

RESEARCH

Open Access



# Single-frame super-resolution for remote sensing images based on improved deep recursive residual network

Jiali Tang<sup>1,2</sup>, Jie Zhang<sup>1</sup>, Dan Chen<sup>1</sup>, Najla Al-Nabhan<sup>3</sup> and Chenrong Huang<sup>4\*</sup>

\* Correspondence: [huangcr@njit.edu.cn](mailto:huangcr@njit.edu.cn)

<sup>4</sup>School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China  
Full list of author information is available at the end of the article

## Abstract

Single-frame image super-resolution (SISR) technology in remote sensing is improving fast from a performance point of view. Deep learning methods have been widely used in SISR to improve the details of rebuilt images and speed up network training. However, these supervised techniques usually tend to overfit quickly due to the models' complexity and the lack of training data. In this paper, an Improved Deep Recursive Residual Network (IDRRN) super-resolution model is proposed to decrease the difficulty of network training. The deep recursive structure is configured to control the model parameter number while increasing the network depth. At the same time, the short-path recursive connections are used to alleviate the gradient disappearance and enhance the feature propagation. Comprehensive experiments show that IDRRN has a better improvement in both quantitation and visual perception.

**Keywords:** Super-resolution, Recursive residual network, Deep learning, Remote sensing

## 1 Introduction

Remote-sensing applications mainly process and analyze remotely sensed images extracted by satellites to analyze useful information on the ground, including disaster monitoring, environmental detection, geology, and resource exploration [1]. As a key indicator for measuring satellite remote sensing performance, the spatial resolution of remote sensing images is very important in practical applications. High-resolution (HR) images are usually desired for remote sensing analysis and processing procedure. However, remote sensing images always distort due to the limitations of remote sensing image sensors and other factors like optical system aberration, atmospheric disturbance, movement, and noise of imaging system. The simplest way to improve the resolution is to increase the sensors' density of remote sensing image acquisition equipment. However, this will generate shot noise, cause a big amount of hardware costs, increase the weight and volume of the sensor, and add the difficulty of satellite launch, which is not conducive to the application and popularization of high-resolution sensors [2–4]. In this respect, SISR is a better approach. It is an image post-processing technology, which is based on digital signal processing theory and can

effectively and conveniently improve image resolution. SISR is mainly divided into two types: reconstruction-based SISR and learning-based SISR. In remote sensing applications, without increasing hardware investment, it can obtain high-resolution images of regions of interest, improve the recognition accuracy of targets of interest in images, and increase the value of image applications [5].

The reconstruction-based method mainly uses the imaging process of low-resolution (LR) images to build a model and proposes a series of constraints on the reconstructed image. The classic algorithms mainly include the iterative backward projection (IBP) [6], projection onto convex sets (POCS) [7], and Bayesian maximum a posteriori (MAP) [8], among which, the MAP method is the most widely used, usually with a regular term [9] to build a MAP solution framework. As for the total variation (TV) regular method [10], it is believed that the total variation of a noisy image is always greater than the total variation of a pure image, so the problem of suppressing noise in reconstruction is solved by constraining the total variation of the image; in general total variation (GTV) regularization [11], the distance relationship between the point of interest and the domain is further accurately described. Gradually, more reasonable and effective regularized [12, 13] image models are used for super-resolution restoration of images. Reconstruction-based SISR algorithms are insufficient in utilization of the prior information of the image itself. Most of these methods use some prior knowledge of the image's edge and local smoothness to form constraints, and then use iterative algorithms to solve the optimization problem, but when the magnification is large, the reconstructed image is often too smooth, which lacks sharpness.

