EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Superresolution reconstruction method for ancient murals based on the stable enhanced generative adversarial network

Jianfang Cao[1,2*], Yiming Jia[2], Minmin Yan[2] and Xiaodong Tian[2]

* Correspondence: caojianfangcn@163.com
[1]Department of Computer Science & Technology, Xinzhou Teachers University, No. 10 Heping West Street, Xinzhou 034000, China
[2]School of Computer Science & Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China

Abstract

A stable enhanced superresolution generative adversarial network (SESRGAN) algorithm was proposed in this study to address the low-resolution and blurred texture details in ancient murals. This algorithm makes improvements on the basis of GANs, which use dense residual blocks to extract image features. After two upsampling steps, the feature information of the image is input into the high-resolution (HR) image space to realize an improvement in resolution, and the reconstructed HR image is finally generated. The discriminator network uses VGG as its basic framework to judge the authenticity of the input image. This study further optimized the details of the network model. In addition, three loss optimization models, i.e., the perceptual loss, content loss, and adversarial loss models, were integrated into the proposed algorithm. The Wasserstein GAN-gradient penalty (WGAN-GP) theory was used to optimize the adversarial loss of the model when calculating the perceptual loss and when using the preactivation feature information for calculation purposes. In addition, public data sets were used to pretrain the generative network model to achieve a high-quality initialization. The simulation experiment results showed that the proposed algorithm outperforms other related superresolution algorithms in terms of both objective and subjective evaluation indicators. A subjective perception evaluation was also conducted, and the reconstructed images produced by our algorithm were more in line with the general public's visual perception than those produced by the other compared algorithms.

Keywords: Superresolution reconstruction of murals, Generative adversarial networks, Dense residual block, WGAN-GP

1 Introduction

Ancient Chinese murals once boasted a glorious history. After thousands of years of accumulation and deposition over different dynasties, splendid art classics such as Dunhuang murals have emerged. Over a long period, most of the artistic research on ancient Chinese murals has been based on the discovery, exploration and summarization of traditional painting techniques. Reproducing the grandeur of these previous works has become the main direction of research. Because ancient murals were previously subject to sabotage and weathering, the original smooth and clear images have been obscured by a layer of mist, and the texture details of the murals have

© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

become blurred. Some murals have been damaged, and the existing ancient murals have sustained various types of damage to varying degrees. Therefore, it is necessary to carry out preventative protection and restoration for the extant murals. Briefly, the target of image restoration is to repair an image damaged by blur or noise, which does not change the original size of the image or increase its number of pixels, whereas image preventative protection refers to superresolution reconstruction, the focus of which is to restore the missing details, i.e., the high-frequency information, of the image. The greater the amount of information after reconstruction, the higher the resolution of the image will be. In recent years, among many computer-aided mural restoration technologies, deep learning has gradually become the main restoration technology. One important method is to carry out superresolution reconstruction of the existing ancient murals to restore the original clear images. This method can also improve the texture details of murals and promise a bright future for research in this area.

As an important image processing method in the field of computer vision, superresolution reconstruction uses a group of low-resolution (LR) images as input and generates a single high-quality, high-resolution (HR) image through a certain procedure, which can improve the recognizability and accuracy of the image. Therefore, superresolution reconstruction plays a critically important role in image-related applications [1]. Common superresolution algorithms are based on either traditional machine learning or deep learning. Traditional superresolution algorithms include interpolation-based methods [2, 3], reconstruction-based methods [4], and learning-based methods [5]. One interpolation-based method is a relatively simple algorithm proposed earlier. By calculating the registration relationship between the LR image and the desired HR image, the actual HR image is obtained by a suitable interpolation algorithm, such as bilinear interpolation [6] and bicubic interpolation (BI) [7]. However, these methods have relatively poor adaptability, can address only monotonous scenes, and are riddled with issues such as blurred edges and the loss of high-frequency details. Reconstruction-based methods, on the basis of the derived registration relationship, obtain the dependence relationship between HR and LR images as a prior and then use the prior to reconstruct the HR image. Farsiu et al. [8] proposed superresolution reconstruction of additional high-frequency information restored from multiframe images in the Fourier transform domain, thereby paving the way for the reconstruction of multiframe images. Irani et al. [9] adopted the iterative back-projection method, and Stark et al. [10] used the method of projection onto convex sets (POCS). These methods can effectively guarantee the quality of the image edges and details. However, the reconstructed images have jagged edges, and the algorithms exhibit slow convergence and entail a large number of calculations. Learning-based methods adopt a large number of high-definition, superresolution images and their corresponding LR images to continuously train the designed model so that the model can not only recover clear images but also allow images to have many high-frequency details. Chang et al. [11] proposed a superresolution reconstruction algorithm based on domain embedding, particularly locally linear embedding (LLE), by using the idea of manifold learning. This method finds the k neighbors that are closest to the input image blocks (the LR image patches) and then solves the constrained least-squares problem to obtain the weights, which are then used for reconstruction. This method has a strong dependence on samples and is prone to overfitting or underfitting. In 2008, Yang et al. [12] used sparse signals to propose a reconstruction algorithm based on sparse representation. This method requires learning and understanding much information about the relationship between HR images and LR images, establishing a complete dictionary, finding LR images and their corresponding matrix arrays and coefficients, and finally completing the reconstruction by weighting the corresponding HR matrix array. The incompleteness of the dictionary used in this method leads to low detail levels in the reconstructed image edges.

In 2014, Dong et al. [13] proposed the "superresolution using a convolutional neural network" (SRCNN) algorithm. The SRCNN algorithm was the pioneering work of deep learning in the field of superresolution. BI was used for the first time to enlarge LR images to the size of the original image. Only a three-layer convolutional network and nonlinear mapping were used to produce HR images. They interpreted the three-layer convolutional structure in three steps, i.e., image block extraction and feature representation, nonlinear mapping of the feature representation, and final reconstruction, and their method had a better effect than those of the common traditional methods. In 2016, Dong et al. [14] improved the SRCNN algorithm. They used the deconvolutional layer to enlarge the image size in the last layer and change the feature dimension and used a smaller convolution kernel and more mapping layers than did Dong et al. Furthermore, a shrinking layer and an expansion layer were added, and the data set was enhanced. Tong et al. [15] proposed a superresolution dense skip connections network (SRDenseNet) algorithm. The dense block was used to input the features of each layer into all subsequent layers so that the features of all layers were connected in series. This structure alleviated the disappearance of gradients, enhanced the transfer of features, and reduced the number of parameters. In 2016, Anagun et al. [6], using a variety of loss functions combined with the Adam optimizer, selected the loss function with the best convergence, which increased the residual module of the network and improved the model performance, and used the Charbonnier or L1 loss functions to reduce the time cost of building the model. In 2017, Zhang et al. [16] proposed a superresolution reconstruction method based on transfer learning and deep learning. This method can not only obtain high-quality, HR images but also reduce the time cost of building the model. Kim et al. [17] proposed a deep-recursive convolutional network (DRCN) by including a recurrent neural network in the overall network, which greatly reduces the number of network parameters. In 2017, Lim et al. [18] proposed the enhanced deep superresolution (EDSR) algorithm, which eliminates batch standardization, reduces the space used during training, removes the unnecessary modules in the conventional residual error network, and achieves good superresolution reconstruction results. In 2017, Ledig et al. [19] proposed the superresolution generative adversarial network (SRGAN) algorithm and applied GANs to solve the superresolution problem. They used the perceived loss and adversarial loss to alleviate the loss of high-frequency details and image smoothness so that images could give good visual perception. However, the peak signal-to-noise ratio (PSNR) obtained based on this strategy was not sufficiently high. In 2017, Arjovsky et al. [20] used the Wasserstein GAN (WGAN) to solve the issue of instability during GAN training; they used the Wasserstein distance instead of JS divergence to measure the distance between the real image and the reconstructed image. Due to the weight pruning strategy, however, there are still issues with this method, such as gradient explosion and disappearance. In 2018, Wang et al. [21]

improved the SRGAN algorithm. They used dense residual blocks for feature extraction from the generative network with the relativistic discriminator as the discriminator. The generator could generate more real texture details and retain more image feature information than could the original SRGAN algorithm. Although all the abovementioned deep-learning-based algorithms have improved the reconstruction results compared with traditional machine learning algorithms, their effects are unsatisfactory when directly applied in mural data set processing. Additionally, during the network model training process, they become unstable. Due to the unclear texture details, lost smoothness and undesired reconstruction effects of the reconstructed images, their research value is low.

Based on the abovementioned literature review, we propose a stable enhanced SRGA N (SESRGAN) algorithm in this study, which is based on the GAN. The main contributions of this study are as follows:

1. A new SESRGAN model is proposed. This model expands the deep learning algorithm and may have reference value for the stable training of other GANs. It utilizes the GAN as its framework and integrates the residual dense blocks with residual scaling (RS-RDB) structure in the generative network to fully capture the feature information of the image and therefore to increase the generalization capacity and robustness of the model. For perceived loss calculation, the SESR GAN model uses the feature information before activation and utilizes the WGAN-gradient penalty (WGAN-GP) to counter loss, thereby enhancing its training stability.
2. The proposed SESRGAN model is applied in the superresolution reconstruction of ancient mural images and improves the overall esthetic and artistic values of the images. In this study, the proposed model completes the superresolution reconstruction of ancient mural images based on the idea of deep learning. It provides a new technical route for ancient mural image restoration, breaks through the existing technical bottleneck of ancient mural digital protection, and provides a technical demonstration of the image content restoration of similar information in the field of cultural heritage digitization.

