IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 December 2021, accepted 7 January 2022, date of publication 11 January 2022, date of current version 25 September 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3142049

REGULAR

An Image Super-Resolution Reconstruction Method Based on PEGAN

CHANG-WEI JING[1], ZHI-XING HUANG[2], AND ZAI-YING LING[1]
[1]Zhejiang Provincial Key Laboratory of Urban Wetland and Regional Change, Hangzhou 311121, China
[2]Zhejiang Key Laboratory of Exploitation and Preservation of Coastal Bio-Resources, Zhejiang Mariculture Research Institute, Wenzhou 325000, China

Corresponding author: Chang-Wei Jing (changweij@hznu.edu.cn)

This work was supported in part by the Scientific Research Foundation of Hangzhou Normal University under Grant 4085C5021820437.

ABSTRACT In order to improve the accuracy and efficiency of super-resolution reconstruction of low-resolution images, a multi-scale and multi-stage self similar fusion image super-resolution reconstruction algorithm is proposed: firstly, the low-frequency features of the image are obtained by using the feature extraction network and used as the input of two sub networks, one of which obtains the structural feature information of low-resolution images through the coding network, Secondly, the high-frequency features are obtained through the multi-path feedforward network composed of stage feature fusion unit, in which the fusion unit fuses the features of several consecutive layers of the network and obtains effective features in an adaptive way; Finally, the residual block and the batch normalization layer unfavorable to the image super-resolution in the discriminator are removed. Spectral normalization is used in the generator and discriminator to reduce the computational overhead and stabilize the model training. The experimental results show that Our method makes the average PSNR of the reconstructed images on the two test data sets of Remo-A dataset and Remo-B dataset reach 27.95 dB and the average SSIM reach 0.771 without losing too much speed. This model draws lessons from the connection mode of dense network to strengthen the connection between network layers and connect the whole network through multipath connection, so as to make full use of the characteristics of hierarchical network, extract more high-frequency information and improve the quality of reconstruction. The texture of the reconstructed results is more real, the brightness is more accurate and more in line with the evaluation of human visual senses, which shows the effectiveness and superiority of the algorithm. It not only performs faster in reconstruction speed, but also improves the quality of reconstructed image effectively.

INDEX TERMS Convolutional neural network, super resolution reconstruction, generate countermeasure network.

I. INTRODUCTION

Image super-resolution reconstruction is a classical problem in the field of computer vision. Image super-resolution reconstruction aims to reconstruct high-resolution (HR) images with rich details by inputting one or more low resolution (LR) images. Therefore, image super-resolution reconstruction technology is widely used in medical images, satellite remote sensing, video surveillance and other fields. However, for any low resolution image, it corresponds to countless high resolution images, so the problem of image super-resolution reconstruction is an ill posed problem. In order to solve this problem, people have proposed interpolation based methods [1], [2], reconstruction based methods [3], [4] and learning based methods [5], [6], [7], [8], [9], [10]. Like. High resolution images have higher pixel density, can provide more details such as hue, shape and texture, and bring better visual experience. According to different input information, the existing SR can be divided into two categories: reconstructing a high-resolution image from multiple low-resolution images (multi frame image super-resolution) and reconstructing a high-resolution image from a single low-resolution image

The associate editor coordinating the review of this manuscript and approving it for publication was Oğuzhan Urhan.

(single frame image super-resolution). Single image super resolution (SISR) uses a single image for super-resolution reconstruction, which overcomes the problems of difficult image sequence and insufficient timing [11]. Due to the lack of correlation information between multi frame images, it is difficult to obtain the prior information of image degradation, which has become the difficulty of image super-resolution reconstruction. With the increasing demand for video and display quality, the image resolution of video, as an important feature of video quality, has gradually transitioned to the 4K level. In order to meet the higher and higher image resolution, the resolution of terminal display equipment has gradually increased [12]. However, due to the low resolution and low resolution (LR) of old video sources Video source can not get better display effect on high resolution (HR) display equipment. Therefore, super resolution technology of video source conversion from LR to HR has been widely concerned. In the learning based method, the idea of completing the reconstruction task by training the neural network has attracted extensive attention with the proposal of SRNN. SRNN stacks three convolution layers to learn the nonlinear mapping between high and low resolution image pairs, so as to complete the super-resolution task. However, due to the small number of network layers of SRNN, only shallow information can be extracted, and the reconstructed image is blurred. Based on SRNN, Xun et al. extended the network depth to 20 layers, extracted deep features [13], and took the high and low resolution image residuals as the learning goal to reduce the amount of model operation. Yue et al. proposed that DRCN (deep recursive revolutionary network) uses recursive network to process data circularly, and adds jump connection in the network layer to alleviate the phenomenon of gradient disappearance and explosion, and make full use of low resolution information [14]. Jieqing et al. used dense blocks to alleviate the problems such as network gradient disappearance, support feature reuse, reduce parameters and reduce the difficulty of training [15]. These methods proved the effectiveness of deepening the number of network layers, learning residual images and inter layer jump connections in reconstructing various types of images, including face reconstruction tasks, but because faces are a class. For objects with regular structure, the best effect can not be obtained by using ordinary super-resolution algorithm. For face images, Wen et al. proposed UR-DGN introduced the generation countermeasure network into the face super-resolution process for the first time, and used the approximately aligned frontal face image training model to solve the problem of artifacts in the reconstruction results due to the low input resolution and difficulty in locating facial features [16]. However, due to the single picture category in the data set, the side and other poses are reconstructed There is a large error in the face image. Chi et al. reconstructed the face image with clear contour by alternately optimizing the face reconstruction task and face intensive field estimation [17]. However, due to the non end-to-end network structure, the learning process is more complex. Du et al. the input image is divided into five different regions according to the facial features, the reconstructed image is generated through the corresponding convolution neural network, and the region enhancement method is used to generate a complete reconstructed image with rich facial features details [18]. However, due to the reason of stitching, the reconstruction effect is uneven and will show obvious unevenness continuous region. Hua et al. Proposed GFRN, which takes the image to be reconstructed belonging to the same person and another high-resolution guide image as common input, estimates the dense flow field to align the guide image with the image to be reconstructed, and uses the aligned image to complete the reconstruction [19]. The reconstruction results can well retain the character features, but this method requires the high-definition image of the same person to assist the reconstruction The condition is difficult to meet in practical application.

With the proposal of generative adversarial networks (GAN), people began to try to use the generated adversary network to process various computer vision tasks [20]. Zhou et al. first used the generated adversary network for image super-resolution reconstruction (SRGAN), which improves the visual effect of the generated image [21]. Cong and Yaxin removed all BN layers in the generator and proposed to use residual dense blocks instead of the original basic blocks to train a very deep network, which further improves the visual quality [22]. Inspired by this, this paper uses support vector machine (SVM) based on SRGAN model Taking the hinge loss in as the objective function, a more stable and noise resistant charbonier loss function is used in the generation network to replace the L2 loss, which improves the problem that the reconstructed image will produce speckle artifacts in the ordinary region. At the same time, the residual block and the BN layer that will normalize the features in the discriminator are removed, and spectral normalization is used in the generator and discriminator to reduce the computational overhead and stabilize the model training [23]. In the selection of activation function, the exponential linear unit ELU activation function with less complexity and better robustness to noise is used in the discriminator to replace the leaky relu activation function. The above methods further improve the visual effect of the image and make it closer to human visual perception.