The learning-based method mainly learns the mapping relationship between the LR and HR images by training on the training set in advance and uses the learned mapping relationship to restore the high-resolution image. Learning-based SISR algorithm was first developed by Freeman et al. [14] and then applied by Baker et al. [15] to reconstruct the face image. Super-resolution reconstruction based on clustering [16, 17] has achieved good results, and the method of learning based on the sparse representation [18, 19] is the most widely used; reference [20] improves image feature extraction and dimensionality reduction during dictionary training so that the reconstructed image retains more high-frequency detail information; reference [21] proposes the sparse representation of the sample database composed of the low-resolution and high-resolution sample image blocks, and the over-complete dictionary corresponding to the training image pair is used. In recent years, super-resolution restoration using deep learning has begun to appear. Reference [22] proposes that the three-layer convolution corresponds to the extraction of image blocks, feature non-linear mapping, and final reconstruction. The interpolation-enlarged LR image is input to reconstruct the image. A method of feedback residual network based on deep edge guidance is proposed in reference [23], and images are trained according to different frequency bands and routes through recursive residual network. Reference [24] puts forward the idea of using residual learning to implement image reconstruction; Reference [25] conducts a convolution operation on a low-resolution image and finally performs an upsampling operation at the end of the network, that is, an operation to improve the resolution; in reference [26], the idea of generative confrontation is introduced into super-resolution, and a confrontation network and a discrimination network are used to simulate the confrontation. The discrimination network is used to judge the predicted high-resolution image

generated by the generation network. However, these learning-based SISR techniques require sufficient HR training examples in order to perform properly and generalize well. In addition, they usually tend to overfit quickly due to the models' complexity and the lack of training data.

To overcome the problems mentioned above, we propose a novel fusion SR method named IDRRN in this paper. A recursive residual network is introduced into the super-resolution restoration of remote sensing images. In this network model, global residual learning and local residual learning are introduced to reduce the difficulty of training deep networks, and a recursive block composed of residual units is used. To learn the residual image between high-resolution and low-resolution images, we can boost the accuracy by increasing the network depth without adding any weight parameters. Without loss of image restoration quality, the deep learning model is improved to make its network structure more concise and compact. By connecting multiple secondary filters in the deep network, the accuracy is significantly improved. This model uses local residual learning instead of global residual learning to train deep networks, which is more conducive to information transmission and gradient flow. The infusion of a recursive structure in the residual block reduces the parameters and makes the model more compact. Taking the uninterpolated LR image as input, and finally using the deconvolution layer at the end of the network to directly upsample to the SR output image, the calculation complexity is greatly reduced.

The algorithm has been adapted to be efficiently executed in parallel and presents some methodological improvements to make the model more efficient and effective. Experimental results show that the proposed method performs significantly against existing methods in evaluation indicators and visual effect.

