ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

Research on high-quality image super-resolution reconstruction algorithm based on deep convolutional generative adversarial network

Zhihao Jiang[1,2,*], *Limi Chen*[1,2] *and Jing Yang*[1]

[1] Hainan Vocational University of Science and Technology, Haikou 571126, China
[2] Institute for Mathematical Research, Universiti Putra Malaysia, Serdang 43400, Malaysia

SUMMARY: *Image is the main carrier of information transmission and presentation today, and high-quality image provides a solid foundation for the development of computer vision field because of its clear picture quality and rich details. In this paper, we take deep learning technology as the basis and optimize the image super-resolution reconstruction algorithm with targeted training, adopt enhanced generative adversarial network, and combine the image cross-level self-similarity features to construct a high-quality image super-resolution reconstruction model based on deep convolutional generative adversarial network. Simulation results show that the algorithm proposed in this paper has better performance, the objective evaluation indexes of the reconstructed image are significantly improved, and the operation speed is relatively fast, the recovered image has more high-frequency detail information, and the visual effect is better.*

KEYWORDS: *super-resolution; generative adversarial network; high quality image; deep convolutional network*

1 Introduction

Science and technology continue to develop, and mankind has entered an informationized big data era. As one of the main data formats, images are widely used in various scenarios, and the spatial resolution of images determines whether urban transportation, medical imaging, satellites, etc. can be intelligently and massively applied [1-3]. Under the application of various technologies images become more refined and rich, laying a high-quality foundation for the development of the computer industry.

Image super-resolution reconstruction task is a fundamental task in the field of computer vision image reconstruction, which refers to the technique of reconstructing a high-resolution (HR) image from a single or multiple low-resolution (LR) images, and recovering a high-resolution image with rich details from a given low-resolution image [4]. Image super-resolution research provides more high-quality input data for computer vision and artificial intelligence applications, and a large number of applications benefit from it to improve such improvements as image style transformation, image denoising, object detection, and image recognition [5-8]. From the hardware point of view, the performance of imaging devices needs to be improved. Currently, the more well-known technologies are mainly focused on the research of image sensor manufacturing, such as improving image sensor performance to enhance the resolution of the image [9]. However, the improvement for image sensor devices and photosensitive units, although more image information and pixel points can be obtained, it

*jiangzhihao256@126.com
https://doi.org/10.65102/is20261300
Published: 8 July, 2026

will affect the charge conversion efficiency and lead to image blurring, and also bring more noise information [10, 11]. Noise, as one of the limiting factors affecting image resolution, leads to image quality degradation, and the improvement of image resolution through the study of instrumentation requires the accumulation of time and the precipitation of technology, and the higher the precision of the imaging equipment, the higher the price of it, which restricts the space for its development [12-14]. Therefore, with the rapid development of information technology, the improvement of the imaging performance of hardware equipment can no longer meet the needs of applications and scientific research, and software-based super-resolution (SR) imaging technology came into being, which can bypass the limitations of the hardware method, effectively improve the image quality, and recover more detailed information of the surface.

In 2015, convolutional neural network (CNN) began to be used for SR (the method is referred to as SRCNN), in view of the poor stability of the image quality reconstructed by this method, the literature [15] designed a SRCNN-based image SR reconstruction method from coarse to fine, which takes the features of the LR image, and reconstructs them from the four modules, namely extraction, enhancement, reconstruction, and refinement, with a step-by-step refinement. Literature [16] accomplished end-to-end multimodal SR MRI reconstruction using CNN and combined with a low-frequency filtering module to screen out the interfering information of the LR image during the high-frequency feature extraction. Residual Learning (RL) was introduced from 2017. Literature [17] utilized the image detail capturing effect of multi-scale global RL and local RL for SR reconstruction of MRI, which outperformed the SRCNN technique. Literature [18] combined SRCNN and RL to develop a single anisotropic magnetic resonance image SR reconstruction method, which not only optimizes the phenomenon of gradient vanishing in SRCNN, but also is effective in reconstructing high-frequency details of images.

The introduction of residual dense networks (RDN) and discriminators was continued in 2018. Literature [19] for SR reconstruction of remote sensing images applied multi-scale convolutional kernel and multi-head self-attention mechanism for multi-scale feature extraction and feature dynamic fusion of remote sensing images, combined with a multi-stage hybrid transformer structure to reconstruct the features step-by-step, and used a multi-fusion discriminator to discriminate the reconstructed images, and added two kinds of loss functions to optimize the reconstruction model. Literature [20] establishes a multi-scale RDN for SR reconstruction of remote sensing images, which is based on splicing and level-by-level transfer of output features from each layer in the network to accomplish an efficient environment for adaptive learning of LR images. Differently, literature [21] innovated a densely sampled SR network for SR reconstruction of remote sensing images at scale, by densely collecting image features at low latitude and optimizing the network by incorporating wide activation and attention mechanisms. Literature [22] developed a residual bottleneck dense network model for single-image SR reconstruction using a joint dense cascade, residual bottleneck structure, and loss function, which not only improves the image resolution in terms of computational speed and cost but also has advantages.

Meanwhile, Generative Adversarial Networks (GANs) have been introduced to perform feature extraction and discrimination with their generator and discriminator respectively for SR image reconstruction. Literature [23] gives two sub-network structures of upsampling and refinement on the generator for feature extraction and refinement of the image, optimizes the generator by combining the perceptual loss, and adds the Wasserstein distance to the discriminator as a way of solving the problems such as feature unambiguity in SR reconstruction of images based on deep learning. Literature [24] designed an each augmented dense GAN in which the generator united multilayer dense connectivity, residual connectivity and Shuffle attention, while the discriminator was designed in PatchGAN style to enhance the

texture details of remote sensing images. Literature [25] improves the image SR reconstruction technique performed by GAN using enhanced U-Net structure to achieve quadruple scale detail reconstruction of remotely sensed LR images. And literature [26] replaced the discriminator to discriminate the feature information in GAN with relative average discriminator, and enhanced the image brightness and texture information by combining the pre-activated features, constructed an improved GAN image SR reconstruction that utilizes the increased shallow information, and the reconstruction speed of this method is far more than the traditional method. The generator and discriminator of literature [27] were created by deep unsupervised and supervised networks, respectively, and the reconstructed SR computed tomography images were obtained by joint parallel 1x1 convolution operation to reduce the dimensionality, and constructing the loss function for the mapping relationship of the generator and discriminator with 4 each loss function. Literature [28] designed the generator and discriminator with recurrent attention network and deep convolutional neural network respectively, combined with the improved loss function to enhance the texture details of infrared images, and realized the SR reconstruction of infrared images. However, the image SR reconstruction of GAN suffers from the defects of unstable generative training, loss of high-frequency features, and non-unique reconstruction results. Literature [29] utilizes the feedback mechanism (generator) and attention mechanism for feature extraction of useful information in the image, and generates SR images in the form of step-by-step feedback, and this method makes the reconstructed images visually expressive and realistic. Deep Convolutional Generative Adversarial Network (DCGAN) is gradually emerging and getting more and more attention. It is a special generative adversarial network structure that can better propagate gradients and improve the efficiency of training by using convolutional and deconvolutional layers, and shows great potential in SR image reconstruction, which fits the needs of super-resolution image generation applications [30].