In order to obtain better super-resolution reconstruction effect, this paper designs PEGAN super-resolution reconstruction method, which can reconstruct four times of down sampled images. PEGAN optimizes the activation function, basic network structure and loss function based on SRGAN algorithm framework and to improve the super-resolution reconstruction algorithm. For face images: end-to-end based on depth residual module A double-layer neural network is constructed to extract the high-frequency features of the image layer by layer to complete the reconstruction; for the unique structure of the face, a face information estimation module is added to the reconstruction network to assist the

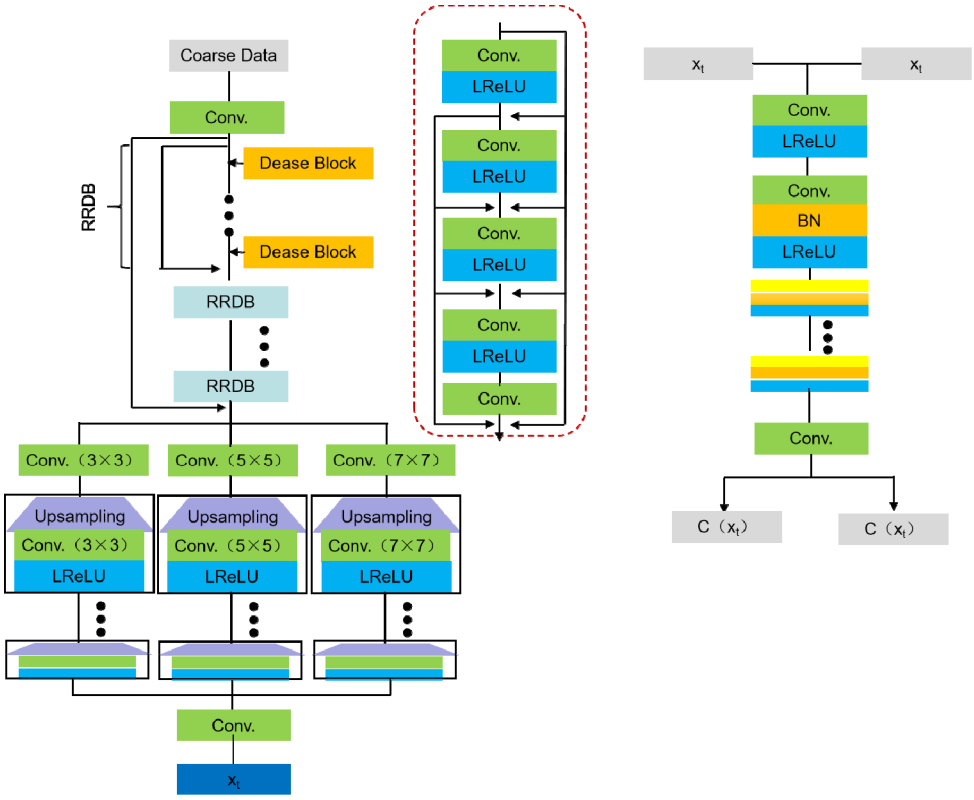

FIGURE 1. Structure diagram of our method.

reconstruction, constrain the reconstruction results from the geometric level, and effectively deal with the reconstruction task of multi pose face; the mixed amplification factor is used to train the model and improve the generalization ability of the model. This paper proposes a super-resolution reconstruction method based on multi-stage feature fusion network, which is mainly used to solve the shortcomings of many classical reconstruction network structures, that is, to improve the reconstruction effect by changing the network structure. On the one hand, this model draws lessons from the connection mode of dense network to strengthen the connection between the network layer and the multi-path connection through the whole network, so as to make full use of the characteristics of hierarchical network, extract more high-frequency information and improve the reconstruction quality.

II. RELATED WORK

A. GENERATE COUNTERMEASURE NETWORK

Inspired by the two person zero sum game, in 2019, Ian goodflow proposed the concept of generating confrontation network, which consists of a special confrontation process in which two neural networks compete with each other. The specific structure is shown in Figure 1. Among them, we are deeply inspired by the representative GAN model [24]. The generating network $G(z)$ captures the potential distribution of real sample data, generates new sample data, and discriminates the network $D(x)$ It is a two classifier that attempts to distinguish between real data and false data created by the generation network [25], that is, to judge whether the input is real data or generated samples. The discrimination network will generate a scalar in the range, representing the probability that the data is real data. The training of generating countermeasure network is actually a minimax game process, and the optimization goal is to achieve Nash equilibrium [26], that is, there is no difference between the sample generated by the generator and the real sample, and the discriminator cannot accurately judge the generated data and the real data. At this time, the probability that the discriminator thinks that the result output by the generator is the real data is 0.5. The confrontation process between the generator and the discriminator is shown in formula (1):

$$\min_G \max_D V(D, G) = E_{x\sim P_{data(x)}}[\log D(x)] + E_{z\sim Pz(z)}[\log(1 - D(g(z)))] \quad (1)$$

where, V (D, g) represents the objective function of Gan optimization, X represents the data obtained from the real data, D (x) represents the probability that the real data is judged to be true through the discriminator, Z is the random noise signal, and the generator g (z) receives the input z from the probability distribution P (z) and inputs it to the discriminator network $D(x)$. The discriminator part is shown

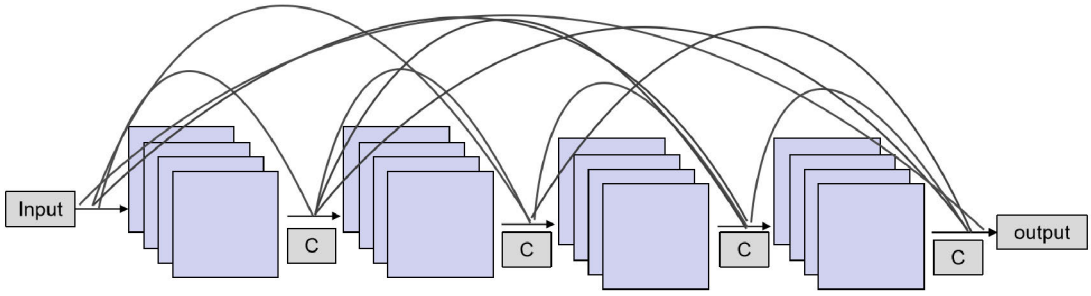

FIGURE 2. DenseNet's dense connection mechanism. (Where c stands for channel-level connection operations.)

in formula (2):

$$\max_D V(D,G) = E_{x\sim P_{data(x)}}[\log D(x)] + E_{z\sim Pz(z)}[\log(1-D(g(z)))] \quad (2)$$

Among them, for the fixed generator, for the real sample x, that is, the former term of the formula, the larger the desired result $D(x)$, the better, because the closer the result of the real sample is to 1, the better. For false samples, the smaller the desired result (DCGCZ), that is, the larger the result of 1, the better, because its label is the o} generator part, as shown in formula ((3):

$$\min_G V(D,G) = E_{z\sim Pz(z)}[\log(1-D(g(z)))] \quad (3)$$

In the optimization process, there is no real sample. We hope that the larger the result with the label $D(G)$, the better, that is, the smaller the result of $g(z)$.