2 Methods

2.1 Relevant theories

2.1.1 Generative adversarial networks

The GAN [22] has become a popular deep learning model in recent years and is one of the most promising methods for unsupervised learning with complex distributions. The generative network first captures the distribution of random noise in an image and then uses the distribution of this noise to generate a sample similar to the real data as the input of the discriminator network, which functions to estimate the probability that a sample comes from the training data to judge whether the input data come from the real data or the generated sample. During the process of training the network, the generative network continuously improves its ability to generate real samples to deceive the discriminator network, while the discriminator network continuously improves its ability to discern the authenticity of samples. Through continuous learning in the mutual game between the generative network and the discriminator network, the model is optimized. As time elapses, the generative network and the adversarial network are continuously trained and optimized, and finally, the two networks reach a dynamic equilibrium: the generated sample approximates the distribution of the real sample, and the discrimination probability of the discriminator network for a given sample is 0.5. The loss formula of the GAN is as follows:

$$\min_{G} \max_{D} V(D,G) = E_{\mathrm{x}\sim \mathrm{P}_x}[\ \log D(x)] + E_{\mathrm{z}\sim \mathrm{P}_{\mathrm{z}}(\mathrm{z})}[\ \log(1 - D(G(z)))] \qquad (1)$$

where X represents the real image data, z represents the noise image data, P_{X} represents the probability distribution of the real image data, $p_{\mathrm{z}}(z)$ represents the probability distribution of the generated data, $G(z)$ represents the reconstructed image data, and $D(x)$ represents the probability that the discriminator network correctly judges whether the image data are real or fake. $D(G(z))$ represents the probability that the discriminator network correctly judges whether the reconstructed image data are real or fake, log $D(x)$ represents the judgment of the discriminator network on the real image data, and log $(1 - D(G(z)))$ represents the judgment on the reconstructed image data. The structure of the generative adversarial network is shown in Fig. 1.

2.1.2 Residual network and dense network

The residual network (ResNet) [23] addresses the fact that the gradient is prone to diffuse or disappear and that the training accuracy and test accuracy decrease due to the saturated and degraded network performance when the convolutional neural network becomes deeper. The ResNet structure is shown in Fig. 2. Suppose that the output of the previous layer is the input x of the current layer; then, the residual of the current layer is $F(x)$, and the output of the next layer is $H(x)$. x is passed on to the next layer $H(x)$ by means of a skip connection. Then, $H(x) = F(x) + x$. The residual of this layer can be expressed as $F(x) = H(x)-x$. This residual structure can improve the performance of a relatively deep network [24]. Therefore, this paper introduces the residual block as a part of the generative network.

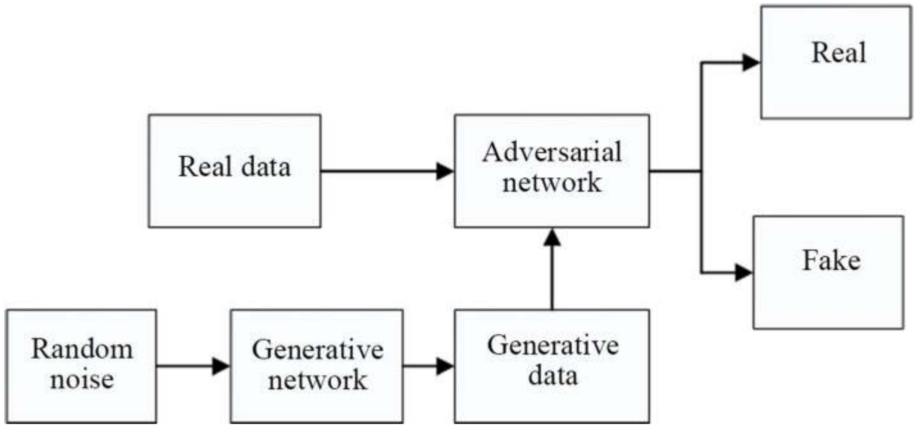

Fig. 1 Basic structure of the GAN

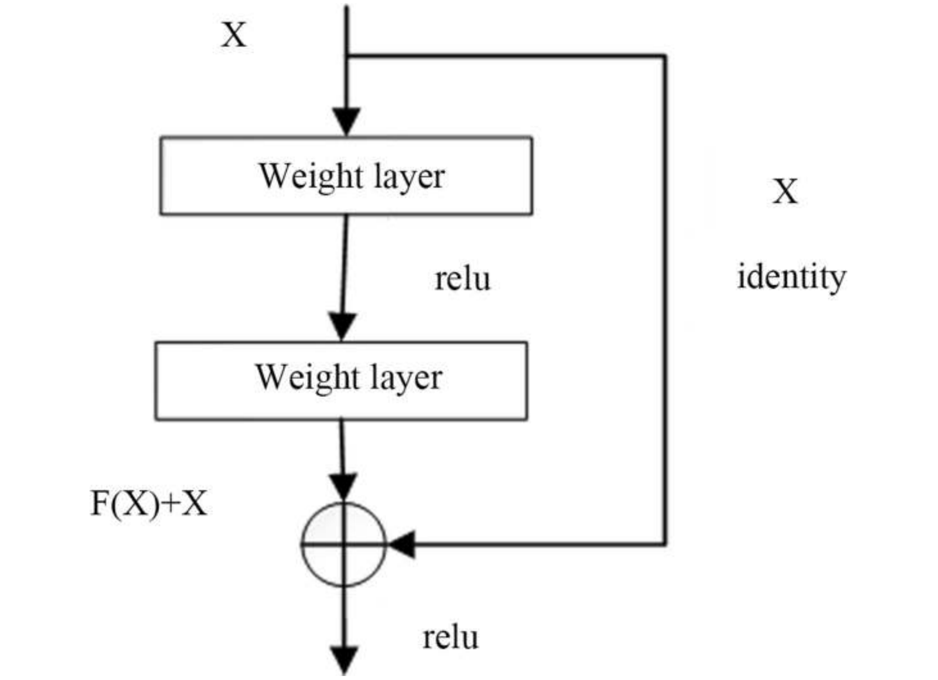

Fig. 2 Structure of the ResNet

The dense network (DenseNet) [25] commits itself to improving the performance of a network from the perspective of feature reuse. It is a convolutional neural network with dense and close connections between any two layers. Each layer of the DenseNet receives input from all previous layers and then outputs its feature mapping to all subsequent layers, thereby improving the transmission efficiency of information and gradients in the network. The merits of this structure are that it can realize the reuse of features, alleviate the disappearance of gradients, strengthen feature transmission, and substantially reduce the number of parameters. The DenseNet structure is shown in Fig. 3.

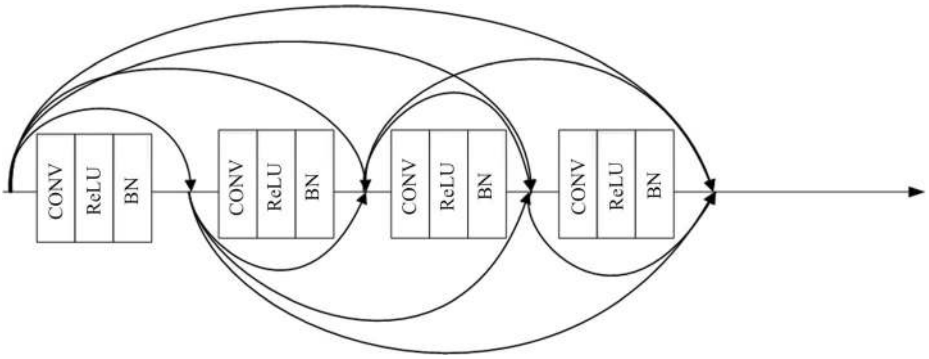

Fig. 3 Structure of the DenseNet

2.1.3 WGAN-GP

The WGAN-GP theory [20, 26] was proposed to solve problems with the WGAN, such as the difficulty in training and slow convergence speed in the real experimental process. Compared with the conventional GAN, its improvement in terms of effects is not obvious. Therefore, the WGAN-GP approach pinpoints the root cause of WGAN defects: the WGAN directly adopts weight clipping in the process of addressing the Lipschitz limitation condition. Every time the parameters of the discriminator are updated once, the condition is then checked to see whether the absolute values of all parameters in the discriminator exceed a threshold n. If any exceed the threshold, these parameters are clipped back to the range of $[-n, n]$. By ensuring that all parameters of the discriminator are bounded during the training process, the approach ensures that there is not a large difference in the discrimination of two slightly different samples by the discriminator; therefore, the Lipschitz condition is realized indirectly. However, this leads to most weights approximating the two extremes. Therefore, the WGAN-GP approach uses a gradient penalty and adopts the Adam optimizer to replace the RMSprop optimizer, which increases Gaussian noise in the generated image. Compared with the WGAN, this method can train the GAN model more stably, requires almost zero repeated adjustments of the superparameters, can enable the model to converge quickly, and can generate images of good quality.

2.2 Stable enhanced generative adversarial network algorithm

The designed GAN can better reconstruct HR images than can other networks. Its architecture is shown in Fig. 4. The generative network takes an LR mural image as input, extracts features through the dense ResNet, reconstructs the image through upsampling and convolution, and outputs the HR mural image. The HR and real HR are input into the discriminator network together. Finally, the discriminator network is responsible for determining whether the input images are real or fake.

