therefore, the reconstructed images with this method were the worst. As the first superresolution GAN algorithm, the SRGAN showed performance improvement. However, the reconstructed images had an edge sawtooth. The EDSR algorithm restored some of the high-frequency details of the image, but the reconstructed image exhibited too much noise. In contrast, the reconstructed image with the algorithm proposed in this study had an improved texture, clarity and overall image performance.

Table 5 shows the PSNR and SSIM indices of the images reconstructed by the above four algorithms. As shown in this table, the algorithm proposed in this study had better evaluation indices than the previous algorithms. This finding indicated that this algorithm improved the mural representation effect to some extent and, therefore, might be helpful in improving the value of mural research.



**Fig. 13** Different mural reconstruction algorithms

**Table 5** PSNR and SSIM after reconstructing murals with different algorithms

Image	Index	BI	Literature [17]	SRGAN	Literature [20]	EDSR	ProGAN
a	PSNR/SSIM	19.18/0.43	25.31/0.66	26.46/0.69	26.41/0.67	26.90/0.72	27.11/0.73
b	PSNR/SSIM	21.64/0.45	26.20/0.62	27.25/0.67	27.20/0.68	27.40/0.68	27.85/0.70
c	PSNR/SSIM	22.57/0.37	27.03/0.52	27.45/0.54	27.49/0.53	27.47/0.55	27.53/0.55
d	PSNR/SSIM	22.76/0.40	26.96/0.52	27.57/0.55	27.61/0.53	27.66/0.56	27.72/0.56
e	PSNR/SSIM	20.23/0.37	28.91/0.79	29.03/0.79	28.97/0.77	29.23/0.81	29.38/0.81
f	PSNR/SSIM	19.62/0.39	25.21/0.63	26.94/0.69	27.08/0.68	27.13/0.71	27.38/0.72
Avg.	PSNR/SSIM	21.00/0.40	26.60/0.62	27.45/0.66	27.46/0.64	27.63/0.67	27.83/0.68

#### Superresolution reconstruction experiment for general images

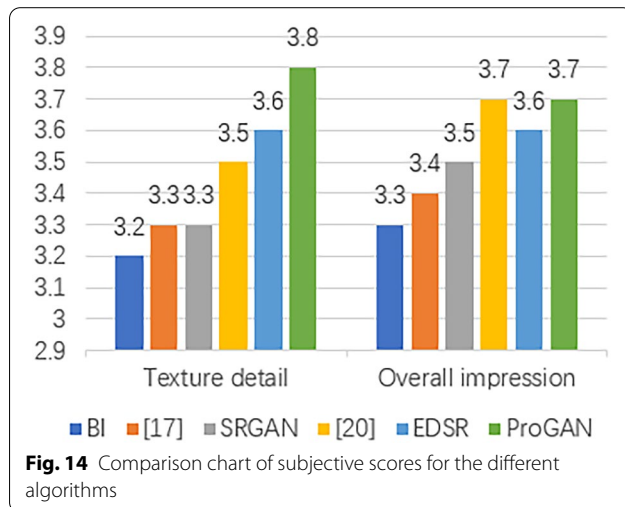
Alongside the superresolution reconstruction experiment for mural images, we also tested the performance of the proposed algorithm on general image datasets, which

included BSD100, URBAN100 and Set14. The results are summarized in Table 6.

Based on the PSNR value obtained in the experiment, the proposed model exhibited satisfactory performance on the public datasets. Compared with the VDSR, which was obtained based on deep convolution training, the

**Table 6** The PSNR (dB) results on three test datasets

Dataset	Upscaling factor	VDSR PSNR/dB	EDSR PSNR/dB	ProGAN (this study) PSNR/dB
BSD100	4	27.29	27.71	27.73
URBAN100	4	25.18	26.89	25.97
Set14	4	28.02	28.80	28.55



performance of the proposed algorithm was not lower on any of the datasets. Compared with the EDSR algorithm, it even showed some advantages.