The combined generation countermeasure network combines two Gan together. Each Gan is for an image domain. In essence, the image distribution $P\ (x)$ learned by the Gan should be close to the training sample distribution $P(x)$. Then any input noise into the trained generator can generate an image that is enough like the training sample. In the traditional domain adaptation, we need to learn or ill practice a domain adapter, and the domain adapter needs to use the source domain and the corresponding target domain Therefore, by adding unsupervised weight sharing constraints to the network and solving the inner product distribution solution of the boundary distribution, COGAN can realize unsupervised learning a joint distribution when there are no corresponding images in the two domains, such as learning the joint distribution of two different attributes of color and depth of the picture.

B. DENSE CONVOLUTION NETWORK

In deep learning networks, the problems of gradient disappearance and gradient dispersion will become more and more serious with the increase of the number of network layers [27]. Resnetworks proposed in highway networks proposed improved networks for the above problems [28]. Although the above algorithms are different in network structure and training process, their key point is to create a short path from the early feature layer to the later feature layer. In order to ensure the maximum transmission of information flow between different layers, Li et al. Put forward dense convolutional networks [29]. The basic idea of the model is that each layer in the network should obtain additional feature input from all its feedforward layers, and transmit its own feature map to all subsequent layers for effective training. Assuming that the model has layer B, there will be connections in densenet instead of the B connection of traditional neural network. Densenet creates a deeper and more effective convolution network, and its dense connection mechanism is shown in Figure 2.

In view of the problems existing in WGAN, researchers at the University of Montreal put forward new improvements on the basis of WGAN to further improve the objective function of the model and punish the gradient of the discriminator to the input.

The performance of the method is better than that of the standard WGAN, can stably train the Gan architecture for image generation and language model, hardly need the adjustment of super parameters, and can generate samples with higher quality than the standard WGAN at a faster convergence speed. WGAN-GP does not essentially change the structure of Gan model, but improves the description of objective function and the selection of optimization method.

C. LOSS FUNCTION

1) DISCRIMINATOR LOSS FUNCTION

Jolicoeur proposed that the probability that the pseudo data is real data should not only be improved, but also the probability that the actual data is real data should be reduced [30]. The loss function of the discriminator is shown in equation (7):

$$\begin{aligned} L_D^{Ra} &= -E_{x_r}[\lg(D_{Ra}(x_r,x_f))] - E_{x_f}[\lg(D_{Ra}(x_f,x_r))] \\ D_{Ra}(x_r,x_f) &= \sigma(C(x_r) - E_{x_f}[C(x_f)]) \\ D_{Ra}(x_f,x_r) &= \sigma(C(x_f) - E_{xr}[C(x_r)]) \end{aligned} \quad (4)$$

where ex and ex represent the average in the real high-resolution batch and the average in the generated high-resolution batch respectively, and 6 is the sigmoid function:

2) GENERATOR MODEL LOSS FUNCTION

Haoran et al. proposed vector I loss and extended it in SRGAN [31]. The loss of perception was previously defined in the activation layer of the pre trained deep network, that is, the distance between the two activation functions was minimized. In addition, PESRGAN introduces the penetration index PI, that is, the variance ratio of the original image and the Laplace operator of the image [27], [32], [33], [34]. The larger the penetration index, the higher the definition of the representative image and the more details. Using the penetration index as the weight is conducive to the generator model to generate clearer images. However, if the PI value is too large, too many pseudo details will be generated. PI formula is shown in formula (8):

$$PI = \frac{\sum_{i=1}^{N_a} (G(x_{i,}y_i) - \bar{G}(x_i, y_i))^2}{\sum_{j=1}^{N_p} (G(x_{j,}y_j) - \bar{G}(x_j, y_j))^2} \tag{5}$$

$$\begin{aligned} G(x, y) = & f(x, y) + c(f(x + 1, y) + f(x - 1, y)f(x, y + 1) \\ & + f(f(x, y - 1) - 4f(x, y)) \end{aligned} \tag{6}$$

PeSRGAN redefines the loss function in combination with the relative discriminator theory. The reason for using the features before the activation layer is that the activated features are very sparse [35], [36], [37]. For example, the percentage of activated neurons in baboon is only 11.17%. after VGG19-54 layer, indicating that the weak supervision provided by sparse activation of advanced features will degrade the performance [38]. Using the activated features will also make the brightness of the image different from the real image; The advantage of fusing two-layer features is that the two-layer features can better clarify the direction of convergence with low loss. The loss function of generator model is defined as equation (5):

$$\begin{aligned} L_G &= PI * (0.5 * L_{conv3-4} + L_{conv5-4}) + \lambda L_G^{Ra} + \eta L_1 \\ L_1 &= IE_{x_h} ||G(x_i) - y||_1 \\ L_G^{Ra} &= -E_{x_r}[1 - \lg(D_{Ra}(x_r, x_f))] - E_{x_f}[\lg(D_{Ra}(x_{f,}x_r))] \end{aligned} \tag{7}$$

where L. Is the total generation loss of the generator, l is the 1-norm distance content loss between the evaluated image g (x;) and the real image y, brother and knife are the coefficients to balance different loss terms, lrag is the loss relative to the generator, and P1 is the permeability index parameter.

3) DISCRIMINANT NETWORK

The discrimination network of PESRGAN algorithm is based on the classical VGG19 network, which can be simplified into two modules: feature extraction and linear classification. The feature extraction module includes 16 convolution layers. After convolution of each layer, leakyrelu is used as the activation function. In order to avoid gradient disappearance and enhance the stability of the model, BN layer is used after convolution of each layer except the first layer. The discriminant network needs to judge the authenticity of the input sample image, that is, the problem of image classification. In order to avoid the slow training speed of network model and the risk of over fitting, this paper uses global average pooling (GAP) to replace the full connection layer used in most image classification models, calculates the pixel average value of each layer of feature map finally obtained by the feature extraction module, and then linearly fuses all the values into the sigmoid activation function, Finally, the network output discriminator classifies the input sample image. The training of discriminant network is helpful to generate the network and reconstruct the results closer to the real high-resolution image.

III. METHOD

A. RECONSTRUCTION ALGORITHM BASED ON ANCHORED NEIGHBORHOOD REGRESSION

In the SCDL algorithm, after learning the high-resolution and low-resolution dictionary pairs and the mapping matrix W, the sparse representation coefficient can be calculated according to the input image, and then the image super-resolution reconstruction can be realized by the linear combination of the high-resolution dictionary DH and the corresponding sparse representation coefficient ah. In the reconstruction process, the objective function for calculating the sparse representation coefficient is:

$$\begin{aligned} \min_{(\alpha_h, \alpha_l)} & ||x - D_h\alpha_h||_F^2 + ||Y - D_l\alpha_l||_F^2 \\ & + \gamma||\alpha_h - W\alpha_l||_F^2 + \lambda_h||\alpha_h||_1 + \lambda_1||\alpha_l||_1 \end{aligned} \tag{8}$$

Obviously, equation (8) must continuously update the high-resolution and low-resolution sparse representation coefficients, which is an iterative process and takes a long time, so the reconstruction speed of SCDL algorithm is very slow. In order to improve the image super-resolution reconstruction speed, the reconstruction algorithm based on anchored neighborhood regression (ANR) is adopted, and the low-resolution sparse representation coefficient "is solved by ridge regression":

$$\min_{\alpha_1} ||y - D_l\alpha_l||_2^2 + \lambda||\alpha_l||_2 \tag{9}$$

where: is the regularization parameter; Y is the input low resolution image block feature. The feature block [' } is obtained by sampling the first-order and second-order operators in the horizontal and vertical directions, and then the dimension of the feature is reduced by principal component analysis. Finally, the numerical solution of "is":

$$\alpha_l = (D_l^T D_l + \lambda I)^{-1} D_l^T y \tag{10}$$

It can be seen that as long as the low resolution dictionary is obtained by off-line learning, the sparse coefficients can be calculated according to the input low resolution image block characteristics, and then the image super-resolution reconstruction is carried out by equation (13), which improves the reconstruction speed. However, in the anr algorithm,

TABLE 1. Comparison of loss function.