This chapter first describes the problem of super-resolution reconstruction of graph-quality images, followed by an in-depth analysis of the related techniques of deep learning. A super-resolution reconstruction model for high-quality images is designed using the ESRGAN algorithm combined with the cross-level self-similarity feature of 3D-MRI images to downscale the reconstruction task to 2D. Then simulation tests are conducted using five datasets to compare the model performance from three aspects: qualitative analysis, quantitative comparison, and efficiency comparison. Finally, objective evaluation and ablation comparison experiments are carried out for this paper's model, and 2x and 4x super-resolution reconstruction tests are performed on Set5, Set14 and BSD100 datasets, and the results are analyzed for the experimental results.

2 Theoretical foundations

2.1 Description of the problem of super-resolution reconstruction of high-quality images

The process of acquiring high quality images is limited by a number of factors that can lead to the observation of high quality images of suboptimal quality. The low-resolution high-quality image imaging model is shown in Figure 1. As can be seen from the figure, in the process of low-resolution image imaging, due to optical contamination, inadequate sensor performance and other reasons, will blur the target of interest high-resolution high-quality images, after sampling high-quality image quality is further reduced. The whole process will also be affected by noise, and ultimately only low-resolution high-quality images can be observed. In this

process, the high quality image will lose a lot of details, resulting in the quality of high quality image degradation.

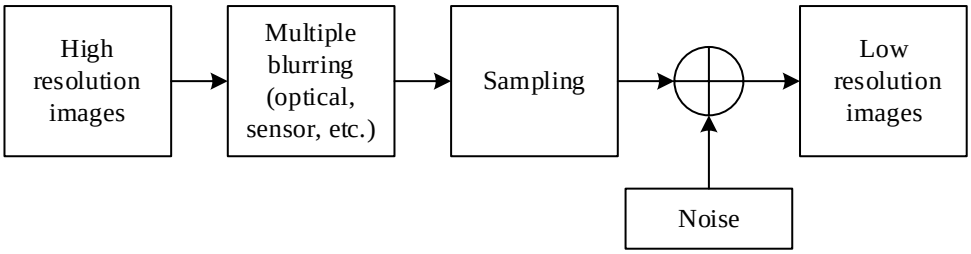


Figure 1: The differential rate image imaging model

In this paper, let I^{HR} and I^{LR} be the original high-resolution high-quality image and the low-resolution high-quality image after the quality reduction, respectively. The relationship between the two can be expressed as follows based on the imaging model of the low-resolution high-quality image of Fig:

$$I^{LR} = D \times B \times I^{HR} + n \tag{1}$$

where D and B are denoted as sampled and blurred matrix forms, and the low-resolution high-quality image is obtained after convolving with the high-resolution high-quality image and superimposing the noise n. This model provides a solid theoretical foundation for the high-quality image super-resolution reconstruction technique. High-quality image super-resolution reconstruction is the inverse solution of the observation model to reconstruct a given low-resolution high-quality image into a high-resolution high-quality image [31]. In this paper high-quality image super-resolution reconstruction refers to super-resolution reconstruction of a single high-quality image. However, high-quality image super-resolution reconstruction is an ill-defined problem due to the presence of many uncertain covariates in this process, which results in the inverse solution not being unique.

2.2 Overview of deep learning techniques

2.2.1 Convolutional Neural Networks

CNN has a powerful ability of representation learning, which is one of the representative algorithms of deep learning. CNN is a kind of neural network containing convolutional layers, and the major breakthroughs made in modern computer vision cannot be achieved without the contribution of CNN. Combined with the research direction of this paper, this section mainly introduces the convolutional layer and pooling layer in CNN.

(1) Convolutional layer

The convolutional layer is to interrelate the convolutional kernel and the feature map, and subsequently output a two-dimensional feature map. In the process of convolutional operation, the first point-to-point multiplication is performed, and the subsequent addition is the final output value. According to the principle of dot product computation, the computational process can be represented as follows:

$$c = a_1 \cdot b_1 + a_2 \cdot b_2 + a_3 \cdot b_3 + a_4 \cdot b_4 \tag{2}$$

Subsequently sliding over the input feature maps in order from left to right and from top to

bottom and repeating the above convolution process. The interval between slides is called the step size s of the convolution.

Other things being equal, different step sizes result in different sizes of the output feature maps. In general, the larger the step size, the smaller the output feature map. Without padding 0 and with a step size of s, let the input feature map size be $H_I \times W_I$ and the convolution kernel size be $k_H \times k_W$, then the output feature map of size $H_O \times W_O$ can be expressed as follows:

$$\left(H_O, W_O\right) = \left(\left\lfloor \frac{H_I - k_H}{s} \right\rfloor + 1, \left\lfloor \frac{W_I - k_W}{s} \right\rfloor + 1 \right) \tag{3}$$

If you need to get a specific size of the feature map, can be in the process of designing CNN will carry out the filling operation, that is, in the feature map of the height and width of the two dimensions of the filling 0. Assuming that the number of filling 0 is p, the size of the input feature map is $H_I \times W_I$, at this time through the size of $k_H \times k_W$ of the convolution for the calculation of the convolution, the size of the output feature map can be be expressed as follows:

$$\left(H_O, W_O\right) = \left(\left\lfloor \frac{H_I + 2p - k_H}{s} \right\rfloor + 1, \left\lfloor \frac{W_I + 2p - k_W}{s} \right\rfloor + 1 \right) \tag{4}$$

For feature maps, they also typically have multiple channels. The number of channels in a high-quality image is denoted as C, and the size of the high-quality image is denoted as $C \times H \times W$. Referring to the two-dimensional convolution, when the input is convolved with multiple channels, each channel is assigned a convolution kernel of $k_H \times k_W$ individually, i.e., the size of the convolution kernel for the convolution is $C \times k_H \times k_W$.

(2) Pooling layer

Through feature selection, the pooling layer can reduce the number of network parameters while effectively preventing overfitting. The pooling layer uses the same sliding window approach as the convolutional layer, sliding over the input array in order from left to right and top to bottom. The step size is kept the same as the step definition in the convolutional layer. The two most commonly used pooling methods are maximum pooling and average pooling.

2.2.2 Activation function

The use of convolutional layers alone can only achieve simple linear mapping capabilities and cannot characterize complex nonlinear transformations from data in the source domain to data in the target domain. In CNNs, activation functions are needed to allow CNNs to approximate complex nonlinear functions, and the introduction of activation functions allows CNNs to be used in many complex nonlinear models.

Sigmoid is one of the commonly used activation functions in CNNs.Sigmoid function, also known as Logistic function, is a special kind of two-end saturation function.Sigmoid function can be expressed as follows:

$$\sigma(x) = \frac{1}{1 + \exp(x)} \tag{5}$$

The Sigmoid function can be viewed as a squeezing function that compresses all data in the

real number domain into the interval $(0,1)$. The smaller the input value, the smaller the output value. This approach is similar to biological neurons in that only specific regions of interest are focused on and disinterested regions are suppressed.

The Tanh activation function is also one of the commonly used activation functions and is represented as follows:

$$\mathrm{Tanh}(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{6}$$

It is not difficult to find that the Tanh function is obtained by scaling the Sigmoid function by a factor of 2 and then translating it one place down along the y-axis, as shown in equation (7):

$$\mathrm{Tanh}(x) = 2\sigma(x) - 1 \tag{7}$$

When the outputs of the above two types of activation functions are close to the limit values at both ends, the gradient of the layer will converge to 0, causing the gradient to disappear and resulting in the values of the weights not continuing to be updated. At the same time, the computational cost of these two types of activation functions is relatively high.