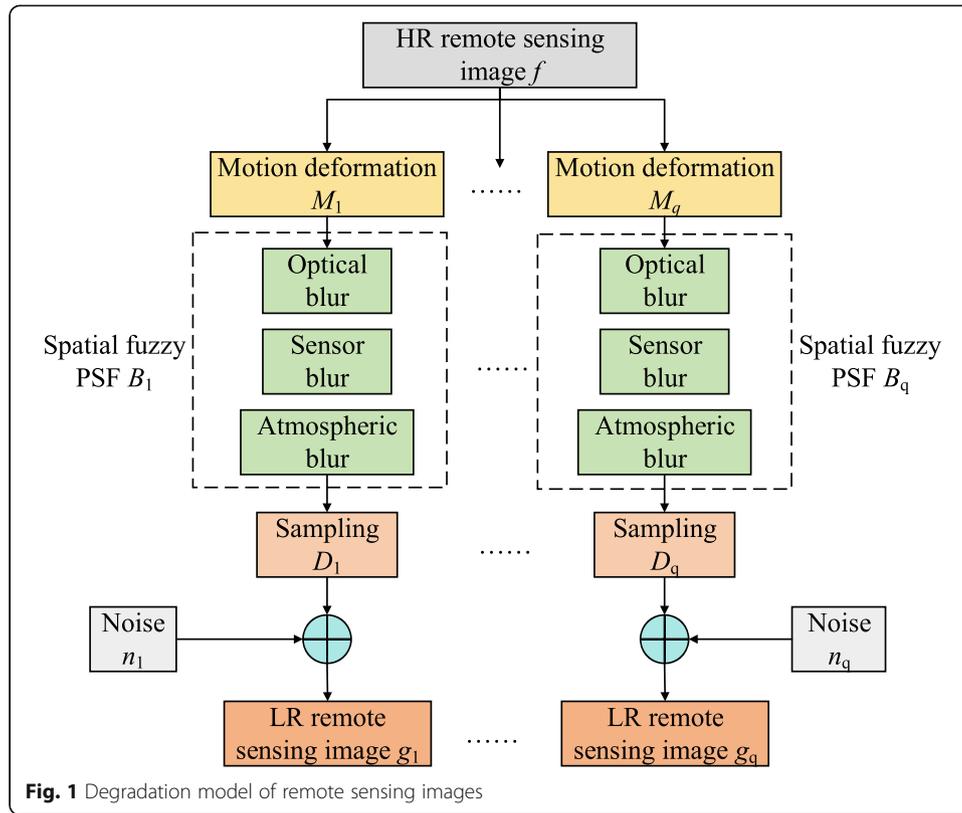
### 1.1 Related works

We briefly review the ideas and work progress related to this paper in this section. Firstly, we discuss the image degradation in remote sensing and get the mathematical model of LR images. Next, we describe the main idea of deep learning and its application in SISR algorithms. Finally, we illustrate the image restoration model of learning the residual by the convolutional neural network (CNN), in which the corruption is considered as "residual information."

### 1.2 Image degradation in remote sensing

The formation of remote sensing images has gone through several links. In these links, the problems of image degradation and quality degradation inevitably occur. In order to obtain high-quality spatial images, the acquired remote sensing images need to be denoised and deblurred [27]. As shown in Fig. 1, a degradation model is first established from the original image to the actual acquired image, where the original image is a high-resolution image and the actual acquired is a LR image.

When each image is taken by remote sensing, the blurry point spread function in different spatial domains  $B_i$  and motion deformation parameters  $M_i$  under different effects  $D_i$ , a LR image sequence can finally be obtained. After the image degradation model is established, the mathematical model of the low-resolution image can be expressed as follows:



$$g_i = D_i B_i M_i f + n_i, i = 1, 2, \dots, q \tag{1}$$

Among them,  $g_i$  is the vectorized representation of the low-resolution image  $i$ ,  $q$  is the number of LR image frames,  $f$  is a vectorized representation of a HR image,  $m$  and  $n$  represent the spatial dimensions of the real image,  $M_i$  is the motion matrix,  $B_i$  is a fuzzy matrix,  $D_i$  is the downsampling matrix, and  $n_i$  is the vectorized representation of the  $(m \times n) \times 1$  dimensional noise.

Make

$$g = \begin{bmatrix} g_1 \\ g_2 \\ \dots \\ g_p \end{bmatrix}, H = \begin{bmatrix} D_1 B_1 M_1 \\ D_2 B_2 M_2 \\ \dots \\ D_p B_p M_p \end{bmatrix}, n = \begin{bmatrix} n_1 \\ n_2 \\ \dots \\ n_p \end{bmatrix}, p = 1, 2, \dots, q \tag{2}$$

then the degradation model of  $q$  LR remote sensing images can be abbreviated as follows:

$$g = Hf + n \tag{3}$$

Among them,  $g$  is a vectorized representation of a LR image,  $H$  is the degradation matrix, and  $n$  is a vectorized representation of noise.

### 1.3 Deep learning for SISR in remote sensing

High-resolution remote sensing images play an important role in agricultural and forestry monitoring, urban planning, and military reconnaissance. As the smallest size that can be distinguished by the spatial details of the target in the image, the spatial resolution of the remote sensing image is one of the key indicators for evaluating the image quality. However,

due to the high-cost and time-consuming development of HR remote sensing satellites, how to obtain HR images economically and conveniently has always been a major challenge in the field of remote sensing. Super-resolution reconstruction technology is a favorite resort to such problems. The general objective in SR is to improve the image resolution beyond the sensor limits, that is, to increase the number of image pixels while providing finer spatial details than those captured by the original acquisition instrument.