Method	Network input	Residual learning	loss function
SRNN	LR+bicubic	NO	L2
FSRNN	LR	NO	L2
ESPCN	LR	NO	L2
VDSR	LR+bicubic	YES	L2
EDSR	LR	YES	L1
OURS	LR+bicubic	YES	Charbonnier

the sparse representation coefficients of high and low resolution are regarded as exactly the same, Considering the difference between the two sparse representation coefficients, this paper improves the anr algorithm and uses the mapping matrix w obtained in the previous section to calculate the high-resolution sparse representation coefficient AR, so as to improve the accuracy of sparse representation and the quality of reconstructed image:

$$\alpha_h = W\alpha_1 = W(D_l^T D_l + \lambda I)^{-1} D_l^T y \tag{11}$$

The high-resolution image block can be obtained by linearly combining the result obtained by equation (14) with the high-resolution dictionary:

$$X = D_h\alpha_h = D_h W(D_l^T D_l + \lambda I)^{-1} D_l^T y = Py \tag{12}$$

It is called projection matrix Since the projection matrix in equation (15) is obtained based on all the features of the whole training set without using the features of the input image, this method can only obtain a global projection matrix, resulting in the imprecision of the projection matrix [39] Therefore, considering the characteristics of the input image, for the trained dictionary, multiply the input image block features with each dictionary atom, and the atom with the largest product is regarded as the most relevant and then select the k nearest neighbors of atom J in the high-resolution and low-resolution dictionaries respectively.

$$P_j = N_{h,j} W(N_{l,j}^T N_{l,j} + \lambda I)^{-1} N_{l,j}^T \tag{13}$$

P_j is used for reconstruction to obtain high-resolution image block; Then, the high-resolution image block x is combined and the overlapping area of the image block is averaged to obtain the high-frequency part of the high-resolution image; Finally, it is added with the interpolated image to obtain the final reconstructed high-resolution image.

B. PERCEIVED LOSS

Different from SRGAN, this paper adds charbonier loss function and TV regularity term to the perceived loss.

In order to ensure the correctness of the low-frequency part of the image constructed by the generator, charbonier loss is introduced. As shown in Table 1, most of the SR methods based on convolutional neural network optimize the network with L2 loss. L2 loss can directly optimize the P SNR value, but it will inevitably produce fuzzy prediction. In contrast, this paper uses a robust charbonier loss function instead of L2 loss to optimize the depth network to deal with outliers and improve the reconstruction accuracy. It not only has less training time, but also has a higher PSNR value of the reconstruction result. The mathematical expression of Charbonnier loss function is shown in formula (17):

$$L_{charbonnier}(y, \hat{y}) = \alpha E_{x,y\sim P_{data(x,y)}} \rho(y - G(x)) \tag{14}$$

where y is the super parameter, Y represents the HR image, $G_{(x)}$ represents the SR image, $P_{(y)}$ is the penalty term of the charbonier loss function, where the value is $\sigma = 10^{-8}$.

C. DOUBLE LAYER CASCADE CONVOLUTIONAL NEURAL NETWORK

In this paper, a super-resolution reconstruction method based on double-layer cascade neural network is proposed, which consists of two parts: a priori recovery network and structure constraint network. The network flow is shown in Figure 3. Given that the low resolution input image is LR, after bicubic interpolation and amplification, it is the same size as the corresponding high resolution image (128 × 128). The input image LR first obtains the preliminarily reconstructed image through the first layer a priori recovery network:

$$I^{MID} = G_1(I^{LR}) \tag{15}$$

Then, the output I ∼ of the first layer obtains the final reconstructed image through the structure constraint network of the second layer:

$$I^{SR} = G_2(I^{MID}) = G(I^{LR}) \tag{16}$$

In the training stage, the reconstructed image ISR and the real high-resolution image IHR are used as the input of the discriminator. The discriminator randomly selects an image to distinguish its authenticity and outputs the discrimination result.

A priori restoration network: in order to extract deep features from low resolution input, reconstruct high-frequency images with rich information to avoid gradient disappearance caused by too deep network. In this paper, the basic structure of a priori recovery network is designed by residual block. The residual block contains two blocks with a size of 3 × 3, each layer is connected with a batch processing layer, and prelu is used as the activation function [40]. Because it is difficult to extract accurate face structure information from the input image when the resolution of the input image is too low, the network first extracts the deep a priori features of the image from the input low resolution face image. The networking structure is shown in figure 4. The low resolution input first passes through a size of 3 × Convolution kernel of 3. In order to reduce the operation, set the moving step to 2 and the size of the feature map to half of the input. Then, the 20 layer residual block extracts the features of the image, enlarges the image from the deconvolution layer to the initial size, and uses the convolution layer to reconstruct the image.

Structural constraint network: by constraining the MSE loss and perceptual loss of the first layer network, a good

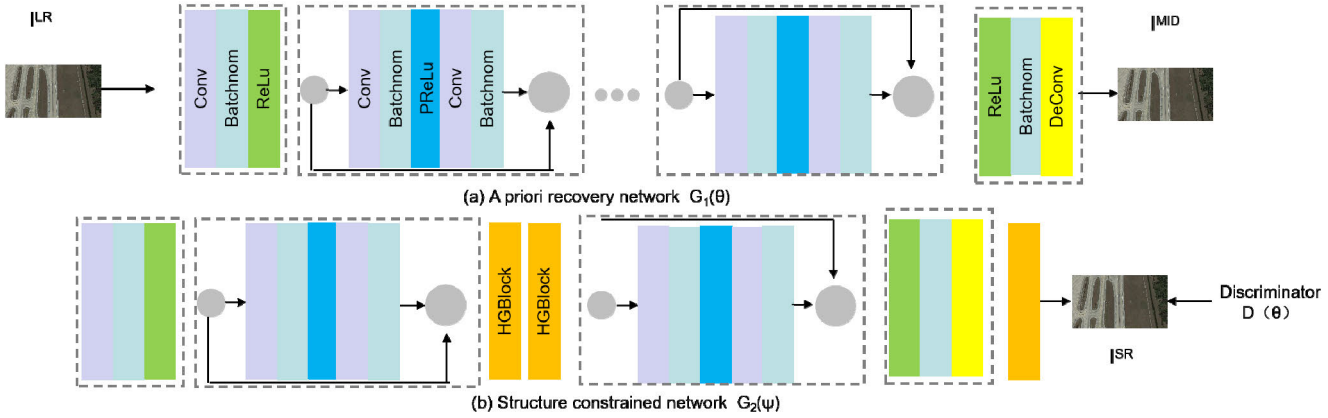

FIGURE 3. Structure diagram of double-layer cascade neural network.