The ReLU activation function is represented as follows:

$$\mathrm{Re}\,LU(x) = \max(0, x) \tag{8}$$

When the input value is less than 0, the output is 0. When the input value is greater than 0, the output value is equal to the input value. Compared to the Sigmoid and Tanh functions, ReLU can alleviate the problem of vanishing gradient to some extent as it is a left saturated function, and at the same time enables the network to converge more quickly. Also the efficiency of ReLU computation is improved due to the use of simple thresholding. However, when the input value is less than 0, the weights can not be updated, the network can not learn some of the features, and the gradient can not be defined when the input value is 0.

2.2.3 Generating Adversarial Networks

The normal iterative operation of the generative adversarial network relies on the generative network G and the discriminative network D. The generative model usually uses a random noise z as its input at the beginning of training, and the output is the target distribution predicted by the model, while the discriminative model uses the output data of the generative model and actual high-quality images as input. In the training process, the generative model and the discriminative model accomplish their own tasks separately, but the tasks of the two are in a game state. In the generative adversarial network, both the generator and the discriminator use a simple forward-structured artificial neural network. Wanting to learn the data distribution P_g of a real high-quality image x, the model acquires the input data as random noise $P_z(z)$, and trains the generator to generate the mapping of the real data on the data space $G(z, \theta_g)$. G is a large differentiable function. The discriminator is $D(z, \theta_g)$, whose role is to continually make judgments about the probability value of the data generated by the generator compared to the real high-quality image, with values ranging from 0 to 1. $D(x)$ is the data distribution of the real data in the model. The entire model is constantly trained G and D

are constantly gamed, training G to minimize $\log\left(1-D\left(G(z)\right)\right)$ [32]. In summary, assuming that the training process about D and G is represented by the function $V\left(G,D\right)$, the total model loss is shown in Equation (9):

$$\begin{aligned}\min_{G}\max_{D}V\left(G,D\right)=&E_{x\sim P_{data}(x)}\left[\log D(x)\right]\\&+E_{z\sim P_{z(z)}}\left[\log\left(1-D\left(G(z)\right)\right)\right]\end{aligned}\tag{9}$$

The network achieves the best results when the discriminator cannot completely distinguish the difference between generated data and real data, which is shown in equation (10):

$$D_G^*(x)=\frac{P_{data}(x)}{P_{data}(x)+P_g(x)}\tag{10}$$

When and only when $P_{data}(x)=P_g(x)$, there is $D(G(z))=0.5$, where D cannot make an effective decision on the input, and G can be considered to have a more desirable generative effect to reach the Nash equilibrium discriminator to reach the best solution. Generating adversarial networks allows for a better and faster grasp of the distribution of training samples in unsupervised machine learning. This is due to mutual conflicts as well as advancing the generative and discriminative models to generate targets from random noise inputs. In the training process, the complete GAN training includes data preprocessing preparation, network structure setup of generator and discriminator, parameter setting, loss function setting, etc., to carry out the training to the target model, keeping only the generator.

2.3 High-quality image evaluation metrics

2.3.1 Peak signal-to-noise ratio

Peak Signal to Noise Ratio (PSNR) is an easy and practical objective calculation for similarity based on the mean square error of two high quality images in decibels, which is defined as:

$$PSNR=10\cdot\log_{10}\left(\frac{I_{\max}^2}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{HR}\left(i,j\right)-I_{SR}\left(i,j\right)\right)^2}\right)\tag{11}$$

where $I_{\max}$ represents the maximum pixel value of the high-quality image, M represents the height of the high-quality image, N represents the width of the high-quality image, I_{HR} represents the original real high-resolution high-quality image, and I_{SR} represents the high-quality image reconstructed by the algorithm. Usually the higher the PSNR value, the better the quality of the reconstructed high quality image. The method is relatively simple and easy to operate, but it may also deviate from the perception of the human eye.

2.3.2 Structural similarity

Structural similarity (SSIM) compares the similarity between high-quality images in terms of their luminance $\mathcal{L}$, contrast C and structure $\mathcal{S}$ with the following mathematical expression:

$$\begin{aligned}\mathcal{L}\left(I_{HR},I_{SR}\right)&=\frac{2\mu_1\mu_2+c_1}{\mu_1^2+\mu_2^2+c_1}\\ C\left(I_{HR},I_{SR}\right)&=\frac{2\sigma_1\sigma_2+c_2}{\sigma_1^2+\sigma_2^2+c_2}\\ \mathcal{S}\left(I_{HR},I_{SR}\right)&=\frac{\sigma_{1,2}+c_3}{\sigma_1\sigma_2+c_3}\\ SSIM\left(I_{HR},I_{SR}\right)&=\left[\mathcal{L}\left(I_{HR},I_{SR}\right)\right]^{\alpha}\left[C\left(I_{HR},I_{SR}\right)\right]^{\beta}\left[\mathcal{S}\left(I_{HR},I_{SR}\right)\right]^{\gamma}\end{aligned} \tag{12}$$

where μ_1, μ_2 denote the mean of I_{HR} and I_{SR}, σ_1, σ_2 denote the standard deviation of I_{HR} and I_{SR}, respectively, and $\sigma_{1,2}$ denotes the covariance of I_{HR} and I_{SR}, c_1,c_2,c_3 are normal numbers with smaller values, and α,β,γ are constants.

3 DCGAN-based super-resolution reconstruction model for high-quality images

3.1 DCGAN-based super-resolution reconstruction model design for high-quality images

3.1.1 Model framework design

In this study, the ESRGAN algorithm is mainly used to design a super-resolution reconstruction model for 3D-MRI high-quality images, and the cross-dimensional self-similarity of 3D-MRI high-quality images is utilized to reduce the dimensionality of the super-resolution reconstruction task for training. The main framework and realization steps of the designed reconstruction model are shown in Fig. 2, in which the anisotropic 3D-MRI high-quality image represents a low-resolution 3D-MRI high-quality image with two dimensions of low resolution, and the isotropic 3D-MRI high-quality image represents a high-resolution 3D-MRI high-quality image with high resolution in all dimensions after reconstruction, and the reconstructed 3D-MRI high-quality image resolution is increased to twice of the original [33].

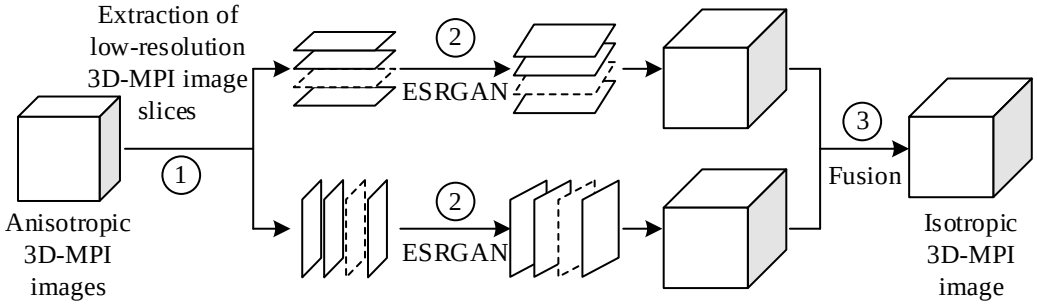


Figure 2: The algorithm framework of this paper

① Extraction of low-resolution 3D-MRI high-quality image slices: low-resolution two-dimensional slices of anisotropic 3D-MRI high-quality images are extracted high-quality images in accordance with different levels respectively, and this process extracts low-resolution high-quality images of two levels.

② Super-resolution reconstruction: the two sets of low-resolution two-dimensional slices

are input into the ESRGAN network separately to reconstruct high-resolution two-dimensional high-quality images, in order to recover the detail information of the high-quality images, and ESRGAN is one of the variants improved from DCGAN.

