J. Meas. Sci. Instrum., 2024, 15(4): 499-508.
DOI: 10.62756/jmsi.1674-8042.2024050
http://jmsi.publish.founderss.cn/
jmsi@nuc.edu.cn

# A super-resolution reconstruction algorithm for mural images based on improved generative adversarial network

GAO Li*, ZHOU Xiaohui

*School of Electrical and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China*

***Corresponding author:** GAO Li (13919801756@163.com)

**Received:** January 14, 2023 **Revised:** March 12, 2023 **Accepted:** May 1, 2023

**Abstract:** In order to solve the problem of the lack of ornamental value and research value of ancient mural paintings due to low resolution and fuzzy texture details, a super resolution (SR) method based on generative adduction network (GAN) was proposed. This method reconstructed the detail texture of mural image better. Firstly, in view of the insufficient utilization of shallow image features, information distillation blocks (IDB) were introduced to extract shallow image features and enhance the output results of the network behind. Secondly, residual dense blocks with residual scaling and feature fusion (RRDB-Fs) were used to extract deep image features, which removed the BN layer in the residual block that affected the quality of image generation, and improved the training speed of the network. Furthermore, local feature fusion and global feature fusion were applied in the generation network, and the features of different levels were merged together adaptively, so that the reconstructed image contained rich details. Finally, in calculating the perceptual loss, the brightness consistency between the reconstructed fresco and the original fresco was enhanced by using the features before activation, while avoiding artificial interference. The experimental results showed that the peak signal-to-noise ratio and structural similarity metrics were improved compared with other algorithms, with an improvement of 0.512 dB—3.016 dB in peak signal-to-noise ratio and 0.009—0.089 in structural similarity, and the proposed method had better visual effects.

**Key words:** mural image; super-resolution reconstruction; generative adversarial network; information distillation block (IDB); feature fusion

## 0 Introduction

Mural painting is a historical picture that reflects thought and real life. It is a precious art property with high academic value and aesthetic value. However, due to the destruction of human factors and natural environment, the ancient frescoes have been damaged to different degrees, so it is particularly necessary to protect and restore the frescoes. The early restoration of mural paintings was done by art lovers through the method of copying, which has great limitations for accurate and efficient restoration of mural paintings. With the development of science and technology, it is increasingly common to apply digital image processing technology to the restoration and enhancement of mural images. One of the important methods is the super-resolution reconstruction of mural images, which improves the viewing experience of tourists by enhancing and repairing mural images. Single image super-resolution (SISR) aims to generate a visually pleasing high-resolution (HR) image from its degraded low-resolution (LR) measurement[1]. Currently, there are three main approaches to the study of high-resolution images: reconstruction-based[2], interpolation-based[3], and learning-based[4,5]. Since the recovery performance of the first two methods tends to decline sharply when the upper scale factor is large, the current super resolution(SR) method belongs to the learning-based method, trying to learn prior knowledge from the LR and HR pairs.

In the deep learn-based approach, the network can automatically extract image features, learn the mapping relationship between high-resolution images and low-resolution images through a large number of datasets, and finally generate high resolution images with rich detail texture and clear visual perception. Dong et al.[6] firstly introduced convolutional neural networks (CNN) into the field of image super-resolution to establish an end-to-end mapping between LR images and HR images, resulting in a significant improvement in the effectiveness and speed of

© The Author(s) 2024. Published by Publishing Center of North University of China. The articles published in this open access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

image super-resolution reconstruction techniques. Kim et al. increased the depth of the network in VDSR[7], which adopted residual learning and adaptive gradient clipping to ease training difficulty. To increase the perceptual field of the network, recurrent neural networks (RNN) was used for the first time to achieve super-resolution of images[8]. Lai et al. proposed a Laplacian pyramid super-resolution network (LapSRN)[9] model using a pyramid structure for image super-resolution reconstruction, which could achieve super-resolution reconstruction of images at different magnifications and achieve better visual effects. Hu et al.[10] proposed an image super-resolution network based on dense connectivity and excitation modules, where the dense connectivity of features improved the visual effect of reconstructed images. Zhang et al.[11] proposed an image super-resolution reconstruction method based on densely connected residual network, which improved the efficiency of feature extraction through the fusion of dense residual blocks. Because generative adversarial network (GAN) is able to learn more information about images, they have been widely used in recent years for super-resolution tasks. Ledig et al. proposed a super-resolution reconstruction algorithm based on generative adversarial networks (SRGAN)[12,13], which solved the problem of lacking high-frequency information in the reconstructed images. Hu et al. proposed a real-time super-resolution generative adversarial network (RTSRGAN)[14], capable of performing image super-resolution reconstruction tasks in real time. Chen et al.[15] introduced the earth-mover distance to redesign the loss function on the basis of SRGAN to improve the stability of network training. Li et al.[16] proposed a deep generative adversarial network for super-resolution image restoration and reconstruction, constructing a relative discriminator to achieve multi-directional data gradient reception of the dataset. Aiming at the problem that the image super-resolution reconstruction method based on generative adversarial network relies on paired training and the results are unstable, Li et al.[17] proposed an NM-SRGAN model based on unpaired images, which had better stability and image quality. Duan et al.[18] proposed an image super-resolution reconstruction algorithm based on Laplacian pyramidal generative adversarial network, which solved problems of poor reconstruction of large scale factors and the need for separate training for all image reconstructions at different scales in the super-resolution reconstruction algorithm.