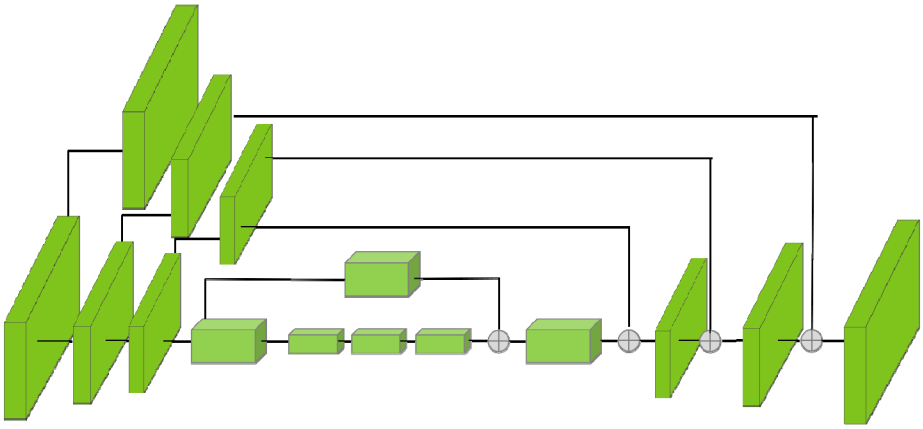

FIGURE 4. Structure diagram of hourglass module.

reconstructed image with rich high-frequency details can be obtained. However, since the facial structure information of the face is not added to the facial reconstruction process, there may be obvious errors in the facial features details of the reconstructed image when reconstructing some images with rich expressions or diverse postures.

After decoding, the two-layer deconvolution layer restores the features to the initial size, and finally the reconstruction is completed by a convolution layer. Through relay supervision of the key point heat map, the facial consistency between the reconstructed image and the real image is strengthened. In addition, during the reconstruction process, the Hg module can also provide more high-frequency information [41], which helps to improve the reconstruction effect. According to the characteristics of the diversity of the scale of the optical remote sensing images of the adaptive multi-scale super-resolution reconstruction, to extract the multi-scale information of optical remote sensing image, reduce some information is missing, but the reconstruction model without considering the visual task the specific needs of target detection, the image is reconstructed and cannot be tested for good, Moreover, both the target detection task and the hyperpartite reconstruction model are optimized independently. We know that the target detection result of optical remote sensing image largely depends on the image clarity and enough texture information to extract specific feature information.

D. MULTISCALE CONCATENATION FRAMEWORK AND HIERARCHICAL SEARCH MATCHING ALGORITHM

The image texture is complex and diverse, and is affected by noise. Wavelet analysis has the characteristics of high and low frequency signal separation and processing. It can filter the image noise and reconstruct the original signal without losing the original important information. In addition, wavelet analysis can analyze the multi-resolution characteristics of the original signal, reasonably allocate the calculation cost, and take into account the differences of signal characteristics at different scales. According to the document ‘‘20’’, the effect of multiple super division reconstruction is better than that of single super division reconstruction. Therefore, a multi-scale series framework is proposed in this paper, as shown in Fig. 5.

According to the multi-scale series framework, the image is decomposed by wavelet three-layer decomposition. Considering that the deeper the wavelet decomposition level is, the

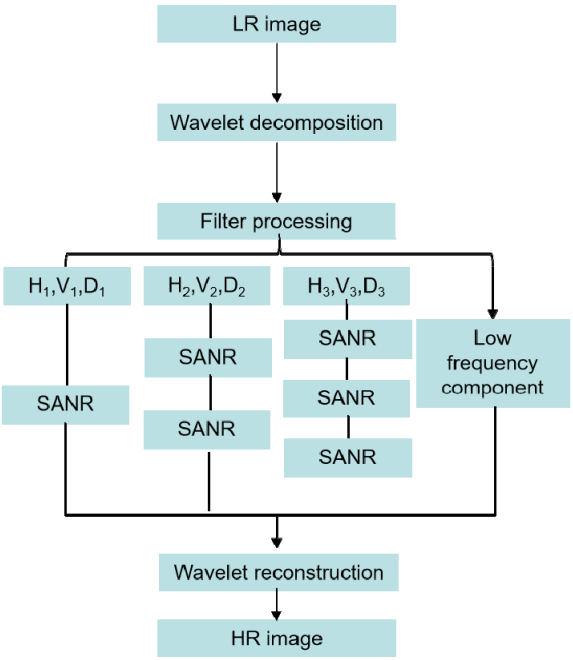

FIGURE 5. Multi-scale series framework.

smaller the wavelet component dimension is, for each layer of components, Sanr methods of different lengths are used in series according to the signal scale for multiple reconstruction, and the reconstruction results of the front Sanr are used as the input image of the rear Sanr. Figure 5, where h; And D are the signal components obtained from the horizontal, vertical and diagonal directions of the i-th layer wavelet decomposition. The series length of Sanr method is I to form the i-layer Sanr. Finally, the reconstructed high-frequency component and the original low-frequency component are reconstructed by wavelet.

IV. EXPERIMENT

A. EXPERIMENTAL SETUP

This section describes the basic setup of the experiment, explores the influence of different components used in this paper in the simplified test, analyzes the selection of K value in the multi-path feedforward structure. Inspired by [42], this paper performs data enhancement on the training set, rotates each image of the training set by 90 °, 180 °, and 270 °, and flips it horizontally. Because the human eye is sensitive to the brightness channel, this paper only deals with the brightness channel y, and the chroma channels CB and Cr are amplified by interpolation. On the other hand, as in [43], the training set in this paper contains image blocks of different sizes ($\times 2$, $\times$ 3 and $\times$ 4), only a single model needs to be trained for super-resolution reconstruction at different scales.

B. NETWORK EFFECTIVENESS EXPERIMENT

Considering that it is difficult to estimate the face structure directly from the low resolution image, this paper takes the a priori recovery network as the first layer to reconstruct

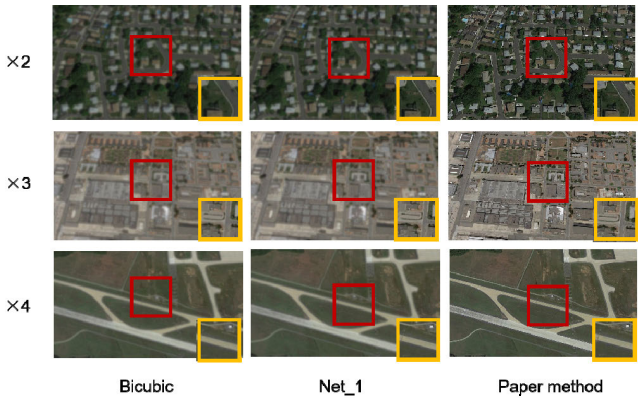

FIGURE 6. Comparison of structural constraint network in visual effect of reconstructed image.