③ Fusion of high-quality images: the high-resolution 2D high-quality images after super-resolution reconstruction are respectively restored in accordance with the original index positions at the time of extracting the slices in order to stack back the 3D-MRI high-quality images, and then the resulting two 3D-MRI high-quality images are fused together to ultimately produce an isotropic high-resolution 3D-MRI high-quality image. Wherein the fusion operation is to take the mean value of the corresponding voxels of the two high-quality images and fuse them.

3.1.2 Model Training Design

Because 3D-MRI high-quality images are multiparameter, native three-dimensional imaging, high-quality images have rich a priori information by themselves, and high-quality images at different scanning levels have cross-level self-similarity, in this study, we constructed a super-resolution reconstruction dataset of anisotropic 3D-MRI high-quality images based on the above characteristics of 3D-MRI high-quality images.Cross-level self-similarity in 3D-MRI high-quality images refers to the fact that there are many similar tissue structures at the same or different scales in two-dimensional sliced high-quality images of the same 3D-MRI high-quality image. Similarity refers to the fact that there are many similarity organizational structures of the same or different scales in different levels of 2D slice high-quality images of the same 3D-MRI high-quality image. Therefore, in this study, three low-resolution anisotropic 3D-MRI high-quality images were simulated for the original high-resolution 3D-MRI high-quality image O (Original) using mean downsampling to generate three low-resolution anisotropic 3D-MRI high-quality images, i.e., the transverse-axial low-resolution 3D-MRI high-quality image A (Axial low resolution), the coronal low-resolution 3D-MRI high-quality image C (Coronal low resolution), and Sagittal low resolution 3D-MRI high quality image S (Sagittal low resolution). Then the following training set is generated based on the above high-quality images, i.e., high- and low-resolution high-quality image pairs: the coronal plane of high-quality image O with the coronal plane of high-quality image A, the sagittal plane of high-quality image O with the sagittal plane of high-quality image A, the cross-section of high-quality image O with the cross-section of high-quality image C, the sagittal plane of high-quality image O with the sagittal plane of high-quality image C, and the cross-section of high-quality image O with the cross-section of high-quality image S. cross section with the cross section of high quality image S, and the coronal plane of high quality image O with the coronal plane of high quality image S. This idea expands the data volume of the training set by dimensionality reduction. For example, one high-resolution 3D-MRI high-quality image can extract 1364 2D-MRI high-quality images as the training set. According to the cross-level self-similarity of 3D-MRI high-quality images, the training set size can be expanded effectively, which plays an important role in fully utilizing the performance of deep neural networks. Using 3D-MRI high-quality image slices with different level orientations to train the network simultaneously enables a network to simultaneously super-resolution reconstruct 3D-MRI high-quality image slices with different level orientations without having to train different super-resolution reconstruction networks for different level-selected orientations separately. In addition, this method can utilize the three-dimensional feature information in 3D-MRI high-quality images to enhance the performance of high-quality image SR reconstruction.

3.2 ESRGAN Network Architecture

ESRGAN is improved on the basis of SRGAN network, which is one of the variants of DCGAN,

which is a product of combining GAN and CNN.ESRGAN contains a generative network and a discriminative network. With continuous feedback from a discriminative network that constantly improves its judgment, the generative parameters of the generative network are continuously improved until the results generated by the generative network can pass the judgment of the discriminative network.

3.2.1 ESRGAN Generative Networks

The generative network part of ESRGAN still adopts the basic network architecture of SRResNet, and the SRResNet structure (generative network of ESRGAN) is shown in Fig. 3. Most of these computations are done in the LR feature space [34]. The computation in the LR feature space will process a smaller amount of data and therefore consume less computation and memory, which can improve the speed of super-resolution reconstruction. The original SRResNet network architecture contains 16 basic residual blocks, and each residual block contains two 3×3 convolutional layers, which are followed by a bulk normalization layer (BN) and parametric ReLU as activation function, and two sub-pixel convolutional layers of scale 2 are used to increase the feature map size.

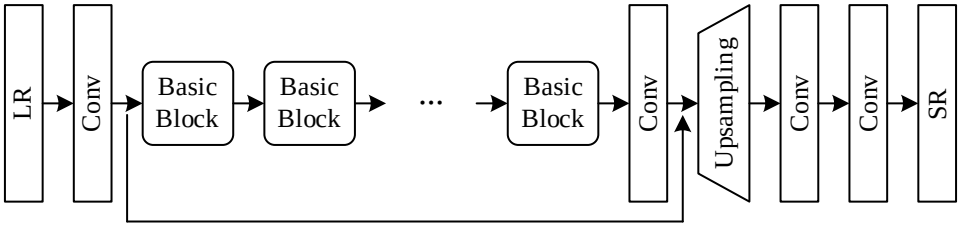


Figure 3: The structure of Generative Network of ESRGAN

3.2.2 ESRGAN Discriminative Networks

ESRGAN uses a relative discriminative network instead of the criterion discriminative network in SRGAN.The criterion discriminative network in SRGAN predicts the probability that an input high-quality image is both a true high-quality image and a pseudo-quality image. Unlike the standard discriminative network in SRGAN, the relative discriminative network attempts to predict the probability that a real high-quality image x_r is relatively more real than a pseudo-quality image x_f. In SRGAN, the standard discriminative network is denoted as:

$$D(x) = \sigma\left(C(x)\right) \tag{13}$$

where σ is the sigmoid function and $C(x)$ is the non-transformed discriminant network output. Thus, the relative discriminative network is given by:

$$D_{Ra}\left(x_r, x_f\right) = \sigma\left(C\left(x_r\right) - E_{x_f}\left[C\left(x_f\right)\right]\right) \tag{14}$$

where $E_{x_f}\left[C\left(x_f\right)\right]$ stands for averaging all false data in a small batch. As a result, the relative discriminative network loss function is defined as:

$$L_D^{Ra} = -E_{x_r}\left[\log\left(D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{x_f}\left[\log\left(1 - D_{Ra}\left(x_f, x_r\right)\right)\right] \tag{15}$$

The adversarial loss function of the corresponding generative network is defined as:

$$L_G^{Ra} = -E_{x_r}\left[\log\left(1 - D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{x_f}\left[\log\left(D_{Ra}\left(x_f, x_r\right)\right)\right] \tag{16}$$

Therefore, the generative network in the ESRGAN network used in this study is able to utilize both the gradient of the generative data and the gradient of the real data for learning during the adversarial training process. Whereas, only the gradient of the generated data in the SRGAN network plays a role in the process of adversarial training.

3.2.3 Loss function based on visual perception mechanism

A more effective method of perceptual loss calculation is proposed in the ESRGAN model, i.e., using the features before the activation layer to calculate the perceptual loss, which can make the reconstructed high-quality images have sharper edges and richer textures, thus improving the quality of high-frequency information in high-quality images. At the same time, the brightness of the reconstructed high-quality image is made closer to the real high-quality image. In order to further improve the visual quality of 3D-MRI high-quality image reconstruction results, this study uses a loss function based on the visual mechanism, which contains three main components: perceptual loss, adversarial loss and content loss. The calculation formula of the loss function is shown in Equation (17):

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \tag{17}$$

where L_1 is the 1-paradigm distance content loss evaluated between the reconstructed high-quality image and the real high-quality image, and λ and η are the coefficients that balance the different loss terms.