2.2.1 Residual dense blocks with residual scaling

To extract as many features of the murals as possible and restore the image quality well, this study makes the following adjustments to the generative network: residual dense blocks with residual scaling (RS-RDB) is used to replace the original residual

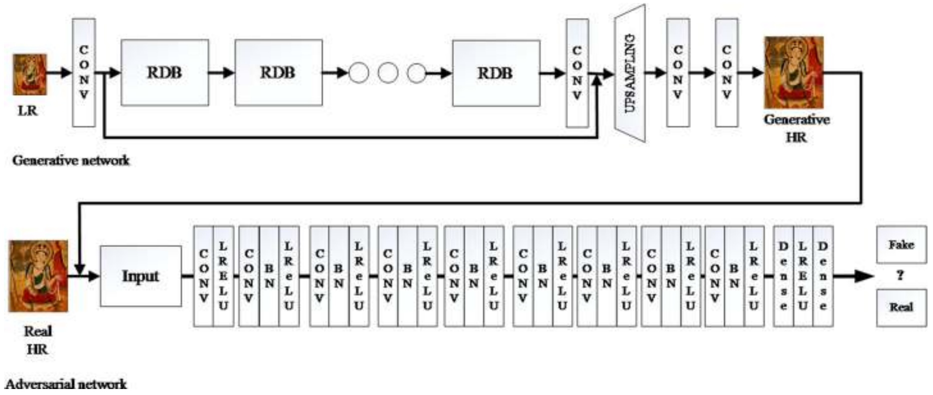

Fig. 4 Stable enhanced generative adversarial network structure

block (RB) for extracting the deep features of the input image. Since increased numbers of layers and connections always improve the performance of a network, the combination of residual networks and dense connections increases the depth of the network used in this study. The structure of the RS-RDB is shown in Fig. 5. To extract the maximum number of image features and prevent overfitting caused by overly deep networks, we use 23 RDBs in this study, which are helpful for restoring the image texture and eliminating noise. One RDB is composed of 3 dense blocks, and residual scaling is introduced. The residual is multiplied by a numerical value β in (0, 1) to increase the stability of the trained deep network. Among them, the dense block is composed of 4 convolutional layers and 4 leaky ReLU layers. For the sake of the consistency, stability and generalizability of the model, the BN layer in the dense block module is removed to reduce the computational complexity and memory usage. For different PSNR-based tasks, such as superresolution and deblurring, removal of the BN layer has been demonstrated to improve the effect of the model and reduce its computational complexity; the BN layer is prone to generating undesired artifacts, thereby limiting the generalization ability of the model [27]. The dense block is shown in Fig. 6. The input image features are extracted through the abovementioned residual dense blocks, and then HR images are generated. Two subpixel convolutional layers (pixel shufflers) are used to enlarge the size of the image, and finally, a 3*3 convolutional layer is used to output a 3-channel HR image.

2.2.2 Design of the discriminator network

The discriminator network first uses a 64-channel convolutional layer to extract shallow features from the input image and then uses 8 convolutional layers, each of which contains a BN layer. Leaky ReLU is used as the activation function. Because the WGAN-GP approach uses the Wasserstein distance instead of JS divergence to measure the distance between the real sample and the generated sample, the task corresponds to a regression model. Therefore, the following adjustments are made in this study: The sigmoid activation function designed by the original discriminator network is not used. Instead, two fully connected layers are used to directly output the real probability of the image. Details about the discriminator network are shown in Table 1.

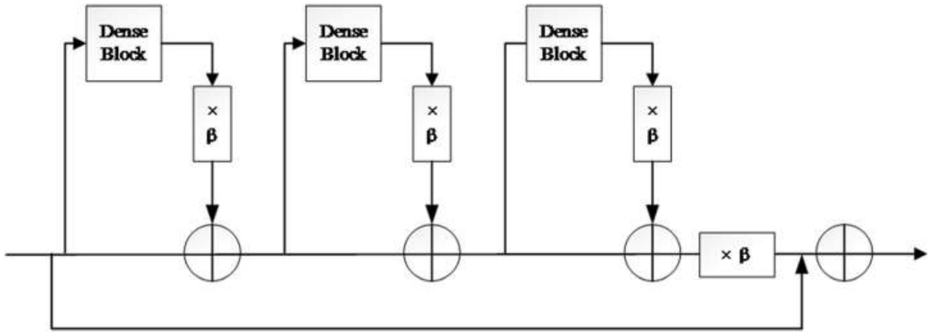

Fig. 5 Structure of an RDB with residual scaling

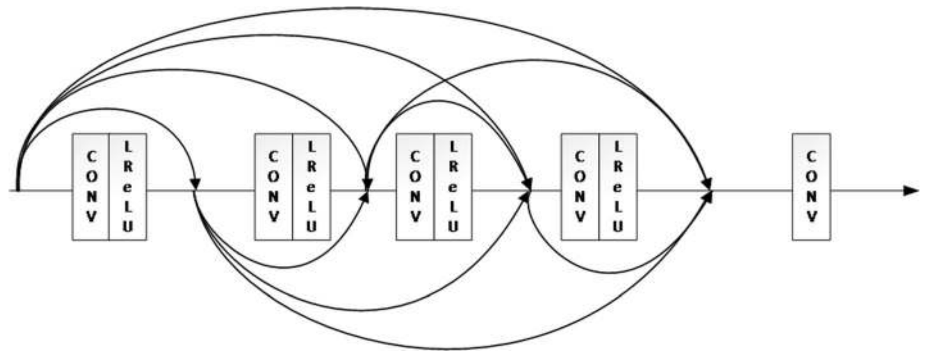

Fig. 6 Structure of the dense block

2.2.3 Loss function

A loss function is used to judge whether the image reconstructed by the model is good or not. To generate a genuine visual effect for the reconstructed image, the perceptual loss and content loss are used to optimize the generative network, and the adversarial loss is used to optimize the adversarial network. The final calculation formula for the loss function is as follows:

$$l_G = l_{MSE}^{SR} + \lambda_1 l_{VGG}^{SR} + \lambda_2 l_{adv} \tag{2}$$

where λ_1 and λ_2 represent the coefficients used to balance the different losses.

First, the content loss is introduced. To ensure the accuracy of the reconstructed image and the original image information, the mean squared error (MSE) loss is taken as the content loss of the generative network. The spatial error between the generated image pixels and the real image pixels is calculated to determine the pixel-level loss. The formula for determining the pixel-level MSE loss is:

$$l_{MSE}^{SR} = \frac{1}{r^2 WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left(I_{x,y}^{HR} - G_{\theta_G}\left(I^{LR}\right)_{x,y}\right)^2 \tag{3}$$

Then, the perceptual loss is introduced. The MSE loss enables the reconstructed image to have a very high PSNR. However, high-frequency content is absent in the

Table 1 Details of the discriminator network

Name	Type	Kernel	Stride	Padding	Output
Conv0_0	conv	3 × 3	1 × 1	1	64
Conv1_0	conv	4 × 4	2 × 2	1	64
Conv1_1	conv	3 × 3	1 × 1	1	64
Conv2_0	conv	4 × 4	2 × 2	1	128
Conv2_1	conv	3 × 3	1 × 1	1	128
Conv3_0	conv	4 × 4	2 × 2	1	256
Conv3_1	conv	3 × 3	1 × 1	1	256
Conv4_0	conv	4 × 4	2 × 2	1	512
Conv4_1	conv	3 × 3	1 × 1	1	512
FC0	FC	–	–	–	100
FC1	FC	–	–	–	1

image, which blurs the details of the image. In the generative network, the perceptual loss in the VGG [23] network is also introduced, and the feature information before activation (instead of after activation) is used for calculation purposes. To calculate the VGG loss, the generated HR image and the real HR image are input into the VGG19 network for feature extraction, and then the RMSE is used to calculate the Euclidean distance on the extracted feature map. The calculation formula for the VGG is:

$$l^{SR}_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}\left(I^{HR}\right) - \phi_{i,j} G_{\theta_G}\left(I^{LR}\right)_{x,y} \right)^2 \tag{4}$$

where $\phi_{i,\ j}$ represents the feature map obtained between the jth convolution (before activation) and the ith pooling layer of the VGG19 network. Then, we define the VGG loss as the feature of the reconstructed image, representing the Euclidean distance between G_{θ_G} (I^{LR}) and the reference image I^{HR}. $W_{i,\ j}$ and $H_{i,\ j}$ are the dimensions of the corresponding feature map in the VGG network.

Finally, the adversarial loss is introduced. To make the model training process more stable, we introduce the WGAN-GP [14] method proposed by researchers at Monterey University to further improve the objective function of the model on the basis of the WGAN. The gradient penalty is applied to the discriminator network. The formula for calculating the loss function of the WGAN is:

$$L(D) = -E_{x\sim p_g}[D(x)] \tag{5}$$

The added Lipschitz condition is:

$$\|\nabla_x D(x)\| \leq K \tag{6}$$

When the input sample does not change significantly after the input sample changes, K is generally set to 1. Through the weighting and merging of the original discriminator loss, the formula to calculate the adversarial loss of the WGAN-GP discriminator can be derived as:

$$l_{adv} = -E_{x\sim p}[D(x)] + E_{x\sim p_g}[D(x)] + \lambda E_{x\sim p_{\hat{x}}}\left[\left(\|\nabla_x D(x)\|_p - 1\right)\right]^2 \tag{7}$$

2.3 Training process of the SESRGAN algorithm

The specific process of the superresolution generative adversarial network (SESRGAN) algorithm proposed in this study is described as follows:

Input: the LR image data set and the corresponding HR image data set.