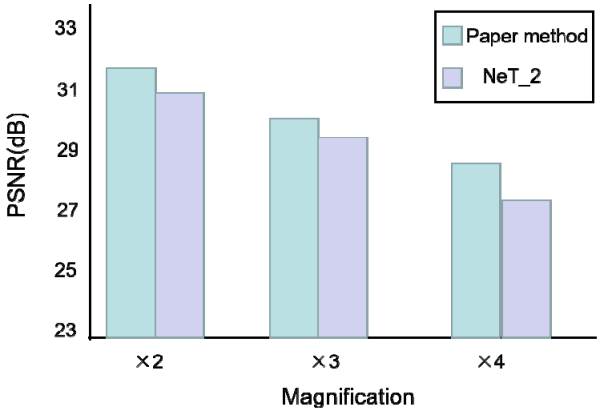

FIGURE 7. Comparison of network layer sequence on PSNR value of reconstruction results.

the image, and then improves the face reconstruction effect through the structure constraint network of the second layer. Therefore, in order to verify the effectiveness of structural constraint network in face reconstruction task and the rationality of the overall structure of the network, Net_1 and Net 2 are built respectively for verification. Net_1 removes the second layer structure constraint network from the original network and directly acts on the output of the first layer network. Net 2 switches the position of the two-layer network, and the rest remains unchanged. Net 1 and Net 2 are trained using the same data sets and parameters as the methods in this paper.

Figure 6 shows the reconstruction effects of net 1 and the method in this paper under 2, 3 and 4 magnification factors respectively. The representative areas in the figure are marked with red boxes and enlarged in the lower right corner. It can be seen that compared with net 1 without structural constraint network, this method has clearer facial contour, more accurate and sharp performance of facial features, and restores more high-frequency details (such as hair, jewelry and background). In addition, this paper can correctly restore the facial features, and the reconstructed images of net 1 have different degrees of errors when the magnification is high.

TABLE 2. Comparison of reconstruction effects of different algorithms in different data sets.

Data set	Amplification factor	Bicubic		SRGAN			SRWGAN			Paper method		
		PSNR(dB)	SSIM	PSNR(dB)	SSIM	Params	PSNR(dB)	SSIM	Params	PSNR(dB)	SSIM	Params
Remo-	2	29.02	0.7890	29.14	0.909		29.26	0.929		29.28	0.949	
A	3	27.56	0.7654	27.68	0.8854	1580K	27.8	0.9054	1356K	27.82	0.9254	897K
dataset	4	25.55	0.6752	25.67	0.7952		25.79	0.8152		25.81	0.8352	
Remo-	2	28.88	0.7123	29	0.8323		29.12	0.8523		29.14	0.8723	
B	3	27.90	0.7680	28.02	0.888	1784K	28.14	0.908	1452K	28.16	0.928	902K
dataset	4	24.44	0.6354	24.56	0.7554		24.68	0.7754		24.7	0.7954	

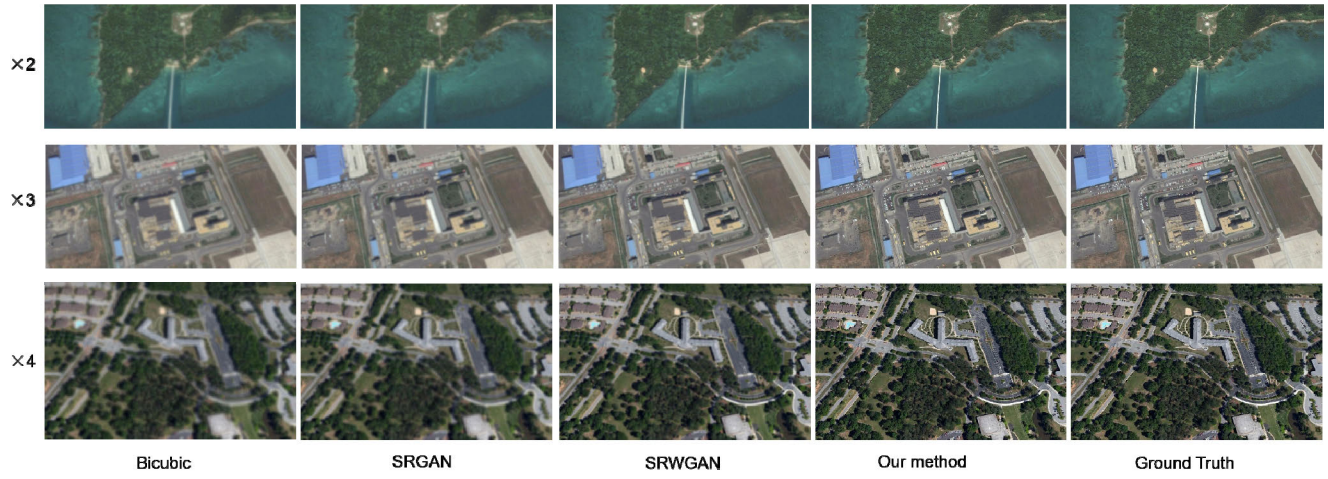

FIGURE 8. Comparison of Reconstruction Effects of different methods under multiple magnification factors on Remo-A dataset.

Figure 7 compares the peak signal to noise ratio (PSNR) of net 2 and the method in this paper under the celeba test set [37]. It can be seen from the figure that under various magnification, the method in this paper has achieved better P SNR value. The higher the magnification, the greater the PSNR gap between this paper and net 2. Therefore, it is necessary to use a priori recovery network to initially reconstruct the input image, and then optimize it to achieve better results.

C. COMPARATIVE EXPERIMENT

In this paper, the low resolution face image amplified by bicubic is reconstructed by using 2-4 magnification factor, and the reconstruction results are compared with bicubic method, SRGAN method and WGAN method. The P SNR value and image structure similarity index (SSIM) often used in image processing tasks are used for quantitative comparison. The results are shown in Table 2. It can be seen from the table that when the magnification factor of Remo-A dataset is 2, 3 and 4 times respectively, the PSNR value in this paper is increased by 1.90db, 2.34db and 2.58db respectively compared with other methods, and the SSIM value is increased by 0.0441, 0.0677 and 0.08270 respectively. It can be seen that although the PSNR and SSIM values in this paper are reduced to varying degrees with the increase of magnification, compared with other methods, This paper can recover more details at high magnification. On the Remo-B dataset, the average PSNR value increased by 1.24db and SSIM value increased by 0.040. Because celeba is selected as the training set, the numerical results of Helen test set improve less. Compared with srgan and srwgan, the model of this method has more lightweight characteristics and spends less training time.

The following compares the above methods with the reconstruction effect of this paper from the visual effect. Figure 8 compares the Remo-A dataset with 2, 3 and 4 magnification factors respectively. It can be seen that under different magnification factors, this paper can reconstruct clear facial contour. Compared with other algorithms, it can also restore facial features closer to the original image and more exquisite texture details. Although the sharpness of the reconstructed image decreases gradually with the increase of the magnification factor. However, the higher the magnification, the greater the reconstruction effect, which is consistent with the quantitative index. Figure 6 compares the Remo-B dataset with a 3 magnification factor. Clear reconstructed images can still be obtained.

D. RESULTS AND ANALYSIS

PESRGAN experiments were conducted on the host R with Intel coretmi7-6800k @ 3.400ghz CPU, NVIDIA gtx1080ti

FIGURE 9. Comparison of reconstruction effects when the magnification factor is 3 on Remo-B dataset.

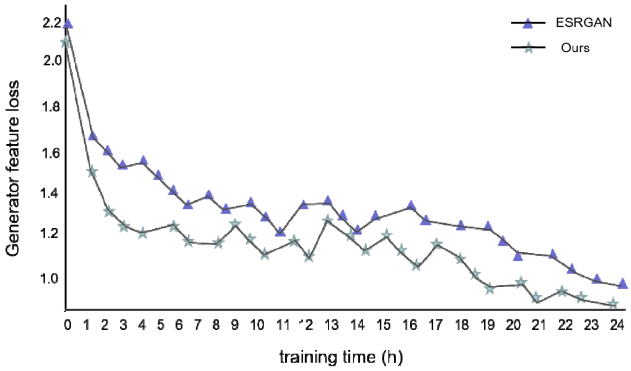

FIGURE 10. Comparison of characteristic loss.