4 Experiments and analysis of results

4.1 Model Performance Testing

4.1.1 Experimental data set

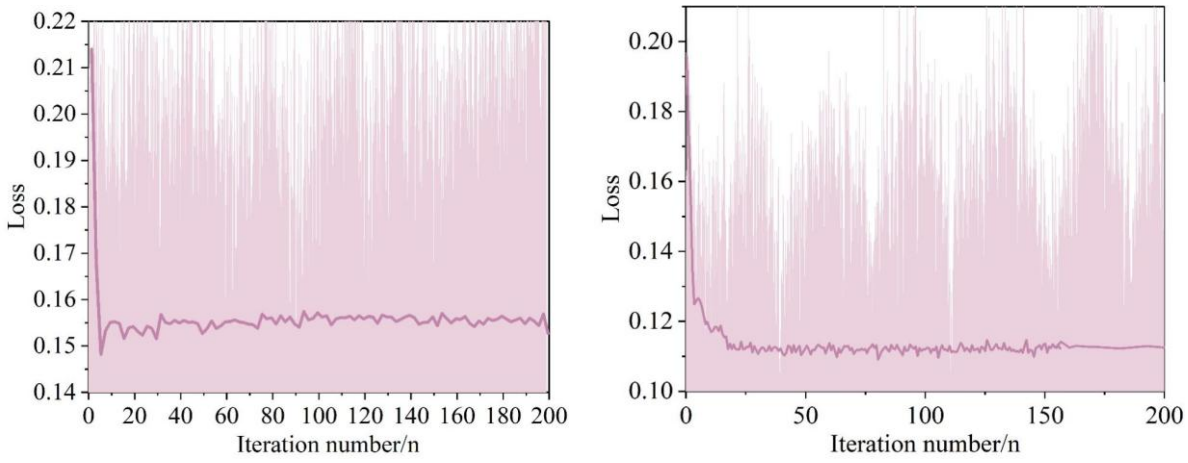
In this section, VOC2012 is used as the training dataset for the network model, and the training and test sets are 16000 and 450 images, respectively. The test dataset is selected from four benchmark datasets widely used to evaluate the quality of super-resolution reconstruction of high-quality images, i.e., Set5, Set14, BSD100, and Urban100, and also in this paper, an other dataset with a total of 14 images is collected for the testing of the algorithm. The test datasets contain a large variety of high-quality images, and the texture, color, and contrast in the high-quality images are also richer, which can fully and effectively evaluate the performance of the model.

4.1.2 Experimental results and comparative analysis

(1) **Comparison** experiment before and after model optimization

Taking the SRGAN model as a benchmark, the stability of the model is verified by the change curve of the loss function. Under 4 times magnification, 200 iterations of the two models were performed respectively, and the loss change curves before and after the improvement are shown in Fig. 4 (Fig. a and Fig. b represent the loss change curves before and after the improvement, respectively). From the figure, it can be seen that the loss function value of the

SRGAN model is close to the steady state after the 60th iteration, but there are still small local fluctuations. The minimum value of 0.1526251 is reached after the 140th iteration, while the loss function value of the optimized model is close to the steady state after the 23rd iteration, and although there are also small local fluctuations, the fluctuations are relatively small. After the 126th iteration, it reaches the minimum value of 0.109821 in the steady state.In addition, it can be seen from the figure that compared with the SRGAN method, the iteration curve of this paper's method is smoother and less fluctuating, especially in the first 40 iterations, this paper's method is more stable and requires less training to obtain the optimal solution. The experimental results show that the optimized network structure used in this paper improves the stability of the network, accelerates the model convergence, and improves the performance of the model.



(a) Before (b) After

Figure 4: Improvement curve

(2) Experimental comparison after optimization of network structure and loss function

The effects of this chapter's optimization of network structure and loss function on the quality of reconstructed high-quality images are explored on Set5, Set14 and Urban100 datasets. The comparison of the average PSNR values of reconstructed high-quality images for models with different optimization measures is shown in Table 1. Comparison of PSNR values of reconstructed high-quality images from the model under four different conditions, namely, dense residual block structure, improved adversarial loss L_g, improved perceptual loss L_{percep}, and texture loss L_{tex} are introduced respectively, where "√ " indicates that the corresponding optimization measures are added, and "×" indicates that the corresponding optimization measures are not added. As can be seen from the table, several optimization measures are able to improve the pixel accuracy of reconstructed high-quality images to different degrees, and the PSNR value of reconstructed high-quality images is the highest and the effect is the best when all the four optimization measures are applied in Model 5.

Table 1: Comparison of the average PSNR value of the model reconstruction image

	Model 1	Model 2	Model 3	Model 4	Model 5
Density block	×	√	×	×	√
L_g	×	×	×	√	√
L_{percep}	×	×	√	√	√
L_{tex}	×	×	√	√	√
Set5	31.41	32.35	32.68	32.48	32.65
Set14	28.18	28.99	29.61	29.74	30.14
BSD100	27.43	27.79	27.55	27.44	27.81

The method of this paper is tested against other methods on the benchmark datasets Set5, Set14, BSD100, Urban100 and other datasets collected in this paper to reconstruct the objective evaluation metric values of PSNR and SSIM for high quality images, and the average PSNR values of the different methods on Set5, Set14, BSD100, Urban100 and other datasets and the average SSIM values are compared as shown in Fig. 5 and Fig. 6. As can be seen from the figures, the test results of the methods proposed in this chapter outperform the SRGAN method on all the five test datasets and achieve the optimal PSNR and SSIM values for reconstructing high-quality images when compared to other deep learning methods (SRCNN, FSRCNN, VDSR, SRGAN, CGAN-SRGAN, WGAN-SRGAN). Compared with SRCNN method, the PSNR values are improved by 2.56dB, 2.39dB, 0.87dB and 2.17dB on Set5, Set14, BSD100 and, Urban100 datasets, respectively, and the SSIM values are improved by 0.0462, 0.052, 0.0832, and 0.112, respectively. Meanwhile, compared with the original SRGAN algorithm compared to the original SRGAN algorithm, the PSNR values are improved by 1.65dB, 1.72dB, 0.42dB and 1.16dB on Set5, Set14, BSD100 and on Urban100 dataset, respectively, and the SSIM values are improved by 0.024, 0.018, 0.0442, and 0.0536. Analyzed from the perspective of the objective evaluation factors, the method of this paper obtains the optimal PSNR value and SSIM value, which indicates that this paper's method improves the pixel accuracy of high-quality images to some extent.