Output: the generative network G and the adversarial network D.

Step 1: Feed the LR image to the generative network G and output the generative HR image HR_g obtained after reconstruction.

Step 2: Calculate the MSE between the HR image and HR_g image and update the parameters of the generative network.

Step 3: Repeat step 1–step 2 n_1 times to obtain and save the pretrained generative network model l_{adv}.

Step 4: Input the LR image into l_{adv} and output the generated HR image l_{adv}.

Step 5: Input HR_1 and the corresponding HR into the adversarial network D, calculate the adversarial loss l_{adv}, and update the parameters of the adversarial network D.

Step 6: Input HR_1 and the corresponding HR into the pretrained VGG network, and then use the eigenvalues before activation to calculate the perceptual loss l_{VGG}.

Step 7: Calculate the content loss l_{MSE} between HR_1 and the corresponding HR.

Step 8: Calculate the total loss l_g, and update and save the generative network G and the adversarial network D.

Step 9: Repeat step 3–step 8 n_2 times, and continuously update and save the generative network G and the adversarial network D.

The training process of the SESRGAN algorithm is shown in Fig. 7.

2.4 Experimental environment and design

The hardware environment set up for the experiment was as follows: an Intel Corei7-7700K CPU, 16 GB of memory, and an NVIDIA GeForce GTX1080Ti graphics card. The software environment set up in the experiment was as follows: CUDA version 9.0, cuDNN version 7.0, and the Windows 10 operating system. Python 3.6 and the PyTorch framework were used to write the experimental test. The software compiler was pycharm2019_3.1_x64.

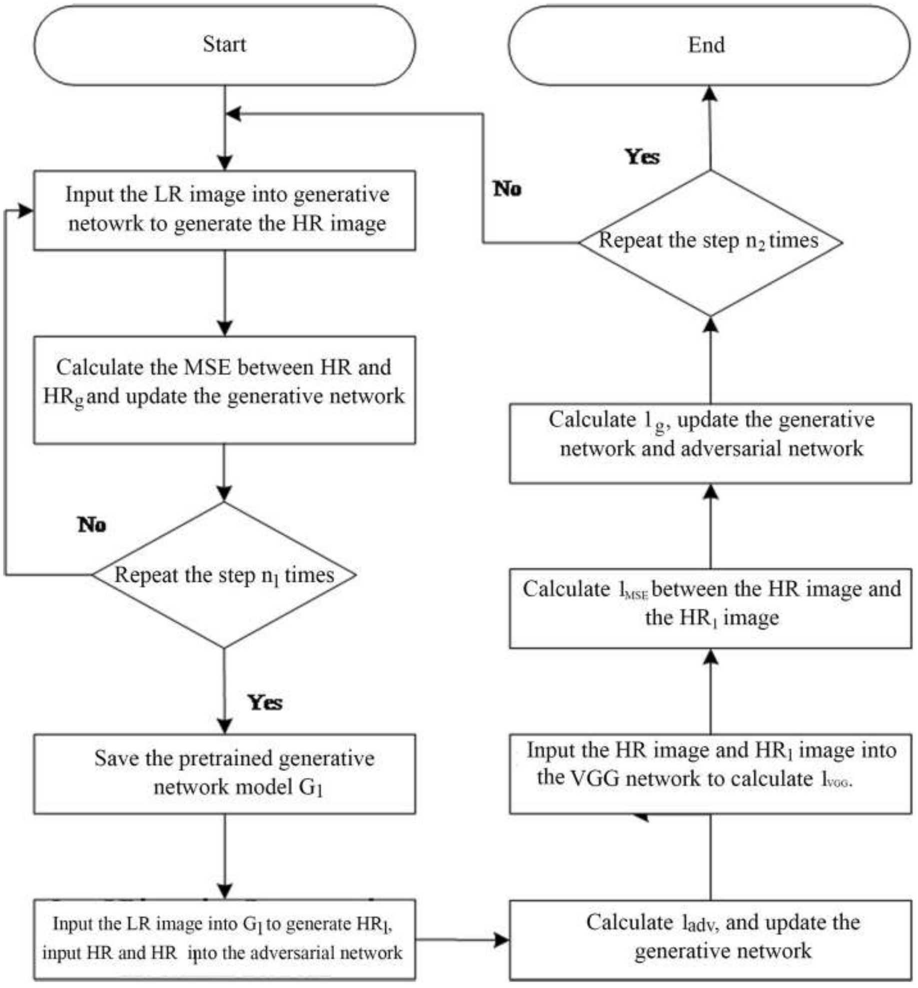
Start
End
Input the LR image into generative netowrk to generate the HR image
Calculate the MSE between HR and HR_g and update the generative network
No
Repeat the step n_1 times
Yes
Save the pretrained generative network model G_1
Input the LR image into G_1 to generate HR_1, input HR and HR ipto the adversarial network
Calculate l_{adv}, and update the generative network
Input the HR image and HR_1 image into the VGG network to calculate l_{VGG}.
Calculate l_{MSE} between the HR image and the HR_1 image
Calculate l_g, update the generative network and adversarial network
Repeat the step n_2 times
Yes
No

Fig. 7 Training flow of the SESRGAN algorithm

The training data set used in the experiment consisted of 800 DIV2K images, 2650 Flickr2K images, and 90 mural images. The test data set was made up of 30 mural images. The mural data set in this study consisted of murals of different styles and types. First, the data set was expanded. The data were enhanced by flipping them and rotating them 180°, as shown in Fig. 8. The downsampling factor for the HR and LR images was 4. To increase the input and output speeds during training, the LR images were clipped to a size of 120*120, and the HR images were clipped to a size of 480*480; all the images were then fed into the model. The optimizer used was the Adam optimizer, with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The initial learning rate was set to 0.0001, and the attenuation parameter of the learning rate was 0.5. The residual scaling strategy was introduced in the generative network model, where $\beta = 0.2$. This study, incorporating the idea of second transfer learning proposed by Yuan [15], used the DIV2K data set to pretrain a

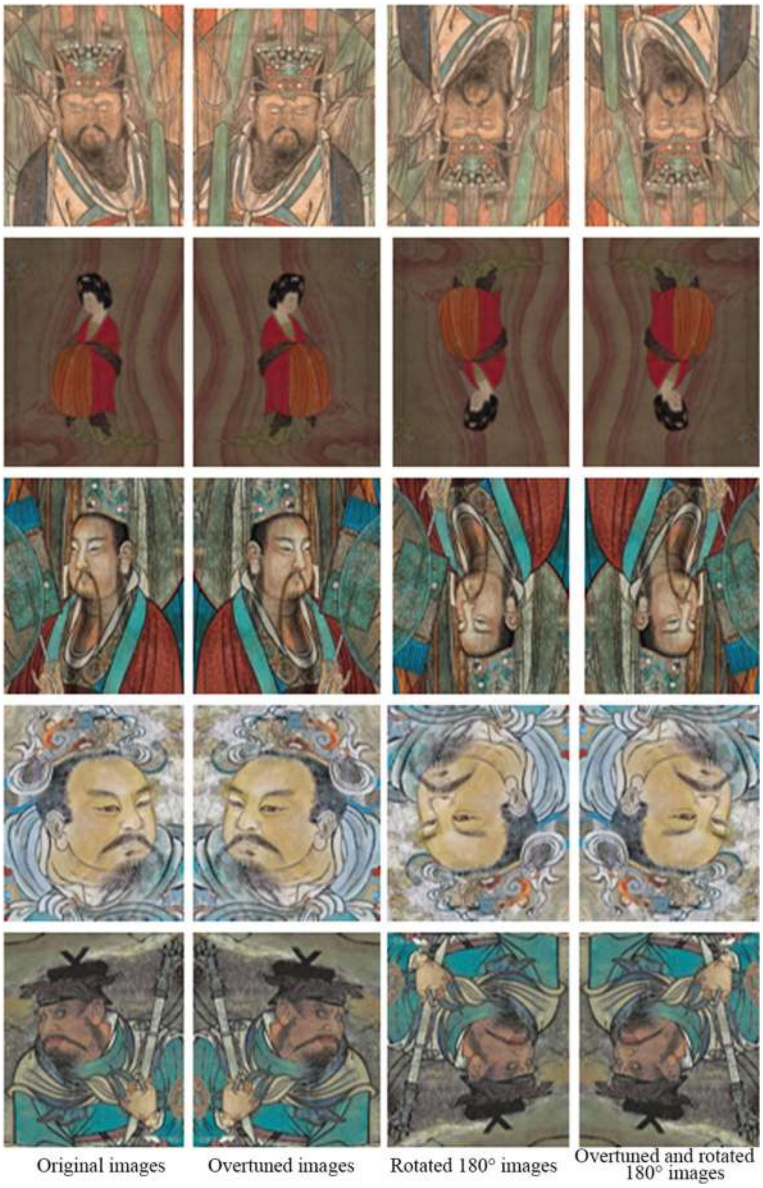

Fig. 8 Enhanced ancient murals

generator on the generative network, which was used to initialize the parameters to obtain high-quality images and fast convergence speeds. Then, the Flickr2K images were fed into the network for training, and finally, the mural data set was fed into the network to complete the training process. The generator and discriminator models were updated alternatively, and finally, the experimental results were obtained and analyzed.