GPU and 16GB memory. The batch size is 16, that is, 16 images are used for one gradient descent. Due to the limitation of GPU video memory used in the experiment, during the training process, the generator network input size is $48 \times 48 \times 3$, the generator network output size is $192 \times 192 \times 3$ high-resolution image through two sampling layers, the discriminator network input size is $192 \times 192 \times 3$ high-resolution image, and the input is binary classification. Finally, in the test process, the input size is mxnx3, and the output size is (4m) × (4N) × 3, That is, a high-resolution image with quadruple r sampling.

Firstly, the experiment combines leakyrelu and swish activation functions, modifies the basic network structure of generator in ESRGAN algorithm, and compares the generator loss of this experiment with the characteristic loss of ESRGAN generator, as shown in Figure 10 below (using pre training model). Compared with the representative ersgan, the model training time in this paper has obvious advantages in training time.

It can be seen from Figure 10 that the generator loss convergence speed of PESRGAN is slightly faster than that of ESRGAN. Although the swish function is more complex than leakyrelu, due to the improvement of the generator basic network RRDB, a large number of parameters are reduced, which improves the speed of the algorithm and introduces the swish activation function, making the model nonlinear, smooth and non monotonic, Finally, the loss of the improved algorithm is slightly lower than that of ESRGAN, which shows that the improvement of activation function and network structure is effective.

In order to further verify the model, the ability field of the model to reconstruct high-resolution flow is statistically checked. Figure 11 shows the probability density function of reconstructed velocity component and pressure. All reconstruction results show that the reconstructed indexes are consistent with the results obtained from the original image, which shows that the reconstruction ability of the model is effective.

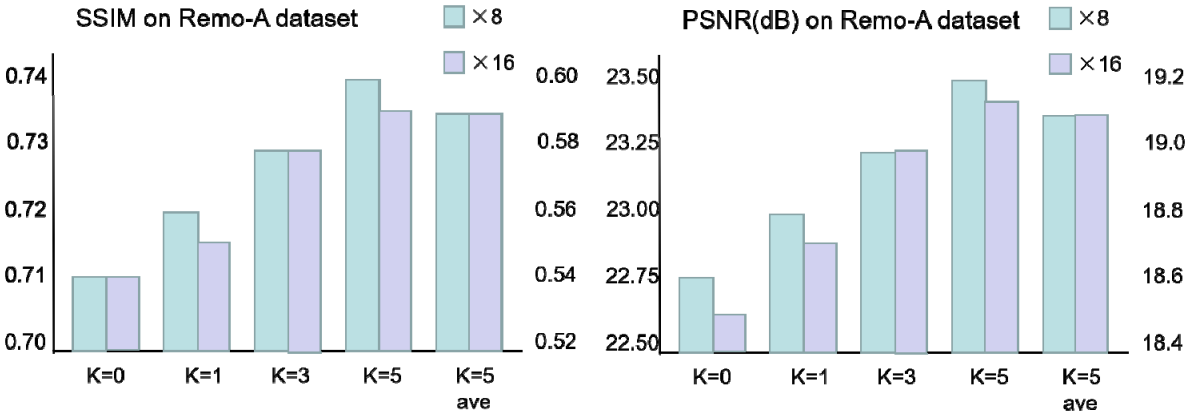

FIGURE 11. Comparison of reconstruction indexes under different magnification factors.

V. CONCLUSION

This paper proposes a super-resolution reconstruction method based on multi-stage feature fusion network, which is mainly used to solve the disadvantages of many classical reconstruction network structures, that is, to improve the reconstruction effect by changing the network structure. On the one hand, this model draws lessons from the connection mode of dense network, strengthens the connection between network layers through the multi-path connection of the whole network, so as to make full use of the layered characteristics of the network, so as to extract more high-frequency information and improve the reconstruction quality. Experimental results show that compared with other methods, this method has certain advantages in PSNR and structural similarity, and the reconstructed image has better visual effect. However, this method also has shortcomings. The existing idea of multi-stage fusion network is relatively simple. In the future work, we need to deeply study the network of feature reuse and the construction method of model. In the future, the network and model construction methods of feature reuse need to be further studied. In the follow-up work, we can adopt the idea of recursive learning to further optimize the model by reducing network parameters and increasing training samples, and explore and improve from the perspective of further improving the use and fusion of hierarchical features. Compared with other methods, the reconstructed image has better visual effect, higher PSNR and SSIM values and faster test time, which proves the effectiveness of the proposed method. Future research will focus on how to reduce training parameters, reduce network structure and obtain better reconstruction quality.

REFERENCES

[1] Z. Haiqi, L. Hong, L. Dingwen, and L. Fu, "Single image super-resolution reconstruction based on generated countermeasure network," *J. Jilin Univ.*, vol. 59, no. 6, pp. 1491–1498, 2021.
[2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
[3] W. Guoying, "Research on low resolution laser image reconstruction of visual communication technology," *Laser J.*, vol. 42, no. 10, pp. 87–92, 2021.
[4] K. Fukami, K. Fukagata, and K. Taira, "Super-resolution reconstruction of turbulent flows with machine learning," *J. Fluid Mech.*, vol. 870, pp. 106–120, Jul. 2019.
[5] C. Shengdi, "Super resolution reconstruction of medical images based on generated countermeasure network," *Comput. Age*, vol. 10, no. 5, pp. 15–19, 2021.
[6] S. Lei, Z. Hongmeng, M. Xiuqing, G. Song, and H. Yongjin, "Detection of strong compression depth forgery video based on super-resolution reconstruction," *J. Electron. Inf.*, vol. 43, no. 10, pp. 2967–2975, 2021.
[7] L. Zhang, H. Zhang, H. Shen, and P. Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Process.*, vol. 90, no. 3, pp. 848–859, Mar. 2010.
[8] Z. Haoguang, Q. Hanshi, W. Xin, S. Yang, L. Ligang, H. Songwei, M. Sen, and W. Ping, "Super resolution reconstruction of high speed micro scanning image," *Opt. Precis. Eng.*, vol. 29, no. 10, pp. 2456–2464, 2021.
[9] Y.-H. Wang, J. Qiao, J.-B. Li, P. Fu, S.-C. Chu, and J. F. Roddick, "Sparse representation-based MRI super-resolution reconstruction," *Measurement*, vol. 47, pp. 946–953, Jan. 2014.
[10] W. Yannian, L. Hangyu, L. Hongtao, and L. Yanyan, "Image super-resolution reconstruction based on wavelet depth residual network," *Foreign Electron. Meas. Technol.*, vol. 40, no. 9, pp. 160–164, 2021.
[11] H. Qiaoling, Z. Xibo, S. Runze, and Z. Yue, "Super resolution reconstruction of soil CT images based on sequence information," *J. Agricult. Eng.*, vol. 37, no. 17, pp. 90–96, 2021.
[12] Y. Chang and B. Luo, "Bidirectional convolutional LSTM neural network for remote sensing image super-resolution," *Remote Sens.*, vol. 11, no. 20, p. 2333, Oct. 2019.
[13] Z. Xun, Y. Degang, and F. Ji, "Single image super-resolution reconstruction algorithm based on multi receptive field Laplace generated countermeasure network," *J. Comput. Aided Design Graph.*, vol. 33, no. 10, pp. 1504–1513, 2021.
[14] Y. Yue, X. Xin, H. Lei, and H. Min, "Super resolution reconstruction of continued fraction interpolation combined with convolution neural network," *J. Hefei Univ. Technol.*, vol. 44, no. 8, pp. 1146–1152, 2021.
[15] T. Jieqing, C. Mengqi, Z. Xingchen, and G. Xianyu, "Super resolution image reconstruction based on local structure similarity and sparse representation," *J. Hefei Univ. Technol.*, vol. 44, no. 8, pp. 1045–1050, 2021.
[16] X. Wen, Y. Xiaomei, X. Qiuyi, T. Qiaoyu, and L. Kai, "FOTV-ADMM super-resolution image reconstruction based on L-curve parameter adjustment," *Telecommun. Technol.*, vol. 61, no. 8, pp. 1034–1042, 2021.
[17] J. Chi, Z. Sun, H. Wang, P. Lyu, X. Yu, and C. Wu, "CT image super-resolution reconstruction based on global hybrid attention," *Comput. Biol. Med.*, vol. 150, Nov. 2022, Art. no. 106112.
[18] J. Du, Z. He, L. Wang, A. Gholipour, Z. Zhou, D. Chen, and Y. Jia, "Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network," *Neurocomputing*, vol. 392, pp. 209–220, Jun. 2020.
[19] J. Hua, M. Liu, and S. Wang, "A super-resolution reconstruction method of underwater target detection image by side scan sonar," in *Proc. 2nd Int. Conf. Control, Robot. Intell. Syst.*, Aug. 2021, pp. 143–148.