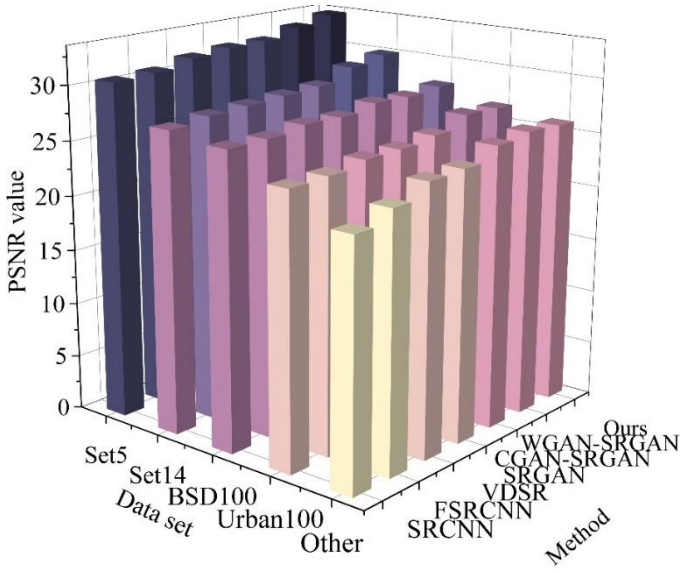


Figure 5: Average PSNR

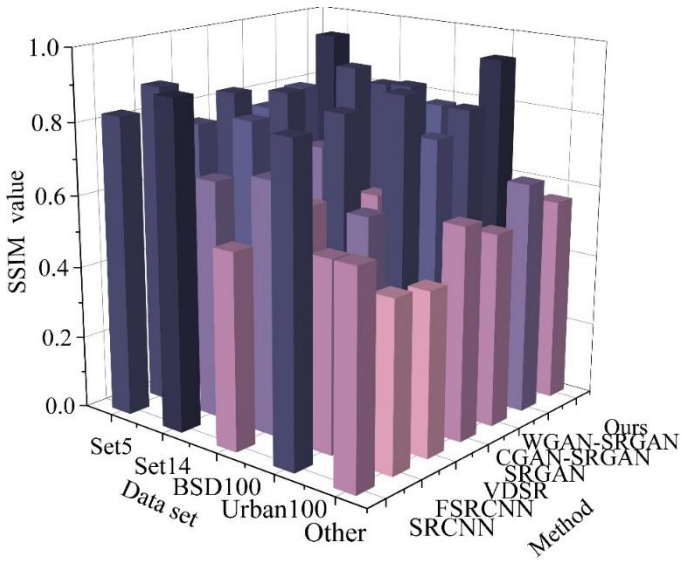


Figure 6: Average SSIM

(3) Efficiency Comparison

The comparison of the running time and accuracy of different models is shown in Fig. 7. As can be seen from the figure, the running time of this paper's method is faster than both the VDSR method and the WGAN-SRGAN model, and it has the highest accuracy in reconstructing high-quality images, which is slightly slower than the SRGAN method and the WGAN-SRGAN model. It can be seen that the model in this chapter can achieve a better compromise and run relatively fast while obtaining the best reconstruction accuracy.

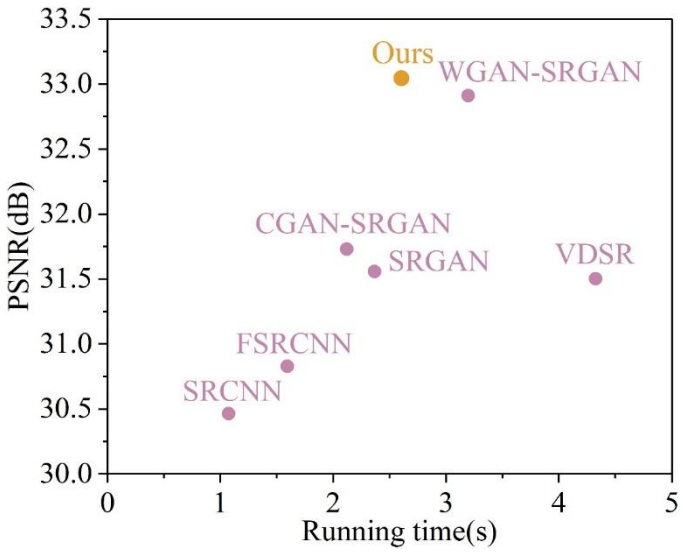


Figure 7: The operating time of different models is compared with accuracy

4.2 Comparison of the two optimization algorithms in this paper

4.2.1 Comparison of objective evaluation indicators

The experiments in this section show the comparison results of PSNR and SSIM values for reconstructing high quality images on five test datasets for the optimized algorithm of this paper (ESRGAN-R), ESRWGAN-R and ESRGAN algorithms. The results of comparison of PSNR

values and comparison of SSIM values are shown in Fig. 8 and Fig. 9, respectively. Observing Fig. 8, it can be seen that the average PSNR value of the ESRGAN-R model optimized in this paper for reconstructing high-quality images is higher than that of the ESRGAN model on the Set5 dataset, and close to the ESRGAN model on the Set14 dataset. The average SSIM value of the optimized ESRGAN-R model in this paper is higher than the ESRGAN model in reconstructing high-quality images on each dataset. It can be seen that the optimized algorithm in this paper improves the quality of reconstructed high-quality images in different degrees in terms of PSNR values compared to the original ESRGAN algorithm.

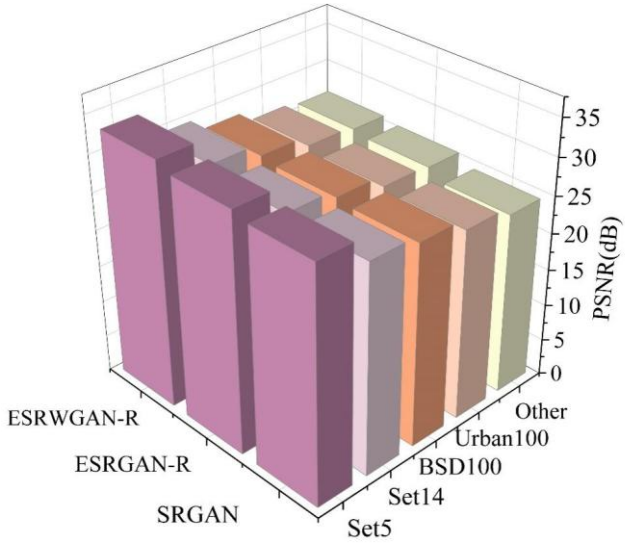


Figure 8: Comparison of PSNR values

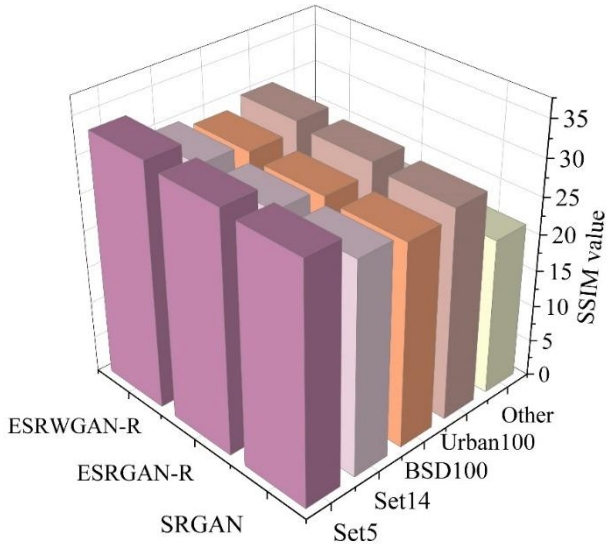


Figure 9: Comparison of SSIM values

4.2.2 Comparison of reconstruction efficiency

The comparison of the running time and accuracy of different models under Set5 dataset is shown in Fig. 10. As can be seen from the figure, the optimized ESRGAN-R model in this paper has a faster running time than the ESRGAN algorithm on the Set5 dataset, and the average PSNR value of reconstructed high-quality images is also higher than that of ESRGAN, which indicates that the ESRGAN-R algorithm has a higher reconstruction efficiency than the

ESRGAN algorithm, and it is more suitable for the high reconstruction speed but relatively low reconstruction accuracy requirement of the application scenarios. The ESRWGAN-R model has slower running time than ESRGAN algorithm on Set5 dataset, but has the highest average PSNR value for reconstructing high-quality images. The ESRWGAN-R algorithm is more suitable for application scenarios that require high reconstruction accuracy but not high reconstruction speed.