Aiming at the problems of artifacts and blurred texture details in the reconstructed images by existing algorithms, a super-resolution reconstruction algorithm for mural images was proposed based on generative adversarial networks. Information distillation blocks (IDB)[19] were introduced into the generative network to enhance shallow image features. The residual dense blocks with residual scaling and feature fusion (RRDB-Fs) module were used to fully extract the color and content information of the murals, and the batch normalization (BN) layer that was not suitable for reconstruction work was removed, and information extraction and feature utilization were enhanced through local feature fusion and local residual learning. The global residual learning and cross-layer feature fusion enabled the information of different frequencies to be effectively transmitted to the deep layers to obtain global mural features. Finally, the VGG perceptual loss before feature activation was used to provide stronger supervision for the brightness consistency and local texture restoration of murals.

## 1 Network structure

The SRGAN algorithm is mainly composed of a generative network and a discriminative network. Firstly, the low-resolution image is fed into the generative network, which generates the image by the improved feature extraction module and reconstruction module. Then the reconstructed image is input into the discriminant network, which judges the authenticity of the input image. The loss function is used to train and optimize the model, so that the resolution of the reconstructed image is constantly approaching the resolution of the high-resolution image. Its basic structure is shown in Fig.1.

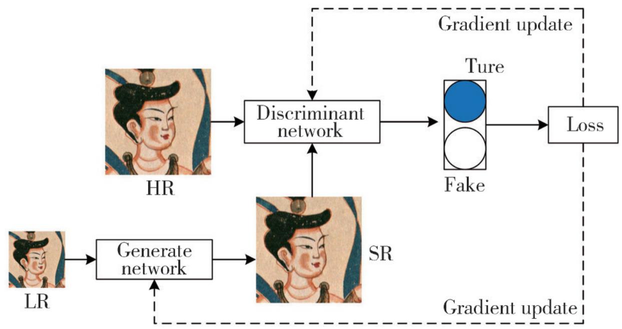


**Fig. 1 SRGAN network**

### 1.1 Generative network

As shown in Fig. 2, the generative network mainly consists of four parts: shallow image feature extraction, RRDB-Fs, global feature fusion, and finally sub-pixel convolution. The high-resolution mural images are pre-processed by down-sampling to obtain the desired low-resolution mural images ($I_{LR}$) before the network is trained.

The $I_{LR}$ is fed into the network and the two convolution and activation modules are used to extract the shallow features of the mural image. The obtained shallow features are fed into the IDB module to enhance the image features. There are four IDBs in the network structure and the input of the current IDB is the output of the previous IDB. Then, the enhanced mural image features are sent to RRDB-Fs, and the hierarchical features are extracted through a set of RRDB-Fs, and then global feature fusion is further performed. After extracting local and global features in low-resolution mural images, the mural image features are fed into sub-pixel convolution to upscale the image. Another method used for image enlargement is deconvolution, which usually involves more human factors. The sub-pixel convolution is a learning process. Compared with deconvolution, the results of sub-pixel convolution are more accurate by continuously learning the parameters and zooming in on the image. Each sub-pixel convolution can magnify the image features by two times, because the magnification factor of the mural image super-resolution reconstruction in this paper is four times, so two sub-pixel convolutions are used to achieve the experimental effect. Finally, a super-resolution mural image ($I_{SR}$) is generated.

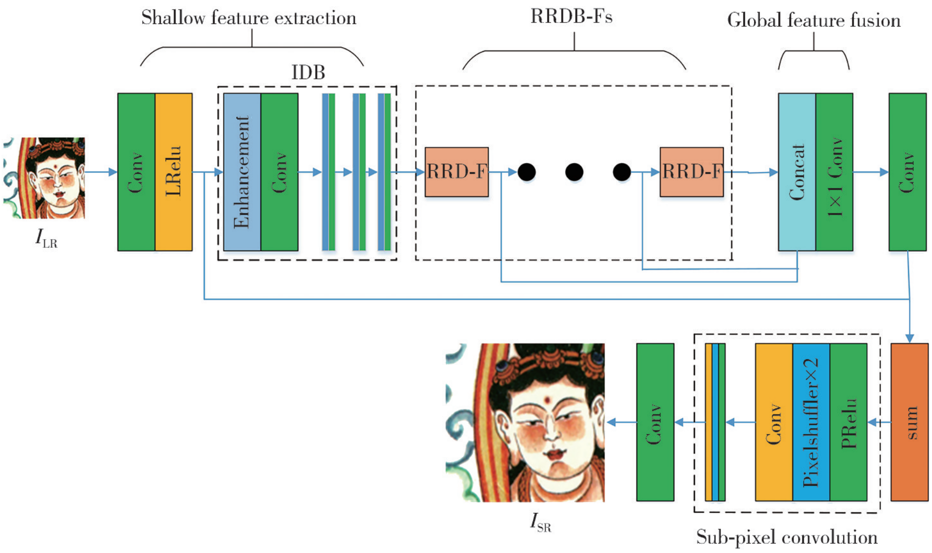


**Fig. 2 Generation network**

1.1.1 Feature enhancement

In the generation network of SRGAN, the shallow features of the image were extracted through a layer of convolution and activation, so the shallow features of the image could not be fully utilized, which affected the output of the network. Therefore, the way of shallow feature extraction was improved, and better results were achieved through the method of information distillation using one convolutional layer and four information distillation blocks to extract and distill shallow mural image features.