2.5 Evaluation indexes

To demonstrate the effectiveness of our algorithm, we use both reference and no-reference indexes to assess the quality of the reconstructed image, with the former type including the PSNR, structural similarity index method (SSIM), multiscale structural similarity index method (MS-SS), and Firchet inception distance (FID) and the latter including the natural image quality evaluator (NIQE). A higher PSNR value (unit: dB) between the two images indicates less distortion between the reconstructed image and the HR image, that is, a higher-quality reconstructed image; the SSIM reflects the similarity of the reconstructed image to the real image in terms of the brightness, contrast and structure. The closer the SSIM is to 1, the higher the similarity between the two images, indicating that the generated image better conforms to the visual perceptual effects of the public. MS-SSIM is the average obtained from the calibration procedures of image assessments with different resolution scales. IS is the scoring system for the generated image in terms of diversity and quality using the training inception-v3 network. A high IS score indicates better diversity and quality of the generated image. FID is a measure of the distance between the feature vector of the authentic image and that of the generated image, based on which the similarity between the authentic image and the generated image is evaluated. A lower FID means a higher similarity. The NIQE does not require distorted images to be visually scored for training. Instead, after calculation of the locally mean subtracted contrast normalized (MSCN) image, part of the image blocks are used as training data according to the local activity. The NIQE takes model parameters that are obtained based on a generalized Gaussian model fitting as the features. It adopts a multivariable Gaussian model to describe the features and assesses the quality of the generated image based on the distance between the model parameter of the image and the pre-established model parameter. A lower NIQE score indicates a higher perception quality of the image.

The subjective evaluation standard commonly used at present is the mean opinion score (MOS), which grades images using 5 levels (bad, poor, fair, good, and excellent). To obtain the MOS, we invited 50 evaluators to grade each algorithm from 1 to 5 points on the basis of the overall perception effects and local details/texture of the image, and then we counted the results of each evaluator’s score of the image and computed the average score, which was taken as the MOS to evaluate the performance of each algorithm.

3 Results and discussion

3.1 Model training loss and analysis

The loss function is a tool to measure the quality of the architecture of a network. During the training process, the SESRGAN algorithm realizes gradient optimization by continuously updating and calculating each parameter. After training the model 160 ×

10^3 times, the parameters achieve their optimal values, thereby minimizing the loss function. The various losses are shown in Fig. 9, where Fig. 9a is the content loss, Fig. 9b is the perceptual loss, Fig. 9c is the adversarial loss, and Fig. 9d is the total loss.

In Fig. 9, the content loss is stable between 4×10^{-4} and 5.5×10^{-4}, indicating that the pixels of the generated images are sufficiently similar to the pixels of the real images, with no excessive pixel errors. The perceptual loss steadily decreases within a range of 0.8–1.2 and finally stabilizes at approximately 0.9, indicating that the error between the feature extraction of the generated image and the feature extraction of the real image is small and demonstrating that the generative model can extract image features well. The adversarial network is stable within a range of 0–0.5. The discriminator loss periodicity begins at a high level and then declines because the discriminator is locked when the generator is being trained, causing the capability of the discriminator to decrease. Then, the generator is locked. The discriminator is continuously optimized during training so that the loss is reduced, indicating that the discriminator finally becomes sophisticated enough that it can successfully judge whether an image is a generated image or a real image. The total loss of the generator is stable within a range of 0.03–0.04, indicating that the generator is relatively mature after training and can successfully generate the reconstructed HR image from the corresponding LR image.

3.2 Analysis of the murals reconstructed by the SESRGAN algorithm

To verify the effectiveness of the algorithm in reconstructing the ancient murals, 4 images were randomly selected from the reconstructed 4-fold HR mural images and compared with the corresponding LR mural images, as shown in Fig. 10.

As shown in Fig. 10, the texture details of each LR ancient mural image after reconstruction by the algorithm in this study are relatively ideal. The reason for this is that a deep network was used to extract high-frequency features, and a pretrained VGG network was used to extract feature information before activation rather than after activation. An observation of the LR images is that those that retain more details are clearer after reconstruction than those that retain fewer details, and these images show good consistency in terms of brightness and contrast. In addition, it is clear that after the reconstruction of LR images that retain few details, the restored images are relatively complete, and they are not inferior to other reconstructed mural images in terms of details. As shown in Table 2, the reconstructed images exhibited satisfactory scores in terms of both the reference and no-reference assessment indicators, which indicates that the algorithm proposed in this study exhibits excellent stability.

3.3 Comparison experiment and analysis

To ensure the desired effect of this experiment, one mural image from each of six styles of mural images was selected as a representative for comparison, as shown in Fig. 11, in which Fig. 11a is the Yonglegong mural, Fig. 11b is the Baigong mural from the Qing dynasty, Fig. 11c is the Chaoyuan mural, Fig. 11d is the Lushui Temple mural, Fig. 11e is the Landscape mural, and Fig. 11f is the Life mural from the Ming dynasty. The BI algorithm in reference [7], the SRGAN algorithm in reference [19], and the ESRGAN algorithm in reference [21] were employed for the comparison experiment. To observe

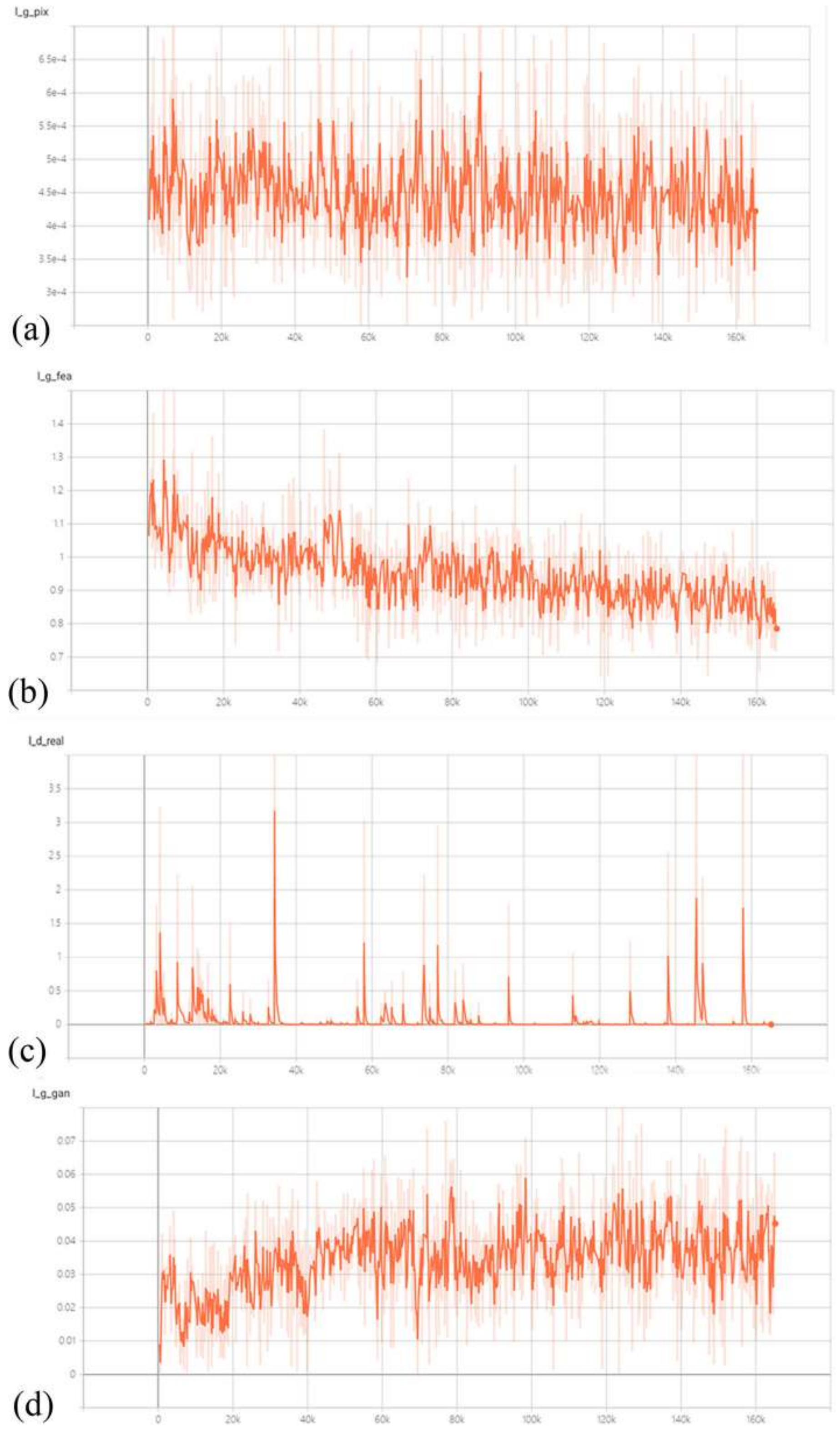

Fig. 9 Four different losses incurred during network training. **a** Content loss. **b** Perceptual loss. **c** Adversarial loss. **d** Total loss