The SISR of remote sensing images is an ill-conditioned inverse problem, so reasonable image feature expression is particularly important in the reconstruction process. Deep learning methods, especially CNN, can perform feature transformation and non-linear mapping on LR images to obtain complex feature expressions of LR images and then build LR images to HR images complex mapping relationship. The essence of deep learning is a self-learning method for data representation, replacing manually extracting features by using unsupervised or semi-supervised feature learning and hierarchical feature acquisition methods.

Super-resolution convolutional neural network (SRCNN) [22] has begun the era of deep convolutional neural networks dealing with super-resolution problems. The algorithm takes the result of LR image interpolation as the network input and obtains a HR image after three convolutional transformations. After three steps of feature extraction, nonlinear transformation, and feature restoration, a very good restoration effect is obtained. The first convolution layer is the extraction of image features. Image blocks are extracted from the LR image and each block is represented as a high-dimensional vector. Given a low-resolution image  $x$ , the process can be expressed as follows:

$$N_1(x) = \max(0, f_1x + d_1) \quad (4)$$

Among them,  $f_1$  is the convolution kernel of the first convolution layer, which can be regarded as a filter.  $d_1$  represents the bias of the first layer.

The second convolution layer is a non-linear mapping between features, mapping each high-dimensional vector to another high-dimensional vector. Each mapping vector is a conceptual representation of HR blocks, which can be expressed as follows:

$$N_2(x) = \max(0, f_2N_1(x) + d_2) \quad (5)$$

Here  $f_2$  and  $d_2$  represent the filter and bias of the second convolution layer.

The third convolution layer is a process of reconstructing an image to generate HR image. This operation stitches the above HR image blocks to generate a final HR image, which can be expressed as:

$$N_3(x) = f_3N_2(x) + d_3 \quad (6)$$

Here,  $f_3$  and  $d_3$  represent the filter and bias of the third convolution layer.

The entire convolutional neural network model continuously reduces the loss of the network through iteration. When the loss value is minimized and stabilized, the corresponding weight and bias of each layer of convolution are the optimal results of the network.

Accompanying the robust development of deep learning algorithms and great success of SRCNN, super-resolution recovery algorithm based on deep convolutional networks developed rapidly, and various improved variants and new network structures appeared accordingly, such as fast super-resolution convolutional neural network (FSRCNN) [28], very deep convolutional networks for image super-resolution (VDSR) [24], super-resolution generative adversarial network (SRGAN) [26], end-to-end deep and shallow

networks (EEDS) [29], and enhanced deep super-resolution network (EDSR) [30]. This greatly improves the practical application of deep learning for SISR.

#### 1.4 Deep residual network

Residual network (ResNet) is proposed to solve the problem of network degradation when the deep neural network has too many hidden layers. Its main idea is to learn the residual function instead of the original function based on the input, which makes the training of the deeper network simpler, and can get better performance from the deeper network [31–33]. Its network structure is shown in Fig. 2.

Reference [34] pointed out that two weight layers and an activation function ReLu are regarded as a basic unit, and then, the input and output of the unit are added at the pixel level through a jump connection, that is, the corresponding pixels in the feature map are added, and the residual operation is performed as follows:

$$H(x) = F(x) + x \quad (7)$$

Among them,  $x$  represents the input of a basic unit,  $H(x)$  represents the result of the residual calculation, and  $F(x)$  represents the basic unit calculation result.