The information distillation block consists of an enhancement unit and a compression unit. The enhancement unit is dedicated to enhancing the shallow features of the image, increasing the number of feature channels, so that the extracted image features contain rich texture information. The structure of the enhancement unit is shown in Fig.3, where S represents the information channel segmentation operation and P represents the information channel connection operation. The enhancement unit is composed of two parts, each part contains three convolutional layers, and each convolutional layer is connected with a PReLU activation function. The input $F_{input}$ of the enhancement unit is the shallow image feature extracted after the first layer of convolution. $F_{input}$ outputs short-path information after passing through the three-layer convolutional network of the first part. The short-path information is divided into two parts after channel division, which are the enhanced short-path information ($F_{boost}$) and the reserved short-path information ($F_{remain}$). The three-layer convolution and activation layers of the second part are represented as M. $F_{boost}$ is the input of the second part of the three-layer convolution, the output of which obtains the long-path information. $F_{remain}$ and $F_{input}$ are connected through the information channel to obtain local short-path information. Long-path information and local short-path information connections are the output of the enhancement unit. The output of the enhancement unit can be expressed by

$$F_{output} = M(F_{boost}) + P(F_{input}, F_{remain}). \quad (1)$$

The compression unit is a 1×1 convolutional layer whose role is to compress the redundant information in the output of the enhancement unit and reduce the complexity of the computation.

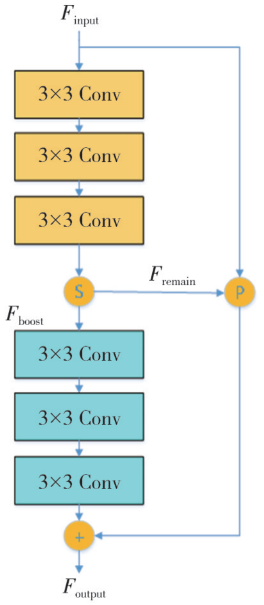


**Fig. 3 Enhancement unit structure**

1.1.2 RRDB

It has been shown that more layers and connections can improve the visual quality of reconstructed images. In order to extract content features and color information at different depths of the mural images and recover more realistic mural images, this paper used dense blocks with residual scaling and feature fusion (RRDB-F) instead of ResNet[20] in the generative network of SRGAN, which increased the network capacity. In order to make full use of the hierarchical features of low-resolution images, this paper added local feature fusion to RRDB-F. The RRDB-F network structure is shown in Fig. 4, which includes densely connected layers, local feature fusion, residual scaling, and local residual learning. The RRDB-F consists of three dense blocks (DB), each containing four convolutional layers, four activation functions, and feature fusion. The output of the previous dense block and the features of each layer in the current dense block are fused together through concat, and then the 1×1 convolution is used to reduce the number of channels and simplify the data. Since there are multiple convolution layers in a database, local residual learning can further improve the utilization of feature information. Residual scaling is introduced in RRDB-F to improve the stability of training deep networks by multiplying dense blocks of residuals by a value between 0 and 1. This network structure allows the learned image information to be retained and utilized multiple times, making the reconstructed image texture more continuous.

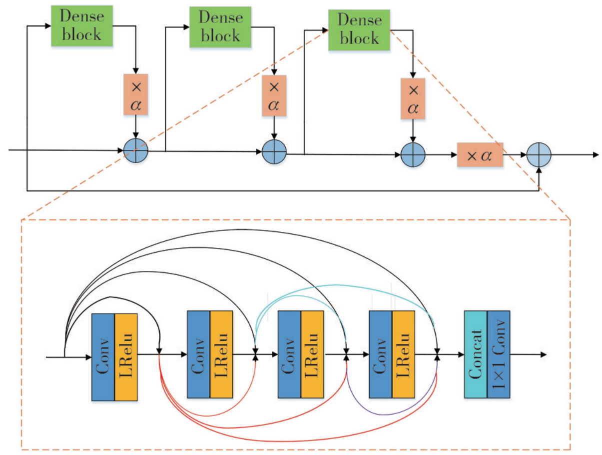


**Fig. 4 RRDB-F network**

In order to enhance the consistency of the network model and the stability of the network during training, the BN layer was removed from the DB module to enhance the reconstructed images and reduce the computational complexity and memory usage, as shown in Fig. 5. For evaluating network performance with peak signal-to-noise ratio, such as super-resolution image reconstruction and deblurring, adding BN layer will affect the stability of network training and the quality of generated images, which limits the generalization ability of the model. Because the color distribution of the image will be normalized after passing through the BN layer, which destroys the original contrast information of the image. But in super-resolution image reconstruction, these contrast information is needed for network training. Therefore, the BN layer is not suitable for image super-resolution.