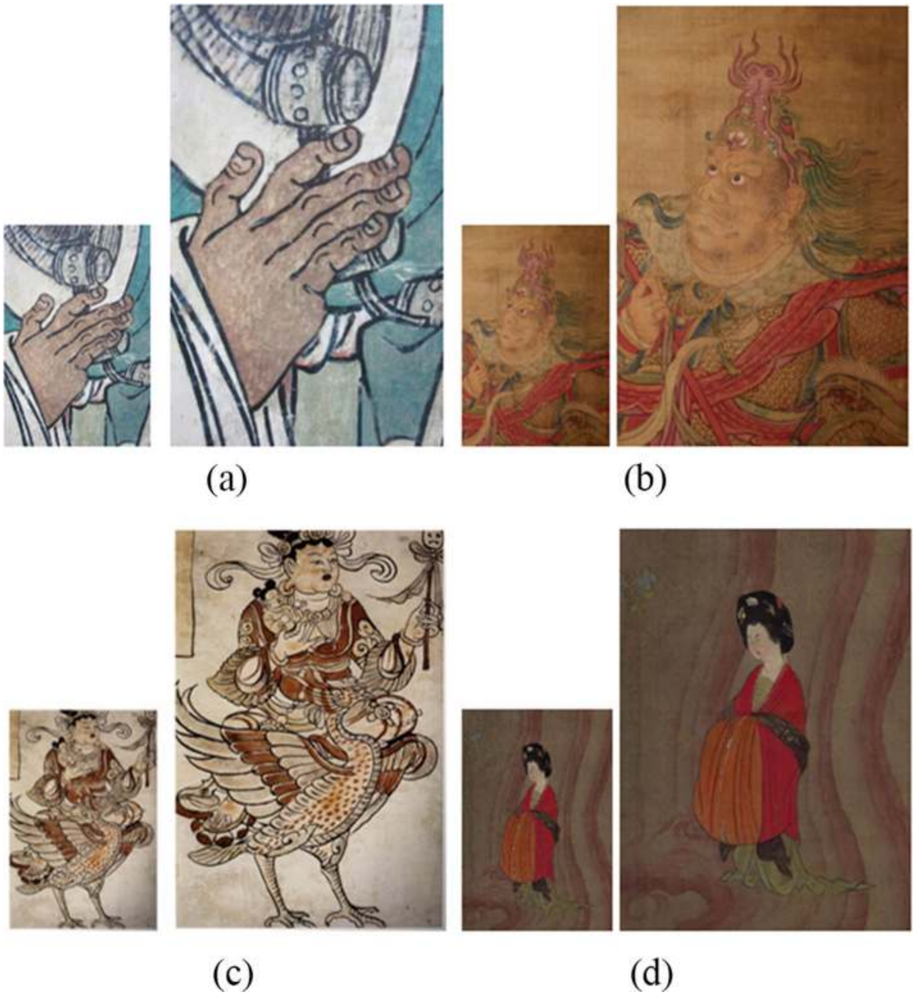

Fig. 10 Reconstruction display effects of ancient murals

the contrast effect clearly, the locally reconstructed and enlarged details of these 6 murals were selected, as shown in Fig. 12.

As shown in Fig. 12, for the Yonglegong mural and Baigong mural, each algorithm restores details well. However, the BI and SRGAN algorithms do not perform as well as the algorithm in reference [28], the ESRGAN algorithm and the SESRGAN algorithm proposed in this study in terms of brightness. For the Chaoyuan mural, an examination of the face of the figure in the mural reveals that the image reconstructed by the SESR GAN algorithm is smoother and that the noise elimination effect of the proposed

Table 2 Different mural evaluation indicators

Indicator	*a*	*b*	*c*	*d*
PSNR	28.71	32.03	30.04	32.11
SSIM	0.876	0.782	0.851	0.768
MS-SSIM	0.868	0.729	0.874	0.770
IS	15.5	20.1	16.3	19.7
FID	12.3	11.9	13.4	10.8
NIQE	6.162	2.575	3.984	2.639

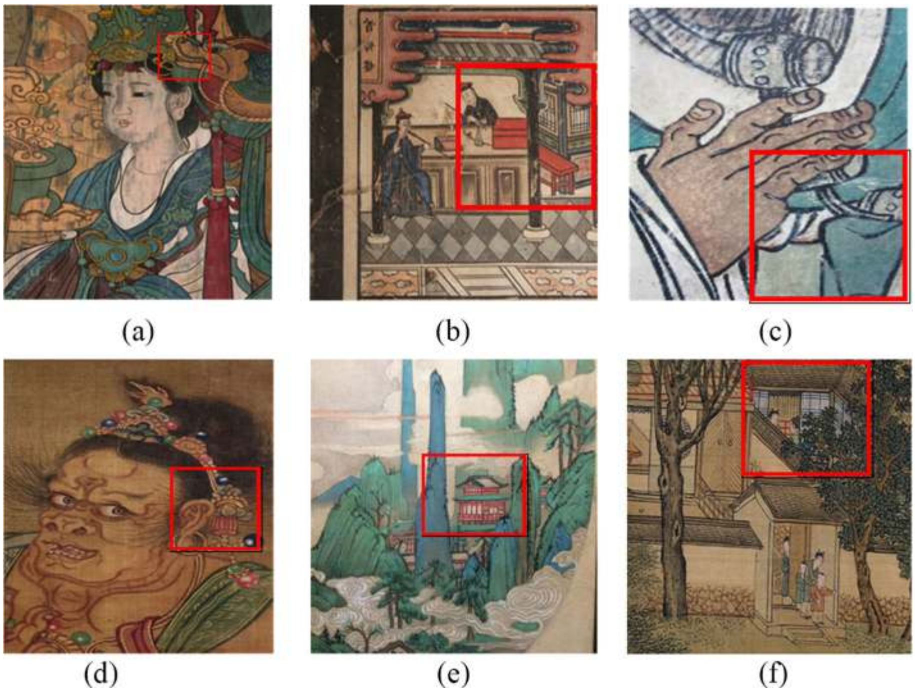

Fig. 11 Ancient murals in different styles. **a** Yonglegong mural. **b** Baigong mural. **c** Chaoyuan mural. **d** Lushui Temple mural. **e** Landscape mural. **f** Life mural

algorithm is better than those of other algorithms. In the Lushui Temple mural, by observing the accessories of the mural portrait, one can find that the texture details of other algorithms are missing, while the algorithm in this study retains these texture details. In the Landscape and Life murals, the BI algorithm does not take the feature information of the entire image into account, resulting in a lack of high-frequency details. The feature information extracted by the SRGAN algorithm is insufficient because it adopts a rather shallow network. Therefore, their details and textures are poorer than those extracted by the algorithms in references [28–30], the ESRGAN algorithm, and the algorithm in this study. The ESRGAN algorithm and that in reference [30] have an excellent restoration effect. However, they introduce some unpleasant noises. In summary, the reconstruction effect of the SESRGAN algorithm is superior to those of the other algorithms in terms of image details. The BI algorithm is inferior to the other algorithms in terms of brightness and detail restoration for various styles of mural images. The SRGAN algorithm restores part of the high-frequency information. Because both the BI and SRGAN algorithms use few network layers, some high-frequency details are not learned, and the edges are seriously sharpened. The overall effect of the ESRGAN algorithm is better than those of the BI and SRGAN algorithms in terms of the restoration of details, but it adds unpleasant artifacts and noise information. Although the reconstruction effects of the algorithms in references [28–30] are satisfactory, they do not perform as well as the SESRGAN algorithm in terms of local details. Therefore, the algorithm proposed in this study is improved to a certain extent in terms of texture details, overall brightness, and resolution compared with the other algorithms.

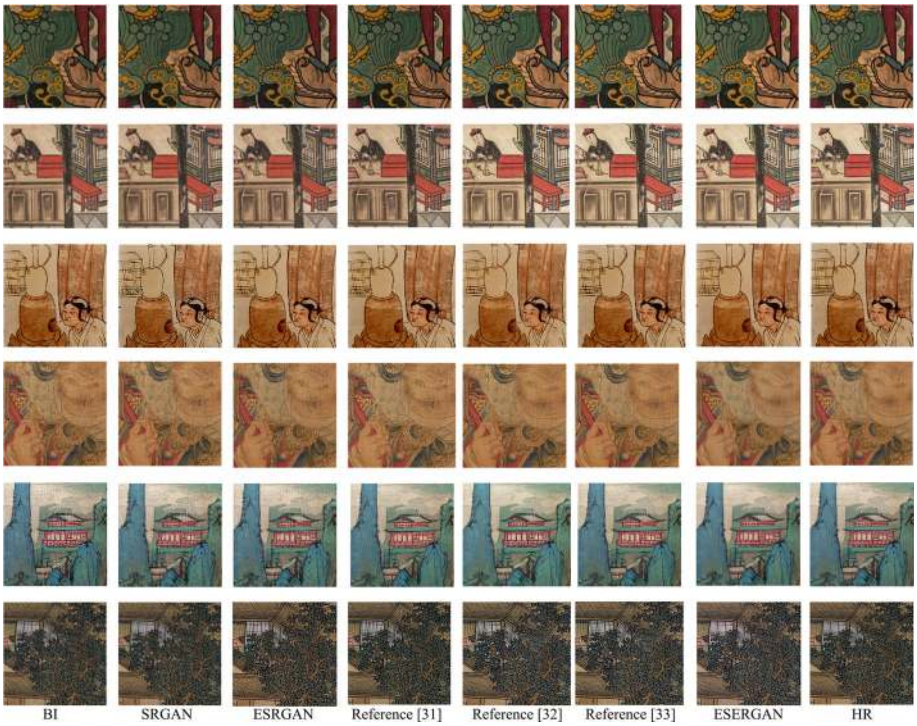

Fig. 12 Comparison of the superresolution reconstruction effects of different mural images obtained with different algorithms

Figure 13 shows the subjective scores given by the 50 evaluators to the reconstruction effects of different styles of ancient murals in Fig. 12. After evaluation and discussion by the evaluators, it was unanimously agreed that the overall perception of the algorithm proposed in this study conformed more to human visual perception than that of the other algorithms, and it achieved better effects in terms of brightness and image smoothness than did the other algorithms. In terms of local texture details, the evaluators believed that the reconstruction effect achieved by this algorithm was more desirable, and more accurate details were restored in LR images than those obtained by the other algorithms. Through comparison, it is shown that this algorithm is also superior to the other algorithms in terms of the overall perception and local detail texture. Therefore, the proposed algorithm also gives a more persuasive outcome than do the other algorithms in terms of subjective scoring.