[20] N. Q. Truong, P. H. Nguyen, S. H. Nam, and K. R. Park, "Deep learning-based super-resolution reconstruction and marker detection for drone landing," *IEEE Access*, vol. 7, pp. 61639–61655, 2019.
[21] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.
[22] L. Cong and W. Yaxin, "Super resolution reconstruction of remote sensing images based on dual parallel residual network," *Pattern Recognit. Artif. Intell.*, vol. 34, no. 8, pp. 760–767, 2021.
[23] L. Jing, L. Zhe, and H. Wenzhun, "Image denoising algorithm based on fast nonlocal mean and super-resolution reconstruction," *J. Ordnance Ind.*, vol. 42, no. 8, pp. 1716–1727, 2021.
[24] Q.-M. Liu, R.-S. Jia, C.-Y. Zhao, X.-Y. Liu, H.-M. Sun, and X.-L. Zhang, "Face super-resolution reconstruction based on self-attention residual network," *IEEE Access*, vol. 8, pp. 4110–4121, 2020.
[25] L. Yida and M. Xiaoxuan, "Image super-resolution reconstruction model based on RDN and WGAN," *Modern Electron. Technol.*, vol. 44, no. 16, pp. 79–84, 2021.
[26] J. Tian and K.-K. Ma, "A survey on super-resolution imaging," *Signal, Image Video Process.*, vol. 5, no. 3, pp. 329–342, 2011.
[27] W. Jue and P. Peisheng, "Research on low resolution expression recognition based on super-resolution reconstruction," *Comput. Technol. Develop.*, vol. 31, no. 7, pp. 47–51, 2021.
[28] L. Shan, L. Cao, and B. Guo, "Identification of flow units using the joint of WT and LSSVM based on FZI in a heterogeneous carbonate reservoir," *J. Petroleum Sci. Eng.*, vol. 161, pp. 219–230, Feb. 2018.
[29] Q. Li, X. Guo, J. Han, X. Zhang, and H. Liu, "Single image super-resolution reconstruction algorithm based on deep learning," in *Proc. 3rd Int. Conf. Comput. Modeling, Simul. Algorithm (CMSA)*, 2021, pp. 224–229.
[30] X. Yongbing, Y. Dong, Y. Dabing, Z. Zhiliang, Z. Zhao, and L. Qingwu, "Binocular image super-resolution reconstruction algorithm guided by multi attention mechanism," *Electron. Meas. Technol.*, vol. 44, no. 15, pp. 103–108, 2021.
[31] L. Haoran, L. Kun, C. Shilong, T. Zhaoxing, Q. Wuxia, and X. Linyan, "Pet super-resolution reconstruction method based on residual mixed domain attention network," *Electron. Meas. Technol.*, vol. 44, no. 14, pp. 103–110, 2021.
[32] G. Jie, C. Yiping, and C. Jin, "On line phase measurement profilometry based on super-resolution image reconstruction," *Acta Photonica Sinica*, vol. 50, no. 7, pp. 160–168, 2021.
[33] Y. Shuguang, "Perceptual enhanced super-resolution reconstruction model based on depth back projection," *Appl. Opt.*, vol. 42, no. 4, pp. 691–697, 2021.
[34] L. Pengwei, G. Yuan, Q. Pinle, Y. Zhe, and W. Lifang, "Generation antagonism based on multiple receptive fields," *Super Resolution Reconstruction Netw. Med. MRI Images Comput. Appl.*, no. 7, pp. 1–8, Nov. 2011.
[35] B. Mingming, Z. Yunjie, and Z. Bin, "Generation of super-resolution images using antagonistic edge learning model," *Sci., Technol. Eng.*, vol. 21, no. 19, pp. 7891–7898, 2021.
[36] A. B. Andhumoudine, X. Nie, Q. Zhou, J. Yu, O. I. Kane, L. Jin, and R. R. Djaroun, "Investigation of coal elastic properties based on digital core technology and finite element method," *Adv. Geo-Energy Res.*, vol. 5, no. 1, pp. 53–63, Mar. 2021.
[37] L. Yunfeng, Y. Jinbiao, H. Jinfeng, P. Yihao, and Z. Hongshan, "Thermal imaging super-resolution reconstruction of power equipment based on edge enhanced generation countermeasure network," *Power construction*, vol. 42, no. 7, pp. 83–89, 2021.
[38] T. Dian, "Research on lensless super-resolution imaging system based on dual line array image sensor," Xi'an Univ. Technol., Xi'an, China, Tech. Rep. 11, 2021.
[39] L. Shuaijun, "Research on low resolution image fusion algorithm based on deep learning," Xi'an Univ. Technol., Xi'an, China, Tech. Rep. 2, 2021.
[40] Y. Nianzeng, "Unsupervised image super-resolution reconstruction based on knowledge distillation," Xi'an Univ. Technol., Xi'an, China, Tech. Rep. 5, 2021.
[41] T. Geng, X.-Y. Liu, X. Wang, and G. Sun, "Deep shearlet residual learning network for single image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 4129–4142, 2021.
[42] H. Xiaolong, "Design and implementation of integrated imaging system based on sparse camera array," Xi'an Univ. Technol., Xi'an, China, Tech. Rep. 3, 2021.
[43] W. Zha, X. Li, Y. Xing, L. He, and D. Li, "Reconstruction of shale image based on Wasserstein generative adversarial networks with gradient penalty," *Adv. Geo-Energy Res.*, vol. 4, no. 1, pp. 107–114, Mar. 2020.