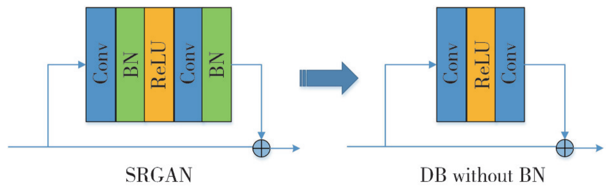


**Fig. 5 Network with BN layer removed**

1.1.3 Global feature fusion and residual learning

In the overall structure of the generative network, the hierarchical features of all RRDB-Fs in the low-resolution mural images are adaptively fused from a global perspective. More characteristic information can be obtained by summing the output of a series of RRDBS through concat of channel dimensions. Then, the optimal information content is filtered out through a 1×1 conversion layer. When low-resolution image features are extracted in the deeper layers of the network, the feature information is reduced after filtering by the convolution and activation function layers. But through global feature fusion, the generative network can use the feature information learned by multiple RRDB-Fs for image reconstruction. The output of the first convolutional layer is element-wise summed with the output of the overall feature fusion by global residual learning. The shallow features and deep features of the mural images are combined to obtain the global dense features of the low-resolution mural images.

Global feature fusion and residual learning can combine image features extracted from different network depths to reconstruct high quality mural images.

### 1.2 Discriminant network

The main purpose of the discriminant network is to discriminate the authenticity of the input image, which is equivalent to a binary classification problem. The basic architecture of the discriminant network is the VGG19 network, including feature extraction and linear classification. The network structure is shown in Fig.6.

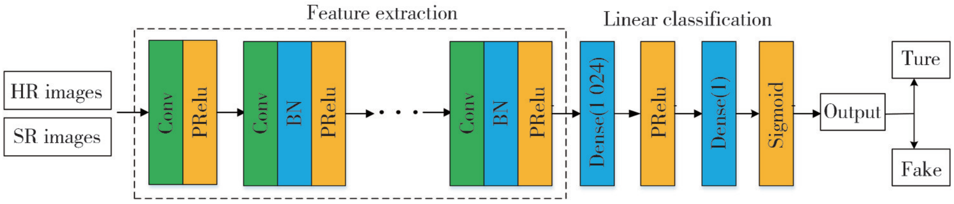


**Fig. 6 Discriminant network**

The feature extraction module has 8 convolutional layers. The activation function after the convolutional layer uses PReLU. In order to enhance network stability and avoid gradient disappearance, a BN layer is set after the convolutional layer. The discriminant network classifies the input real HR mural image and reconstructed mural image, and continuously trains and optimizes the reconstructed image to get closer and closer to the real high-resolution image.

## 2 Loss function and evaluation index

### 2.1 Loss function

The network model is optimized and constrained using a content loss function, a perceptual loss function, and an adversarial loss function. The loss function is defined as

$$L^{\mathrm{SR}} = L^{\mathrm{SR}}_{\mathrm{MSE}} + \lambda_1 L^{\mathrm{SR}}_{\mathrm{VGG}} + \lambda_2 L^{\mathrm{SR}}_{\mathrm{adv}}, \tag{2}$$

where $L^{\mathrm{SR}}_{\mathrm{MSE}}$ is the content loss, which is defined as

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

where $I^{\mathrm{LR}}$ denotes LR image; $W$ and $H$ are the width and height of the LR image, respectively; $r$ is the down-sampling factor from $I^{\mathrm{HR}}$ to $I^{\mathrm{LR}}$; $G_{\blacktriangledown_G}$ is the generative network constructed by parameter $\blacktriangledown_G$.

A good peak signal-to-noise ratio can be obtained using content loss, but the reconstructed mural image is too smooth in the parts of the image where the texture detail changes dramatically, losing the high frequency information of the mural image. SRGAN adds a perceptual loss ($L^{\mathrm{SR}}_{\mathrm{VGG}}$) based on the VGG network to reconstruct images more in line with human vision. $L^{\mathrm{SR}}_{\mathrm{VGG}}$ is defined as the distance between two activation features minimized in the activation layer of the pre-trained network, that is

$$L^{\mathrm{SR}}_{\mathrm{VGG}} = \frac{1}{W_{ij}H_{ij}}\sum_{x=1}^{W_{ij}}\sum_{y=1}^{H_{ij}}\left(\varphi^{ij}\left(I^{\mathrm{HR}}\right)_{x,y} - \varphi^{ij}\left(G\left(I^{\mathrm{LR}}\right)\right)_{x,y}\right)^2, \tag{4}$$

where $W_{ij}$ and $H_{ij}$ are the dimensions of the feature map in the VGG network; $\varphi^{ij}$ represents the feature map obtained before the $i$th pooling after the $j$th convolution (before activation) in the VGG network.

The feature information was used before activation instead of after activation for calculation. Because the activated mural features are very sparse in deeper networks. The sparse features only provide weak supervision and the reconstructed mural images are not ideal. And the perceptual loss after using activation makes the reconstructed mural images inconsistent in brightness.