To validate the superiority of the SESRGAN model, the averages in terms of the considered assessment indicators that are obtained by different algorithms on the mural data sets are calculated, and the results are summarized in Table 3.

As shown in Table 3, the SESRGAN overperforms all other algorithms in terms of the PSNR, SSIM, FID, and NIQE. However, its performance in terms of the MS-SSIM is poorer than that of the algorithm in reference [29], which may be attributed to more detailed features being extracted by the algorithm used in reference [29] because of the adoption of layered RBs. Therefore, in the future, layered feature extraction will be incorporated into the proposed algorithm to obtain more subtle mural texture features. In addition, the SESRGAN exhibits excellent performance in terms of the IS, being second only to the ESRGAN. In summary, because the SESRGAN adopts the RS-RDB

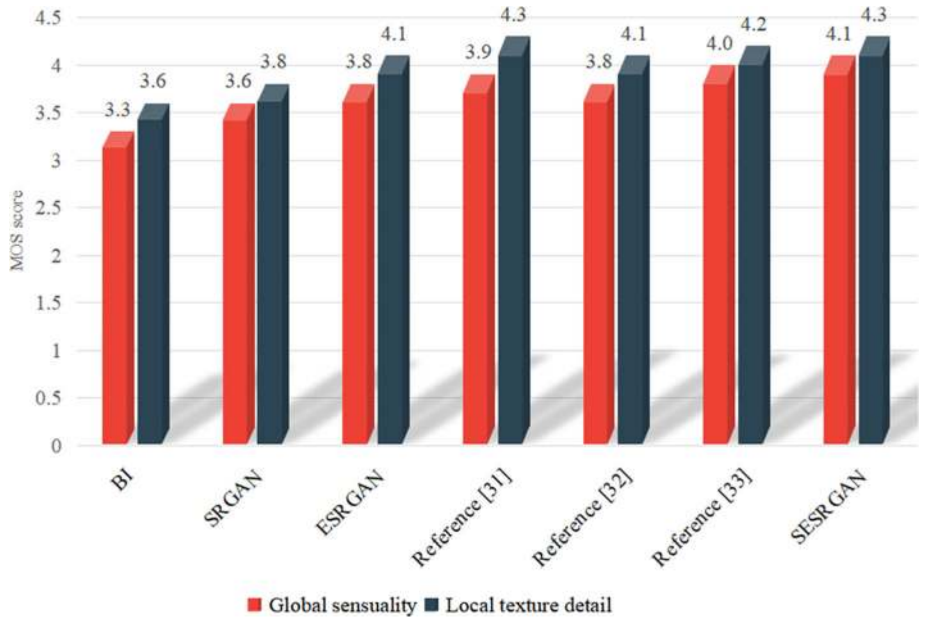

Fig. 13 Subjective scoring of the reconstruction effects of the ancient murals of different styles in Fig. 12

structure in a generative network to fully extract features and calculates the perceived loss before rather than after activation, it exhibits excellent performance in terms of both the reference and no-reference image indicators.

3.3.1 Ablation experiment: influence of modules on model performance

To validate the influence of the modules on the performance of the proposed model, we design the modules, as given in Table 4, to evaluate their influence on the assessment indicators of the model.

As shown in Table 4, RS-RDB has the largest influence on the performance of the model. Compared with the model with the classic RDB, that with RS-RDB achieves noticeable improvements in all indicators. The preactivation perceived loss calculation has the second-largest influence, which is primarily manifested in the NIQE. Although the WGAN-GP presents the smallest influence, it improves the performance of the model to some extent.

Table 3 The average value of various types of indicators achieved by different algorithms on the images of the mural data set

	BI	SRGAN	ESRGAN	Reference [28]	Reference [29]	Reference [30]	Algorithm proposed in this study
PSNR	25.58	29.24	31.76	31.02	31.99	29.06	**32.53**
SSIM	0.697	0.751	0.774	0.791	0.796	0.764	**0.816**
MS-SSIM	0.706	0.749	0.777	0.782	**0.812**	0.773	0.805
FID	23.4	16.7	13.5	12.6	14.1	16.9	**11.5**
IS	14.3	17.5	**21.2**	20.1	19.4	18.3	20.8
NIQE	11.202	7.416	4.154	5.125	5.261	5.311	**3.969**

Table 4 The average value of various types of indicators achieved by different algorithms on the images of the mural data set

	PSNR	SSIM	FID	IS	MS-SSIM	NIQE
RDB + Calculate the perceptual loss after activation + No WGAN-GP	28.59	0.751	17.1	17.9	0.768	7.521
RDB + Calculate the perceptual loss before activation + No WGAN-GP	29.93	0.762	15.1	18.5	0.770	4.621
RDB + Calculate the perceptual loss before activation + WGAN-GP	29.99	0.769	14.8	18.7	0.771	4.623
RS-RDB + Calculate the perceptual loss after activation + No WGAN-GP	30.45	0.784	12.34	19.6	0.774	6.497
RS-RDB + Calculate the perceptual loss before activation + No WGAN-GP	32.41	0.801	11.6	20.5	0.801	4.016
RS-RDB + Calculate the perceptual loss before activation + WGAN-GP	32.53	0.816	11.5	20.8	0.805	3.969

3.4 Ablation experiment: influence of superparameters on model performance

To validate the influence of superparameters on the performance of the proposed model, we design the residual scaling factor (Table 5) and the learning rate (Table 6) to evaluate their influence on the model performance.

As shown in Table 5, when β = 0.2 and β = 0.3, the model exhibits excellent performance, and when β = 0.2, the model is in the optimal state. Therefore, β = 0.2 is most suitable for the SESRGAN model. Table 6 shows that when the learning rate is 0.5, the model exhibits satisfactory performance.

The algorithm proposed in this study achieved a good superresolution reconstruction effect for the ancient mural data set. However, during the process of mural reconstruction, mural images whose reconstruction effects were not ideal also existed, as shown in Fig. 14. The image on the left is the generated image, while the image on the right is the HR image. An observation of the facial features and eave texture in the generated image reveals differences from those in the HR image. This is mainly because there was an insufficient number of murals of this style; thus, the model did not learn sufficient feature information about murals of this style and therefore lacked high-frequency detail information. As a result, the model failed to restore the complete murals when reconstructing them.

4 Conclusions

We proposed the SESRGAN algorithm in this study to reconstruct ancient murals and solve the problems of blurring and low resolution, which are responsible for the low appreciation and research value of ancient murals. In our GAN, the RDB module was

Table 5 Influence of the residual scaling factor on the model evaluation indicators

	PSNR	SSIM	FID	IS	MS-SSIM	NIQE
β = 0.1	30.59	0.731	19.2	16.3	0.781	4.941
β = 0.2	**32.53**	**0.816**	**11.5**	20.8	**0.805**	**3.969**
β = 0.3	31.11	**0.816**	14.3	**21.6**	0.804	4.623
β = 0.5	29.45	0.784	12.34	19.6	0.774	6.497
β = 0.7	30.16	0.784	15.5	18.9	0.780	5.651
β = 0.9	30.05	0.783	16.4	19.1	0.779	5.481

Table 6 Impact of the learning rate on the model evaluation indicators

	PSNR	SSIM	FID	IS	MS-SSIM	NIQE
LR = 0.1	29.54	0.721	18.3	15.9	0.781	5.642
LR = 0.5	**32.53**	**0.816**	**11.5**	**20.8**	**0.805**	**3.969**
LR = 0.01	32.11	0.806	15.3	19.4	0.791	4.625
LR = 0.05	31.45	0.781	15.5	18.1	0.780	6.551
LR = 0.001	30.13	0.784	15.5	18.9	0.780	5.651
LR = 0.005	28.16	0.761	20.1	17.5	0.764	7.162

LR learning rate

used for extracting deep features, which was beneficial for extracting feature information. Then, the WGAN-GP theory was introduced to improve the adversarial loss. The Wasserstein distance was adopted to calculate the distance between the reconstructed image and the real image. The gradient penalty was applied to the discriminator network to improve the stability of the network training process. The preactivation features were used to calculate the perceptual loss, which was conducive to restoring texture details and a satisfactory level of brightness. Second transfer learning was adopted to train the model, and finally, a stable enhanced mural reconstruction model was obtained. Compared with other superresolution algorithms, the algorithm proposed in this study achieved improvements in terms of different types of objective assessment indicators. In terms of the subjective evaluation index, it outperformed other algorithms in terms of the MOS and obtained a higher score in terms of local detailed texture than that obtained by other algorithms. This model achieved a good reconstruction effect for both the overall mural images and the local mural images. After reconstruction, the HR mural images were clear and bright. The reconstructed images retained rich texture details, and the ornamental and research values of the murals were enhanced to a certain extent. This model can help prevent the loss of ancient mural images.