The residual block structure is as follows:

$$x_o = U(x) = \sigma(F(x, W) + h(x)) \quad (8)$$

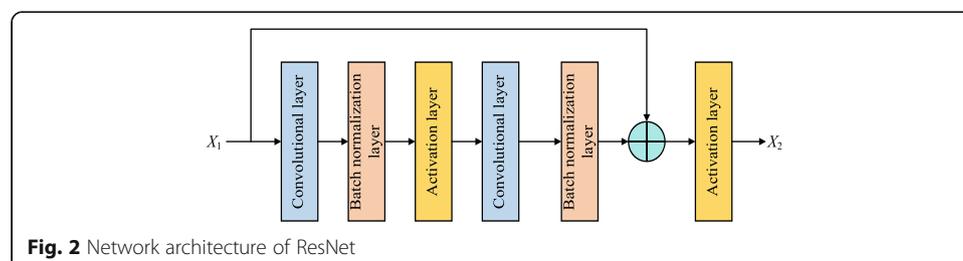
Among them,  $x_o$  represents the output of the residual block,  $h(x)$  is an identity mapping and  $h(x) = x$ ,  $W$  is a set of weights,  $F(x, W)$  is the residual mapping to be learned,  $\sigma$  represents Relu activate function, and  $U$  represents a residual block function. The residual mapping is easier to optimize than the original mapping.

The proposed residual network breaks the argument that deepening the number of layers in the network cannot improve performance. Moreover, the structure of the deep residual network is simple, which solves the problem of performance degradation of deep convolutional neural networks under extremely deep conditions, and the classification performance is excellent.

#### 1.5 Proposed improved method

##### 1.5.1 Recursive structure

Reference [35] proposed deeply-recursive convolutional network (DRCN) algorithm, which introduced recursive algorithm in residual network. The recursive structure consists of 16 chain structures. DRCN passes the recursive results through the reconstruction layer each time, generating intermediate results of HR images. DRCN's recursive structure allows weight parameters to be shared in the convolutional layer, effectively controlling model parameters. However, in order to solve the problem that the training



**Fig. 2** Network architecture of ResNet

deep model is prone to vanish or explode gradients, each recursive learning needs to be supervised, which undoubtedly increases the burden on the network.

In response to the above issues, in this paper, the improved recursive structure is introduced into the residual block to reduce the network scale and make the model more compact. At the same time, the weights are shared among the residual blocks, reducing the number of model parameters. The residual block function is defined as:

$$H^\mu = R(H^{\mu-1}) = F(H^{\mu-1}, W) + H^0 \quad (9)$$

$H^\mu$  is the  $\mu$ th output of the first residual block,  $R$  represents the residual block function,  $F(H^{\mu-1}, W)$  is the residual mapping to be learned,  $W$  is the set of weights, and  $H^0$  is the feature image output through the first convolution layer.

A convolution layer and a Relu layer are introduced at the beginning of the recursive block and then superimpose multiple residual blocks, which forms a recursive structure. Among them,  $H^0$  refers to the identity mapping of each residual block, and  $B$  represents the number of residual blocks contained in the recursive structure. The algorithmic recursive structure is shown in Fig. 3.