The adversarial loss can be expressed as

$$L^{\mathrm{SR}}_{\mathrm{adv}} = \sum_{n=1}^{N} -\log D_{\theta D}\left(G_{\theta D}\left(I^{\mathrm{LR}}\right)\right), \tag{5}$$

where $D_{\blacktriangledown D}\left(G_{\blacktriangledown D}\left(I^{\mathrm{LR}}\right)\right)$ denotes reconstructed image; $D_{\blacktriangledown D}$ denotes the discriminatory network constructed from parameter $\theta_D$; $N$ is the number of samples for one batch training.

### 2.2 Evaluation index

The validity of the experiment was proved by two objective evaluation indicators: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). PSNR is used to indicate the degree of information loss between the reconstructed image and the high-resolution image. The higher PSNR value means less distortion and the higher quality of the reconstructed image. Its unit is dB, and the calculation can be expressed as

$$PSNR = 10\lg\left(\frac{225^2 WH}{\sum_{i=1}^{W}\sum_{j=1}^{H}\left[\bar{x}(i,j) - x(i,j)\right]^2}\right), \tag{6}$$

where $W$, $H$ are the width and height of the image, respectively; $\bar{x}(i,j)-x(i,j)$ denotes the Euclidean distance between the pixels of the two images.

SSIM determines the similarity between the reconstructed image and the original image by three properties: brightness, contrast, and structure of the image. SSIM is more consistent with human vision, and the closer its value is to 1, the more similar the two images are. The SSIM is defined as

$$SSIM(X,Y)=\frac{(2\mu_X\mu_Y+C_1)(2\sigma_{XY}+C_2)}{(\mu_X{}^2+\mu_Y{}^2+C_1)(\sigma_X{}^2+\sigma_Y{}^2+C_2)}, \tag{7}$$

where $X$ represents the reconstructed image and $Y$ represents the original high-resolution image; $\mu_X$ and $\mu_Y$ are means; $\sigma_X$ and $\sigma_Y$ are variances; $\sigma_{XY}$ is the covariance, and $C_1$ and $C_2$ are constants.

# 3 Experiment

## 3.1 Experimental environment

About 150 frescoes each from the Buddha and Flying Sky of Dunhuang frescoes were collected, of which 120 fresco images from each category were used as the training dataset and the rest as the test dataset. To enrich the data sample, the training images of the murals were scaled and rotated. The training dataset also had the public dataset DIV2K. The HR images were uniformly cropped to 96 pixel×96 pixel in front of the input network, depending on the computer performance. The corresponding low-resolution images were then obtained by four times down-sampling. The training batch size was set to 32 and num-work was set to 4. The experimental environment is shown in Table 1.

**Table 1 Experimental environment**

| Project name | Specific configuration |
|---|---|
| Operating system | Windows 10 |
| GPU | GTX 1080Ti |
| Graphics memory | 12 G |
| CUDA version | 10.0 |
| Deep learning framework | Pytorch 3.6 |
| Compiled software | Pcharm 2020 Matlab 2013 |

The operating system was Windows 10. Matlab was used for data processing in the early stage, and Pycharm was used for building and training network model in the later stage. The deep learning framework was Pytorch.

## 3.2 Results analysis

Both the generative network and the overall network were analysed to show the advantages of the propsed algorithm.

### 3.2.1 Generate network analysis

The generative network was first pre-trained, and then the whole network was trained. The variation of the loss function of the generative network in this paper was observed through tensorboard. The horizontal coordinate is the number of iterations (epoch) and the vertical coordinate is the loss function MSE (Mean squared error), as shown in Fig. 7. It is observed that as the number of iterations increases, the value of the loss becomes smaller and smaller, and when the number of iterations is 300, the value tends to be stable.

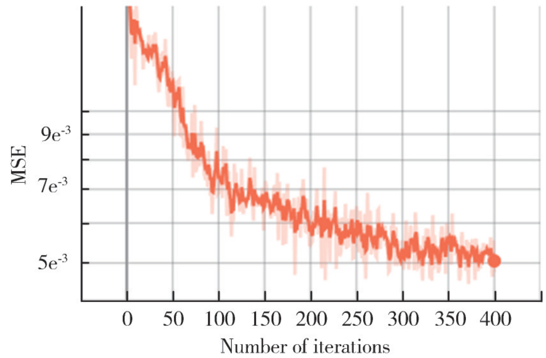


**Fig. 7 MSE loss of generated network in this paper**

An experimental comparison was made between the generative network of this paper and the original generative network (SRResNet). 6 test images of the murals were randomly selected and the model effects were assessed by calculating the PSNR and SSIM of the two model reconstructions of the test images. The results are shown in Table 2.