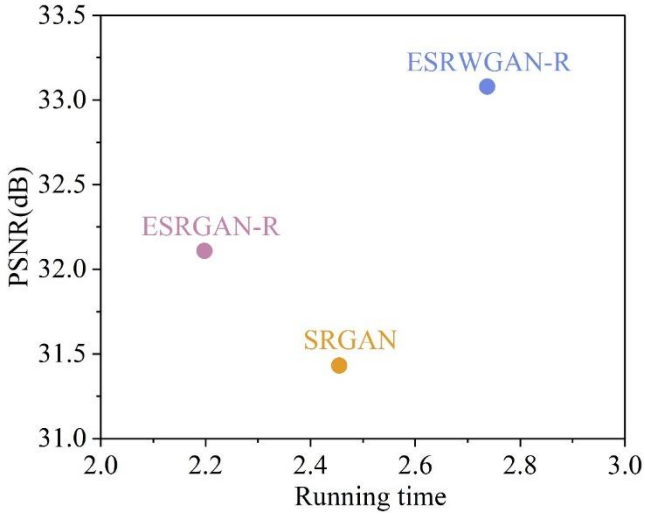


Figure 10: The operating time and accuracy of different models in the set5 data set are compared

4.3 Objective evaluation

The results of the objective evaluation of the 2x and 4x super-resolution reconstruction algorithms on the public test sets Set5, Set14, and BSD100, the comparison of PSNR and SSIM of the 2x and 4x reconstruction results of the algorithms are shown in Table 2 and Table 3. From the table, it can be found that the objective metrics obtained by the algorithm proposed in this paper are not the highest, this is because the algorithm in this paper is trained based on the antagonistic loss and perceptual loss, instead of PSNR and SSIM oriented training, which is mainly concerned with the recovery of details in high quality images, and the PSNR and SSIM do not represent the real visual effect of high quality images.Ours-L1 is trained by pixel-level loss oriented model trained on PSNR, and the objective metrics obtained are all maximized, but the visual effect of reconstructing high-quality images is prone to blurring and is far inferior to the algorithm in this paper.

Table 2: Two times the algorithm and four times the reconstruction results of PSNR comparison

Dataset	Scale	Bicubic	FSRCNN	ESRGAN	Ours-L1	Ours
Set5	2×	31.753	33.296	32.387	35.855	33.028
	4×	26.937	27.593	27.863	29.801	28.197
Set14	2×	28.464	29.538	28.73	31.605	29.381
	4×	24.268	25.012	24.768	26.599	24.824
BSD100	2×	26.848	27	25.909	27.44	26.035
	4×	23.585	23.948	22.501	24.113	22.68

Table 3: Two times the algorithm and four times the reconstruction results of SSIM comparison

Dataset	Scale	Bicubic	FSRCNN	ESRGAN	Ours-L1	Ours
Set5	2×	0.958	0.933	0.847	1.007	0.907
	4×	0.896	0.748	0.809	0.792	0.89
Set14	2×	0.901	0.771	0.838	1.07	0.898
	4×	0.563	0.586	0.591	0.755	0.818
BSD100	2×	0.608	0.749	0.86	0.565	0.598
	4×	0.571	0.356	0.643	0.592	0.534

4.4 Comparison of ablation experiments

In order to explore the impact of the improvements of the generator and the discriminator on the performance of the ESRGAN network structure, respectively, two new model networks are proposed in this paper. Among them, Model I makes changes to the generator network and the discriminator network uses the relative discriminator in ESRGAN networks. Model II only makes changes to the discriminative network and loss function. Model I and Model II are trained and ESRGAN are compared with the values of PSNR and SSIM on the four validation sets, respectively. The performance comparison of Model I, Model II and ESRGAN networks are shown in Table 4 and Table 5. From the table, it can be seen that both models are numerically improved compared to the ESRGAN network, which is specifically analyzed that Model I improves the PSNR and SSIM values on the four benchmark test sets by 0.61% and 5.48%, respectively, when scaled up by a factor of 2, and when scaled up by a factor of 3, it improves the PSNR and SSIM values on the four benchmark test sets by 0.095% and 7.33%, and in the case of zooming in 4 times, the PSNR and SSIM values on the four benchmark test sets are improved by 0.14% and 11.37%, respectively.

Model II improves the PSNR and SSIM values on the four benchmark test sets by 0.108% and 6.05%, respectively, at a magnification of 2x, 0.68% and 8.45%, respectively, at a magnification of 3x, and at a magnification of 4x, the PSNR and SSIM values are improved by 0.28% and 3.8%, respectively. Based on the results of ablation experiments it is illustrated that the design of residual feature aggregation network based on improved generative adversarial network proposed in this paper contributes to the metrics and visualization enhancement of high-quality images for super-resolution reconstruction of high-quality images.

Table 4: Model 1 and ESRGAN network performance comparison

Method	Scale	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM
ESRGAN	×2	35.63/0.9336	30.82/0.909	30.46/0.9747	29.89/0.9489
Model 1		36.18/0.9622	30.63/0.8809	30.85/0.8029	29.96/0.9086
ESRGAN	×3	31.77/0.9928	28.49/0.8138	27.1/0.9247	26.25/0.7489
Model 1		31.55/0.9906	28.24/0.6954	27.35/0.6715	26.52/0.845
ESRGAN	×4	29.22/0.8725	26.02/0.759	26.09/0.7455	24.56/0.5845
Model 1		29.46/0.7936	26.12/0.9775	25.81/0.7564	24.66/0.7263

Table 5: Model 2 and ESRGAN network performance comparison

Method	Scale	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM
ESRGAN	×2	35.63/0.9336	30.82/0.909	30.46/0.9747	29.89/0.9489
Model 2		35.62/0.9399	30.96/0.8699	30.3/0.8391	29.79/0.8857
ESRGAN	×3	31.77/0.9928	28.49/0.8138	27.1/0.9247	26.25/0.7489
Model 2		31.41/0.853	28.11/0.8021	27.15/0.7245	26.13/0.774
ESRGAN	×4	29.22/0.8725	26.02/0.759	26.09/0.7455	24.56/0.5845
Model 2		29.17/0.8262	26.37/0.7163	25.73/0.6571	24.89/0.6288

5 Conclusion

Based on the theory related to deep learning technology, the article proposes a high-quality image super-resolution reconstruction model based on DCGAN, which solves the problem of unclear details in low-resolution images. The final experimental results show that the image texture details reconstructed by the algorithm proposed in this paper are clearer, which greatly improves the visual effect of the image.

In the model performance comparison experiments, the loss function value of this paper's model is close to the steady state after the 23rd iteration, and reaches the minimum value of 0.109821 in the steady state after the 126th, and in the first 40 iterations, this paper's method is more stable, and less training is needed to obtain the optimal solution.

In the case of scaling up by a factor of 3, Model I and Model II (Model I makes changes to the generative network. Model II only makes changes to the discriminative network and loss function.) The PSNR and SSIM values on the four benchmark test sets are improved by 0.095%, 0.68% and 7.33%, 8.45%, respectively. The experimental results demonstrate the superiority of the model in this paper.

Funding

This research was supported by the Youth Program of the Natural Science Foundation of Hainan Province, China, grant number 625QN349

References

[1] Deng, Y., Zhou, L., Wang, L., Su, M., Zhang, J., Lian, J., & Wei, J. (2020). Radio environment map construction using super-resolution imaging for intelligent transportation systems. IEEE Access, 8, 47272-47281.

[2] Hasegawa, H. (2017). Improvement of range spatial resolution of medical ultrasound imaging by element-domain signal processing. Japanese Journal of Applied Physics, 56(7S1), 07JF02.