### Subjective evaluation

In addition to using the PSNR and SSIM indices to evaluate the experimental results, we also selected 50 volunteers with normal vision to score the reconstructed images. The score interval was set from 0 to 5. The subjective evaluation indices included the overall impression and texture detail retention of the reconstructed images. Averages were obtained, and a high score indicated a better superresolution reconstruction effect. The final scores are shown in Fig. 14.

Based on the evaluation results and the intuitive feelings of the evaluators, the proposed algorithm outperformed the remaining algorithms in terms of reconstruction, including the overall appearance and texture detail. In the overall appearance, the images reconstructed by the proposed algorithm had a satisfactory performance in terms of brightness and smoothness. In evaluating the effect of fine superresolution handling for low-resolution murals, volunteers carefully compared the images before and after handling, and they all confirmed that the proposed algorithm had more advantages

in retaining the original texture details of the image and had a better overall impression. In general, the algorithm proposed in this study successfully restored the texture of the original images while preserving the overall performance, improving the research value of the related image.

### Conclusions

To address the problems of low resolution and ambiguity in ancient mural images, this study proposed a ProGAN algorithm for superresolution reconstruction of low-resolution mural images. With a GAN as the basic framework, this model integrated asymmetric pyramid modules for training, which reduced gradient explosions and disappearances during the GAN network training process. Furthermore, the perceived loss of the VGG network was introduced to train the discriminant network. Our results showed that the proposed algorithm produced a satisfactory reconstructed image in terms of texture detail and feature performance.

The proposed algorithm-generated high-resolution images of different sizes with low-resolution inputs. Although it performed well in terms of the PSNR, the structural similarity can still be improved.

Based on the results of this study, the following work should be conducted in the future:

- (1) Improve the model training to increase the structural similarity of the generated images to make the generated images more meaningful;
- (2) In the model construction, be more concise to reduce the training time; and
- (3) Because the dataset used in this study was a publicly available dataset, the pictures and murals had features and textures. Therefore, the performance of the network in reconstructing mural images was not as good as its performance in reconstructing ordinary images. In the future, we will consider further optimizing the network model to improve the ability of the network to learn image features to improve the reconstructed mural images.

### Abbreviations

GAN: Generative adversarial network; SRCNN: Superresolution convolutional neural network; SCN: Sparse coding-based network; LISTA: Learned iterative shrinkage and thresholding algorithm; LR: Low-resolution.

### Acknowledgements

None.

### Author contributions

All authors contributed to the current work. SM devised the study plan and led the writing of the article; JFC, ZXL, and ZYC conducted the experiments



and collected the data, and XHH performed the analyses. JFC reviewed the article and supervised the whole process. All authors read and approved the final manuscript.

### Funding

This study was supported by the Humanities and Social Sciences Research Project of the Ministry of Education (Planning Fund Project; Grant No., 21YJAZH002) and the Key Research Base Project of Humanities and Social Sciences in Colleges and Universities of Shanxi Province (Grant No., 20190130).

### Availability of data and materials

All data used for analysis in this study are included within the article and additional file.

### Declarations

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Computer Science & Technology, Xinzhou Normal University, No. 10 Heping West Street, 034000 Xinzhou, China. <sup>2</sup>School of Computer Science & Technology, Taiyuan University of Science and Technology, 030024 Taiyuan, China. <sup>3</sup>No. 10 Heping West Street, Xinzhou 034000, China.