The deficiency of this study lies in the superresolution reconstruction process being carried out only after the murals were enlarged 4 times. However, the reconstruction

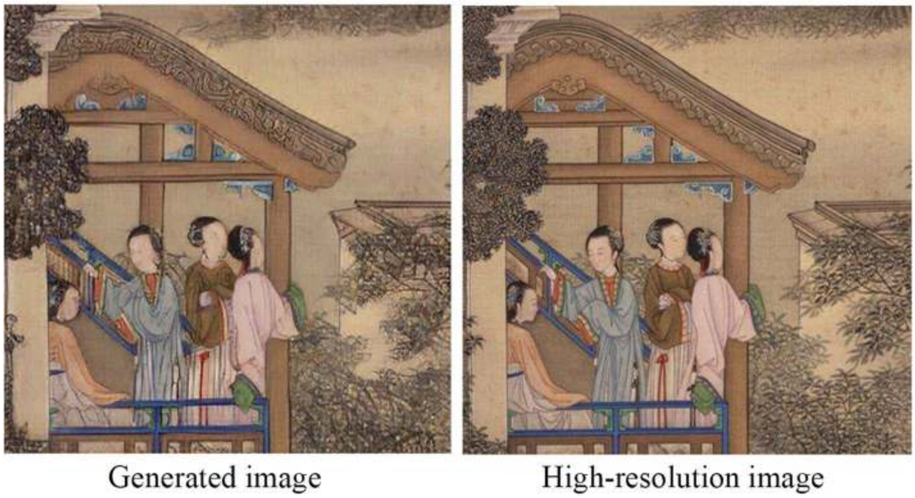

Fig. 14 Reconstructed mural images with unsatisfactory results

process was not carried out at multiple scales. Next, the global feature information and local feature information were not fully utilized. In addition, the texture details of some reconstructed murals were not recovered completely, and the training period of the model was long. The main tasks in the next stage of our research include (1) collecting more mural data sets of different styles to adapt to the depth of the generative network and increase the generalizability and stability of the model; (2) conducting multiscale superresolution reconstruction of murals, showing the artistry and research value of murals at different scales; and (3) simplifying the network structure to reduce the training time and integrating the global and local features into the training process to obtain more detailed texture features and higher-quality mural images than before.

Abbreviations

SESRGAN: Superresolution generative adversarial network; HR: High-resolution; PSNR: Peak signal-to-noise ratio; BI: Bicubic interpolation; POCS: Projection onto convex sets; CNN: Convolutional neural network; SRGA
N: Superresolution generative adversarial networks

Acknowledgments

None.

Authors’ contributions

All authors have made contributions to the current work. JC devised the study plan, led the writing of the article, and was responsible for project management; YJ and MY conducted the experiment and collected the data; and XT conducted the analysis.

Funding

This work was supported by the Natural Science Foundation of Shanxi Province [grant number 201701D21059], a Project of the Key Basic Research in Humanities and Social Sciences of Shanxi Colleges and Universities (20190130), an Art and Science Planning Project of Shanxi Province [grant number 2017F06], and the Platform and Personnel Specialty of Xinzhou (20180601).

The funders had no roles in the design of the study, the collection, analysis, and interpretation of data, and the writing of the manuscript.

Availability of data and materials

The data sets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 9 October 2020 Accepted: 18 July 2021
Published online: 30 July 2021

References

1. JC Guo, J Wu, C Guo, MH Z, Image super-resolution reconstruction based on residual connection convolutional neural network. J. Jilin Univ. 49(5), 1726-1734 (2019).
2. Z.F. Lu, B.J. Zhong, Image interpolation with predicted gradients. Acta Automat. Sin. **44**(6), 1072–1085 (2018)
3. X. Zhang, X. Wu, Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation. IEEE T. Image Process. **17**(16), 887–896 (2008)
4. J. Sun, Q. Yuan, J.W. Li, C.P. Zhou, H.F. Shen, License plate image super-resolution based on intensity-gradient prior combination. J. Image Graphic. **23**(06), 802–813 (2018)
5. R. Timofte, S.V. De, G.L. Van, *Adjusted anchored neighborhood regression for fast super-resolution. Paper presented at the Asian conference on computer vision* (Springer, Cham, 2014), pp. 111–126
6. Y. Anagun, S. Isik, E. Seke, SRLibrary: Comparing different loss functions for super-resolution over various convolutional architectures. J. Vis. Commun. Image R. **61**, 178–187 (2019)
7. D. Zhou, in *Paper presented at the International Congress on Image and Signal Processing*. An Edge-Directed bicubic Interpolation Algorithm (IEEE, 2010), pp. 1186–1189
8. S. Farsiu, D. Robinson, S. Elad, Advanced and challenges in super-resolution. Int. J. Imaging Syst. Technol. **14**(2), 47–57 (2004)
9. M. Irani, S. Peleg, Improving resolution by image registration. CVGIP: Graph. Models Image Process. **53**(3), 231–239 (1991)
10. H. Stark, P. Oskoui, High resolution image recovery from image-plane arrays, using convex projections. J. Opt. Soc. Am. A **6**(11), 1715–1726 (1989)
11. H. Chang, D.Y. Yeung, Y. Xiong, *Super-resolution through neighbor embedding*, vol 1 (Paper presented at the IEEE Conference on Computer Vision and Pattem Recognition (CVPR), Washington, DC, USA, 2004), pp. 275–282

12. J. Yang, J. Wrigbt, T.S. Huang, Y. Ma, Image super-resolution via sparse representation. IEEE T. Image Process. **19**(11), 2861–2873 (2010)
13. C. Dong, C. Loy, K. He, X. Tang, in *Paper presented at the Proceedings of European Conference on Computer Vision*. Learning a deep convolutional network for image super-resolution (Springer, Cham, 2014), pp. 184–199
14. C. Dong, C.L. Chen, X. Tang, *Accelerating the superresolution convolutional neural network. Paper presented at the Proceedings of the 2016 14th European Conference on Computer Vision* (Berlin, Springer, 2016), pp. 391–407
15. T. Tong, G. Li, X. Liu, Q. Gao, *Image super-resolution using dense skip connections. Paper presented at the 2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE Computer Society, 2017), pp. 4809–4817
16. Y.N. Zhang, M.Q. An, Deep learning and transfer learning-based super resolution reconstruction from single medical image. J. Healthcare Eng, 1–20 (2017). https://doi.org/10.1155/2017/5859727
17. J. Kim, L.J. Kwon, L.K. Mu, *Deeply-recursive convolutional network for image super-resolution. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1637–1645
18. B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, *Enhanced deep residual networks for single image super-resolution. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), pp. 136–144
19. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, *Photo-realistic single image super-resolution using a generative adversarial network. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4681–4690
20. M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2017)
21. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. Loy, *Esrgan: Enhanced super-resolution generative adversarial networks. Paper presented at the Proceedings of the European Conference on Computer Vision (ECCV)* (2019), pp. 63–79
22. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Generative adversarial nets. Paper presented at the Proceedings of the 27th International Conference on Neural Information Processing Systems* (MITPress, Cambridge, 2014), pp. 2672–2680
23. K.M. He, X.Y. Zhang, S.Q. Ren, J. Sun, *Deep residual learning for image recognition. Paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, 2016), pp. 770–778
24. W.S. Lai, J.B. Huang, N. Ahuja, M.H. Yang, *Deep laplacian pyramid networks for fast and accurate super-resolution. Paper presented at the Proc of IEEE International Conference on Computer Vision* (IEEE Computer Society, 2017), pp. 5835–5843
25. G. Huang, Z. Liu, L. Weinberger, *Densely connected convolutional networks. Paper presented at the CVPR 2017: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Piscataway, 2017), pp. 2261–2269
26. Y. Gong, J. Hou, Y.J. Ying, Night-time aerial vehicle recognition method based on transfer learning and image enhancement. J Comp. Aided Design Graph. **31**(3), 467–473 (2019)
27. B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, *Enhanced deep residual networks for single image super-resolution. Paper presented at the Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (IEEE Computer Society, Washington, DC, 2017), pp. 136–144
28. K. Jiang, Z.Y. Wang, P. Yi, J.J. Jiang, Hierarchical dense recursive network for image super-resolution. Pattern Recogn. **107**, 107475 (2020)
29. Dharejo F A, Deeba F, Zhou Y, Das B, Jatoi MA, Zawish M, Du Y, Wang XZ. TWIST-GAN: Towards wavelet transform and transferred GAN for spatio-temporal single image super resolution. arXiv preprint arXiv:2104.10268 (2021).
30. M. Zhang, Q. Ling, Supervised pixel-wise GAN for face super-resolution. IEEE Transac. Multimedia (2020) https://ieeexplore.ieee.org/document/9132630

Publisher's Note

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

Submit your manuscript to a SpringerOpen 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