[3] Fisher, J. R., Acosta, E. A., Dennedy-Frank, P. J., Kroeger, T., & Boucher, T. M. (2018). Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. Remote Sensing in Ecology and Conservation, 4(2), 137-149.

[4] Yanqiu, T., Hong, P., Yaping, Z., & Xinde, L. (2020). A review of image super-resolution

reconstruction research. J. Electron, 48, 1407-1420.

[5] Park, S. H., Moon, Y. S., & Cho, N. I. (2022). Flexible style image super-resolution using conditional objective. IEEE Access, 10, 9774-9792.

[6] El Mourabit, I., El Rhabi, M., Hakim, A., Laghrib, A., & Moreau, E. (2017). A new denoising model for multi-frame super-resolution image reconstruction. Signal Processing, 132, 51-65.

[7] Wang, Y., Bashir, S. M. A., Khan, M., Ullah, Q., Wang, R., Song, Y., ... & Niu, Y. (2022). Remote sensing image super-resolution and object detection: Benchmark and state of the art. Expert Systems with Applications, 197, 116793.

[8] Cao, F., & Su, M. (2022, May). Research on face recognition algorithm based on CNN and image super-resolution reconstruction. In 2022 IEEE 8th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) (pp. 157-161). IEEE.

[9] Li, Y., Zhao, K., Ren, F., Wang, B., & Zhao, J. (2020). Research on super-resolution image reconstruction based on low-resolution infrared sensor. IEEE Access, 8, 69186-69199.

[10] Grishentsev, A., Elsukov, A., Korobeynikov, A., & Arustamov, S. (2020). Method for achieving the super resolution of photosensitive matrices. In Proceedings of the Fourth International Scientific Conference “Intelligent Information Technologies for Industry”(IITI’19) 4 (pp. 573-580). Springer International Publishing.

[11] Almasri, F., & Debeir, O. (2019). Multimodal sensor fusion in single thermal image super-resolution. In Computer Vision–ACCV 2018 Workshops: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers 14 (pp. 418-433). Springer International Publishing.

[12] Goyal, B., Agrawal, S., & Sohi, B. S. (2018). Noise issues prevailing in various types of medical images. Biomedical & Pharmacology Journal, 11(3), 1227.

[13] Liu, J., Malekzadeh, M., Mirian, N., Song, T. A., Liu, C., & Dutta, J. (2021). Artificial intelligence-based image enhancement in PET imaging: noise reduction and resolution enhancement. PET clinics, 16(4), 553.

[14] Wei, Z., Zhang, J., Xu, Z., Liu, Y., Huang, Y., & Fan, X. (2019). Improving the signal-to-noise ratio of superresolution imaging based on single-pixel camera. IEEE Photonics Journal, 11(1), 1-16.

[15] Tian, C., Xu, Y., Zuo, W., Zhang, B., Fei, L., & Lin, C. W. (2020). Coarse-to-fine CNN for image super-resolution. IEEE Transactions on Multimedia, 23, 1489-1502.

[16] Kang, L., Tang, B., Huang, J., & Li, J. (2024). 3D-MRI super-resolution reconstruction using multi-modality based on multi-resolution CNN. Comput. Methods Programs Biomed., 248, 108110.

[17] Shi, J., Liu, Q., Wang, C., Zhang, Q., Ying, S., & Xu, H. (2018). Super-resolution

reconstruction of MR image with a novel residual learning network algorithm. Physics in Medicine & Biology, 63(8), 085011.

[18] Du, J., He, Z., Wang, L., Gholipour, A., Zhou, Z., Chen, D., & Jia, Y. (2020). Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. Neurocomputing, 392, 209-220.

[19] Liu, D., Zhong, L., Wu, H., Li, S., & Li, Y. (2025). Remote sensing image Super-resolution reconstruction by fusing multi-scale receptive fields and hybrid transformer. Scientific Reports, 15(1), 2140.

[20] Kong, D., Gu, L., Li, X., & Gao, F. (2024). Multi-scale residual dense network for the super-resolution of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing.

[21] Dong, X., Sun, X., Jia, X., Xi, Z., Gao, L., & Zhang, B. (2020). Remote sensing image super-resolution using novel dense-sampling networks. IEEE Transactions on Geoscience and Remote Sensing, 59(2), 1618-1633.

[22] An, Z., Zhang, J., Sheng, Z., Er, X., & Lv, J. (2021). RBDN: Residual bottleneck dense network for image super-resolution. IEEE Access, 9, 103440-103451.

[23] Wu, Y., Lan, L., Long, H., Kong, G., Duan, X., & Xu, C. (2020). Image super-resolution reconstruction based on a generative adversarial network. IEEE Access, 8, 215133-215144.

[24] Guo, M., Xiong, F., Zhao, B., Huang, Y., Xie, Z., Wu, L., ... & Zhang, J. (2024). TDEGAN: A texture-detail-enhanced dense generative adversarial network for remote sensing image super-resolution. Remote Sensing, 16(13), 2312.

[25] Yue, X., Liu, D., Wang, L., Benediktsson, J. A., Meng, L., & Deng, L. (2023). IESRGAN: enhanced U-net structured generative adversarial network for remote sensing image super-resolution reconstruction. Remote Sensing, 15(14), 3490.

[26] Sun, X., Zhao, Z., Zhang, S., Liu, J., Yang, X., & Zhou, C. (2020). Image super-resolution reconstruction using generative adversarial networks based on wide-channel activation. Ieee Access, 8, 33838-33854.

[27] Jiang, X., Liu, M., Zhao, F., Liu, X., & Zhou, H. (2020). A novel super-resolution CT image reconstruction via semi-supervised generative adversarial network. Neural Computing and Applications, 32, 14563-14578.

[28] Liu, Q. M., Jia, R. S., Liu, Y. B., Sun, H. B., Yu, J. Z., & Sun, H. M. (2021). Infrared image super-resolution reconstruction by using generative adversarial network with an attention mechanism. Applied Intelligence, 51, 2018-2030.

[29] Wang, Y., Li, X., Nan, F., Liu, F., Li, H., Wang, H., & Qian, Y. (2022). Image super-resolution reconstruction based on generative adversarial network model with feedback and attention mechanisms. Multimedia Tools and Applications, 81(5), 6633-6652.

[30] Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., ... & Shen, D. (2018).

Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering, 65(12), 2720-2730.

[31] Fuqiang Lu, Xia Xiao, Zengxiang Wang, Yu Liu & Jiannan Zhou. (2025). DFFKI-Net: A dual-domain feature fusion deep convolutional neural network for under-sampled MR image reconstruction. Biomedical Signal Processing and Control,106,107732-107732.

[32] Luoyi Li, Lintao Zheng,Chunlei Yang & Yongsheng Dong. (2025). GAL-GAN: Global styles and local high-frequency learning based generative adversarial network for image cartoonization. Computers and Electrical Engineering,123(PC),110164-110164.

[33] Nitisukanan Nidchapan, Boonthaweechok Chotika, Tiawpanichkij Prapatsorn, Pissakul Juthamas, Maneesawangwong Naliya & Siriborvornratanakul Thitirat. (2024). StyleGAN2-ADA and Real-ESRGAN: Thai font generation with generative adversarial networks. Advances in Computational Intelligence,4(1).

[34] Choi Yoonsil & Park Hanhoon. (2022). Improving ESRGAN with an additional image quality loss. Multimedia Tools and Applications,82(2),3123-3137.