Received: 17 January 2022 Accepted: 21 April 2022

Published online: 09 May 2022

### References

- Glasner D, Bagon S, Irani M. Super-resolution from a single image. 2009 IEEE 12th Int Conf Comput Vis. 2009. <https://doi.org/10.1109/ICCV.2009.5459271>.
- Dong C, Loy CC, He KM, Tang XO. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell*. 2016;38(2):295–307. <https://doi.org/10.1109/TPAMI.2015.2439281>.
- Wang Z, Liu D, Yang J, Han W, Huang T. Deep networks for image super-resolution with sparse prior. In: *Proceedings of the IEEE International Conference on Computer Vision*. USA: IEEE; 2015. p. 370–8.
- Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, et al. Real-time single image and video super-resolution using an efficient sub pixel convolutional neural network. In: *2016 IEEE Conference on Computer Vision and Recognition*. USA: IEEE; 2016. <https://doi.org/10.1109/CVPR.2016.207>.
- Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Recognition*. USA: IEEE; 2016. p. 1646–54. <https://doi.org/10.1109/CVPR.2016.182>.
- Zhang YL, Li KP, Li K, Wang LC, Zhong BN, Fu Y. Image super-resolution using very deep residual channel attention networks. In: *Proceedings of the European Conference on Computer Vision*. USA: IEEE; 2021. p. 286–301.
- Zhou FY, Jin LP, Dong J. A review of the study of reel neural networks. *J Comput Sci*. 2017;40(06):1229–51.
- Chang L, Deng XM. The cosmic neural network in image understanding. *J Autom*. 2016;42(09):1300–12.
- Luo JH, Wu JX. An overview of fine-grained image classification based on depth reuter characteristics. *J Autom*. 2017;43(08):1306–18.
- Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. *IEEE Comput Soc*. 2017. <https://doi.org/10.1109/CVPR.2017.19>.
- Liu HD, Gu XF, Samaras D. Wasserstein gan with quadratic transport cost. 2019 IEEE CVF Int Conf Comput Vis. 2019. <https://doi.org/10.1109/ICCV.2019.00493>.
- Han SS. Research on image super-resolution algorithm based on deep learning. Kaifeng: Henan University; 2018. (in Chinese).
- Yang J, Li WJ, Wang RG, Xue LX. Generative counter-super-resolution algorithm fused with perceptual loss. *J Image Graph*. 2019;24(08):1270–82.
- Su JM, Yang LX. Single-frame remote sensing image super-resolution based on generative confrontation network. *Comput Eng Appl*. 2019;55(12):202–7 + 214 (in Chinese).
- Mi H, Jia ZT. Image super-resolution reconstruction based on improved generative confrontation network. *Comput Appl Softw*. 2020;37(09):139–45.
- Wang H. Inpainting of potala palace murals based on sparse representation. *Int Conf Biomed Eng Informatics*. 2015. <https://doi.org/10.1109/BMEI.2015.7401600>.
- Ma Q. Research on super-resolution reconstruction of a single-color image based on sparse representation. Lanzhou: Lanzhou University of Technology; 2019. (in Chinese).
- Cao JF, Jia YM, Yan MM, Tian XD. Super-resolution reconstruction of murals by stably enhanced generative confrontation network. *J Syst Simul*. 2021. <https://doi.org/10.16182/j.issn1004731x.joss.20-0989> (in Chinese).
- Han M, Liu H. Super-resolution restoration of degraded image based on fuzzy enhancement. *Arab J Geosci*. 2021. <https://doi.org/10.1007/s12517-021-07218-9>.
- Xu ZG, Yan JJ, Zhu HL. Super-resolution reconstruction algorithm of mural image based on multi-scale residual attention network. *Prog Laser Optoelectron*. 2020;57(16):152–9.
- Lai WS, Huang LB, Ahuja N, Yang MH. Deep laplacian pyramid networks for fast and accurate super-resolution. *IEEE Conf Comput Vis Pattern Recognit*. 2017. <https://doi.org/10.48550/arXiv.1704.03915>.
- Tong T, Li G, Liu XJ, Gao QQ. Image super-resolution using dense skip connections. In: *IEEE International Conference on Computer Vision*. USA: IEEE; 2017. p. 4809–17.
- Huang G, Liu Z, van Der Maaten L, et al. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Recognition*. USA: IEEE; 2017. p. 4700–8.
- Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced deep residual networks for single image super-resolution. *IEEE Conf Comput Vis Pattern Recognit*. 2017. <https://doi.org/10.48550/arXiv.1707.02921>.
- Simonyan K, Zisserman A. Very deep convolutional net-works for large-scale image recognition [EB/OL]. *Comput Vis Pattern Recognit*. 2015. <https://doi.org/10.48550/arXiv.1409.1556>.
- Huang G, Liu Z, Weinberger KQ, van der Maaten L. Densely connected convolutional networks. *Comput Vis Pattern Recognit*. 2020. <https://doi.org/10.48550/arXiv.1608.06993>.

### Publisher's note

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

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)