**Table 2 Comparison of generating network reconstruction indicators**

| Murals image | Model | PSNR/dB | SSIM |
|---|---|---|---|
| 1 | SRResNet | 27.302 | 0.761 |
| | Ours | 29.022 | 0.805 |
| 2 | SRResNet | 26.985 | 0.867 |
| | Ours | 27.973 | 0.869 |
| 3 | SRResNet | 30.690 | 0.857 |
| | Ours | 31.180 | 0.863 |
| 4 | SRResNet | 28.794 | 0.842 |
| | OursOurs | v29.489 | 0.847 |
| 5 | SRResNet | 30.517 | 0.82 |
| | Ours | 31.466 | 0.834 |
| 6 | SRResNet | 29.167 | 0.825 |
| | Ours | 29.704 | 0.823 |
| Average | SRResNet | 28.909 | 0.829 |
| | Ours | 29.805 | 0.840 |

It can be seen that the generation network in this paper outperforms the SRResNet algorithm in both PSNR and SSIM metrics. Among them, the average increase in PSNR is 0.896 dB and the average increase in SSIM is 0.011. It was shown that the proposed method, which made full use of the shallow features of the mural images and adaptively blends features from different levels together, was suitable for the image reconstruction task.

3.2.2 Overall network analysis

During the network training process, the visual effect of the reconstructed images changes as the number of iterations increases. The experimental comparison is shown in Fig.8.

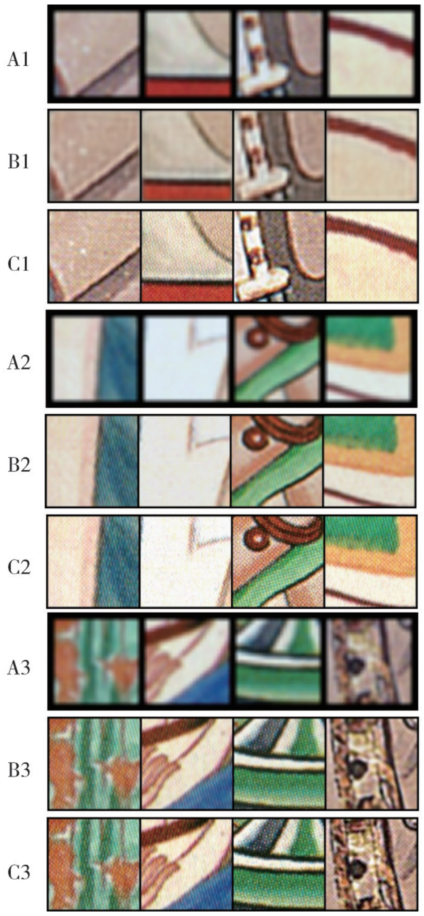


**Fig. 8 Reconstruction effect of proposed algorithm. (A1–A3) Low-resolution image; (B1–B3) Reconstructed image; (C1–C3) Original image**

Fig.8 shows a comparison of the reconstructed images during the training of the proposed algorithm with four different images in each group. A1—A3 are the low-resolution mural images. C1—C3 are the original mural images. B1 is the reconstructed image when the number of iterations is 1, B2 is the reconstructed image when the number of iterations is half of the total number of iterations, and B3 is the final reconstructed image. It can be found that at first the reconstructed images are blurred because the network model learns few image features. As the number of iterations increases, the reconstructed mural images become more and more detailed. The final reconstructed images of the mural have clearer details, and the color brightness are closer than the original image.

A classical traditional method of Bicubic and 3 deep learning methods, namely SRCNN, VDSR, and SRGAN were selected for experimental comparison with the proposed algorithm on 5 local mural images. The 5 murals are all Dunhuang murals. The experimental results are shown in Fig. 9, which respectively shows the reconstruction effect of each algorithm. There are 5 sets of images, the top part of each set of images is the reconstructed image of each algorithm, and the bottom part is the local zoomed image. Fig. 9(a) is the image of Guanyin in Cave 57; Fig.9(b) is the Yakushi Buddha in the Oriental Medicine Buddha Sutra variation on the south wall of Cave 220; Fig.9(c) is the image of the bodhisattva who hit the pipa in cave 112; Fig.9(d) is the image of the Flying Buddha in Cave 321; and Fig. 9(e) is the image of the Flying Buddha in Cave 332.

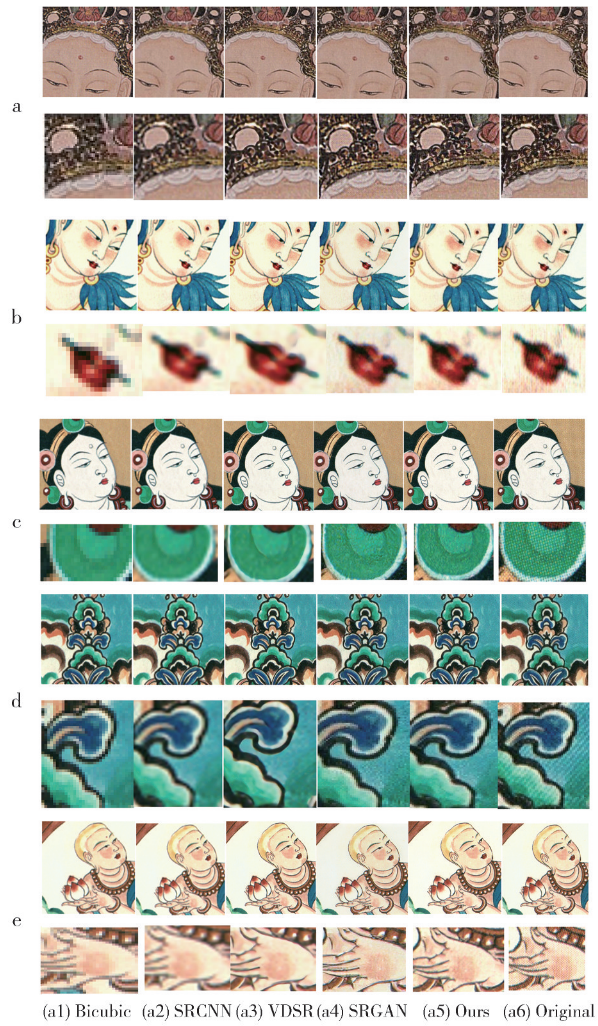


**Fig. 9 Mural images reconstructed by each algorithm**

From left to right in the Fig.9 are the low-resolution images Bicubic, SRCNN, VDSR, SRGAN, the algorithm in this paper, and the original image. Enlarged images show that Bicubic has performed only a simple interpolation algorithm, and the reconstructed mural images are relatively smooth, fuzzy, and of the lowest quality. SRCNN is a three-layer convolutional network, and the reconstructed images are not as smooth and visually better than Bicubic. The VDSR algorithm deepens the number of network layers and extracts more image features. And the network structure is more stable. The reconstructed image has better visual effect than the previous two algorithms. But the edge sharpening is more obvious. SRGAN and the proposed algorithm are models based on perceptual loss, and the overall visual effect far exceeds the previous 3 kinds of algorithms.

However, there is a difference in the brightness between the reconstructed image of SRGAN and the original image. It can be seen in the partially enlarged images of group b and group e that the color and brightness of images reconstructed by the proposed algorithm are closer to the original image. In addition, the proposed algorithm makes full use of images detail features, and the reconstructed images are more detailed than the previous algorithms. It is observed in the group c and group d that the images reconstructed by SRGAN have artifacts. However, the reconstructed images of the proposed algorithm have no artifacts and are closer to the original images.

Then the PSNR and SSIM indexes of the reconstructed images are calculated in the above 5 test images, and the advantages and disadvantages of the algorithms are accurately compared by using objective evaluation indexes. The results are shown in Tables 3 and 4.

**Table 3 Comparison of PSNR values among super-resolution reconstruction algorithms**

| Mural image | PSNR/dB | | | | |
|---|---|---|---|---|---|
| | Bicubic | SRCNN | VDSR | SRGAN | Ours |
| 1 | 25.107 | 26.659 | 27.815 | 27.357 | 28.362 |
| 2 | 26.964 | 28.742 | 29.306 | 29.928 | 30.278 |
| 3 | 25.129 | 26.911 | 27.976 | 27.366 | 28.056 |
| 4 | 23.612 | 25.169 | 26.297 | 26.943 | 27.115 |
| 5 | 25.817 | 27.015 | 27.262 | 28.028 | 28.621 |
| Avg | 25.326 | 26.899 | 27.731 | 27.924 | 28.486 |

**Table 4 Comparison of SSIM values among super-resolution reconstruction algorithms**

| Mural image | SSIM | | | | |
|---|---|---|---|---|---|
| | Bicubic | SRCNN | VDSR | SRGAN | Ours |
| 1 | 0.679 | 0.722 | 0.758 | 0.752 | 0.764 |
| 2 | 0.711 | 0.776 | 0.801 | 0.814 | 0.821 |
| 3 | 0..681 | 0.732 | 0.751 | 0.749 | 0.768 |
| 4 | 0.736 | 0.788 | 0.801 | 0.823 | 0.817 |
| 5 | 0.675 | 0.707 | 0.712 | 0.741 | 0.753 |
| Avg | 0.696 | 0.745 | 0.765 | 0.776 | 0.785 |

The proposed algorithm achieves higher PSNR and SSIM metric values for the 5 test images than the other algorithms at the four-fold magnification reconstruction scale factor. Compared with the interpolation method, the improvement is particularly obvious, because deep learning has more parameters that can be learned by itself, which can better reconstruct the learning model and achieve better expressive ability. The effect values are also significantly better compared to SRGAN. The proposed algorithm built a more reasonable generative network, which could effectively increase the depth of the network and generate a better-fitting network, so that the trained network model could find a balance between the improvement of the visual effect of the image and the improvement of the objective evaluation value. Compared with Bicubic, SRCNN, VDSR, and SRGAN, the average PSNR of this algorithm has improved by 3.16 dB, 1.587 dB, 0.755 dB, and 0.562 dB, respectively. The average SSIM of this algorithm has improved by 0.089, 0.04, 0.02 and 0.009, respectively. It can be seen that the overall reconstruction effect of the proposed algorithm is better in the objective quality evaluation.

## 4 Conclusions

Based on SRGAN, an improved generative adversarial network was proposed to achieve super-resolution reconstruction of mural images. Firstly, a generative network was constructed using information distillation blocks and dense residual blocks containing residual scaling and feature fusion to extract pixel feature information from the mural images with high efficiency. Then, the representational power of the network was improved by global and local feature fusion. Finally, the model was jointly optimized using perceptual loss before feature activation, content loss, and adversarial loss to enhance the consistency of the reconstructed mural brightness and to preserve the high-frequency details of the mural images.

After the previous discussion, the PSNR and SSIM values of the test images by the proposed algorithm were higher than those of other methods, and also had better visual effects, which proved that the proposed algorithm could achieve a balanced improvement of image quality and objective evaluation indicators. In future research, the network will be further improved in training network stability to build a better network model to further improve the image reconstruction quality. There is still the problem of a single dataset in this paper. In the future, more mural images of different styles will be collected to increase the generalization ability of the model.

## Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

## References


[1] GUO J, WU J, GUO C, et al. Image super-resolution reconstruction based on residual connection convolutional neural network. Journal of Jilin University (Engineering and Technology Edition), 2019, 49(5): 1726-1734.

[2] SU H, ZHOU J, ZHANG Z H. Survey of super-resolution image reconstruction methods. Acta Automatica Sinica, 2014, 39(8): 1202-1213.

[3] HASAN M S, HAQUE S T. Single image super-resolution using back-propagation neural networks//2017 20th International Conference of Computer and Information Technology (ICCIT), December 22-24, 2017, Dhaka, Bangladesh. New York: IEEE, 2017: 1-5.

[4] PENG Y F, GAO Y, DU T T, et al. Single image super-resolution reconstruction method for generative adversarial network. Journal of Frontiers of Computer Science and Technology, 2020, 14(9): 1612-1620.

[5] WANG Z W, LIU D, YANG J C, et al. Deep networks for image super-resolution with sparse prior//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 370-378.

[6] DONG C, LOY C C, HE K M, et al. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.

[7] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1646-1654.

[8] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1637-1645.

[9] LAI W S, HUANG J B, AHUJA N, et al. Deep Laplacian pyramid networks for fast and accurate super-resolution//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 5835-5843.

[10] HU S Y, WANG G D, ZHAO Y, et al. Image super-resolution network based on dense connection and squeeze module. Laser & Optoelectronics Progress, 2019, 56(20): 201005.

[11] ZHANG Y L, TIAN Y P, KONG Y, et al. Residual dense network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(7): 2480-2495.

[12] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks. Communications of the ACM, 2020, 63(11): 139-144.

[13] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 105-114.

[14] HU X Y, LIU X J, WANG Z C, et al. RTSRGAN: real-time super-resolution generative adversarial networks//2019 Seventh International Conference on Advanced Cloud and Big Data (CBD), September 21-22, 2019, Suzhou, China. New York: IEEE, 2019: 321-326.

[15] CHEN Z H, HU H L, YAO J M, et al. Single frame image super-resolution reconstruction based on improved generative adversarial network. Chinese Journal of Liquid Crystal and Displays, 2021, 36(5): 705-712.

[16] LI Y H, MU X, ZHU Y L, et al. Super resolution image restoration and reconstruction of deep generative countermeasure network. Journal of Xi'an University of Technology, 2021, 35(4): 1-8.

[17] LI X X, CAO Q, LIU C M. Image super-resolution based on no match generative adversarial network. Journal of Zhengzhou University (Engineering Science), 2021, 42(5): 1-6.

[18] DUAN Y X, ZHANG H X, SUN Q F, et al. Image super-resolution reconstruction algorithm based on laplacian pyramid generative adversarial network. Journal of Computer Applications, 2021, 41(4): 1020-1026.

[19] HUI Z, WANG X M, GAO X B. Fast and accurate single image super-resolution via information distillation network//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 723-731.

[20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.

# 基于改进生成对抗网络的壁画图像超分辨率重建算法

高　丽*，周晓慧

兰州交通大学 电子与信息工程学院，甘肃 兰州 730070

**摘　要：**针对古代壁画分辨率低、纹理细节模糊导致壁画观赏性不足和研究价值不高的问题，本文基于生成对抗网络（GAN）提出了一种能更好地重建细节纹理的壁画图像超分辨率方法。首先，针对浅层图像特征利用不充分，引入信息蒸馏块提取图像浅层特征，增强后面网络的输出结果。其次，用RRDB-Fs提取深层图像特征，去除了残差块中影响图像生成质量的BN层，提高了网络的训练速度。再者，引入局部特征融合和全局特征融合，将不同层次的特征自适应地融合在一起，使重建出的图像含有丰富的细节信息。最后，在计算感知损失时使用激活前的特征，增强重建壁画图像亮度和原壁画图像亮度的一致性，同时避免伪影的产生。实验结果表明，本文所述方法和其他算法相比，具有较好的视觉效果，且重建图像的峰值信噪比和结构相似性指标均获得提高：峰值信噪比提高了0.512 dB—3.016 dB，结构相似性提高了0.009—0.089。

**关键词：**壁画图像；超分辨率重建；生成对抗网络；信息蒸馏块；特征融合

**引用格式：**GAO Li, ZHOU Xiaohui. A super-resolution reconstruction algorithm for mural images based on improved generative adversarial network. Journal of Measurement Science and Instrumentation, 2024, 15(4): 499-508.