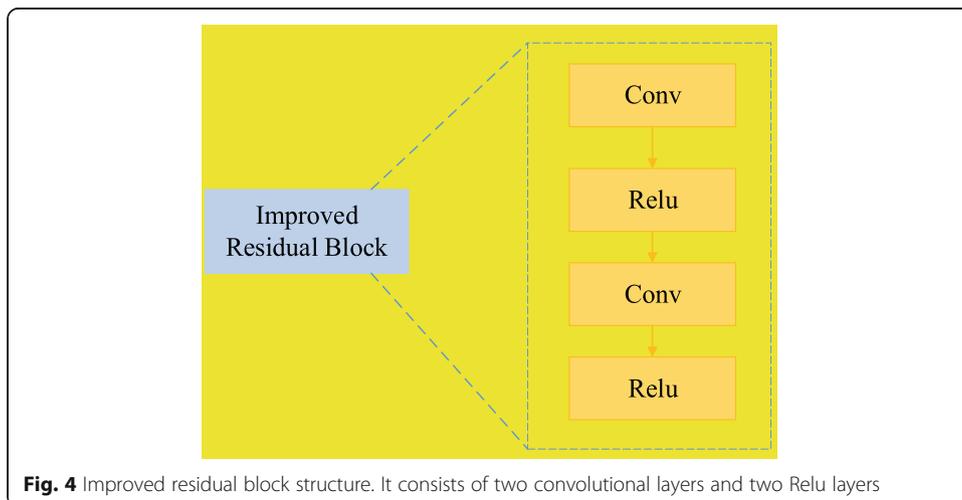
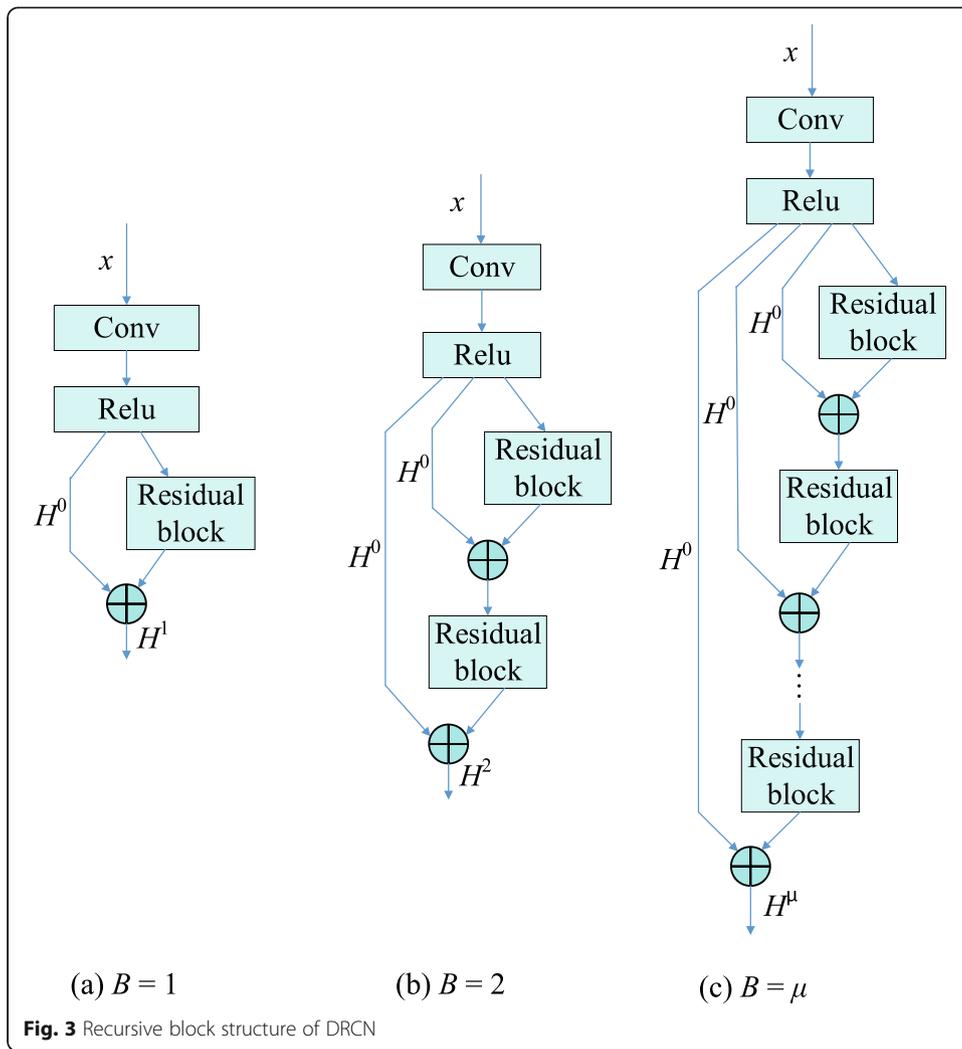
The result of the  $\mu$ th residual block can be obtained by the residual block function  $R$  recursively.

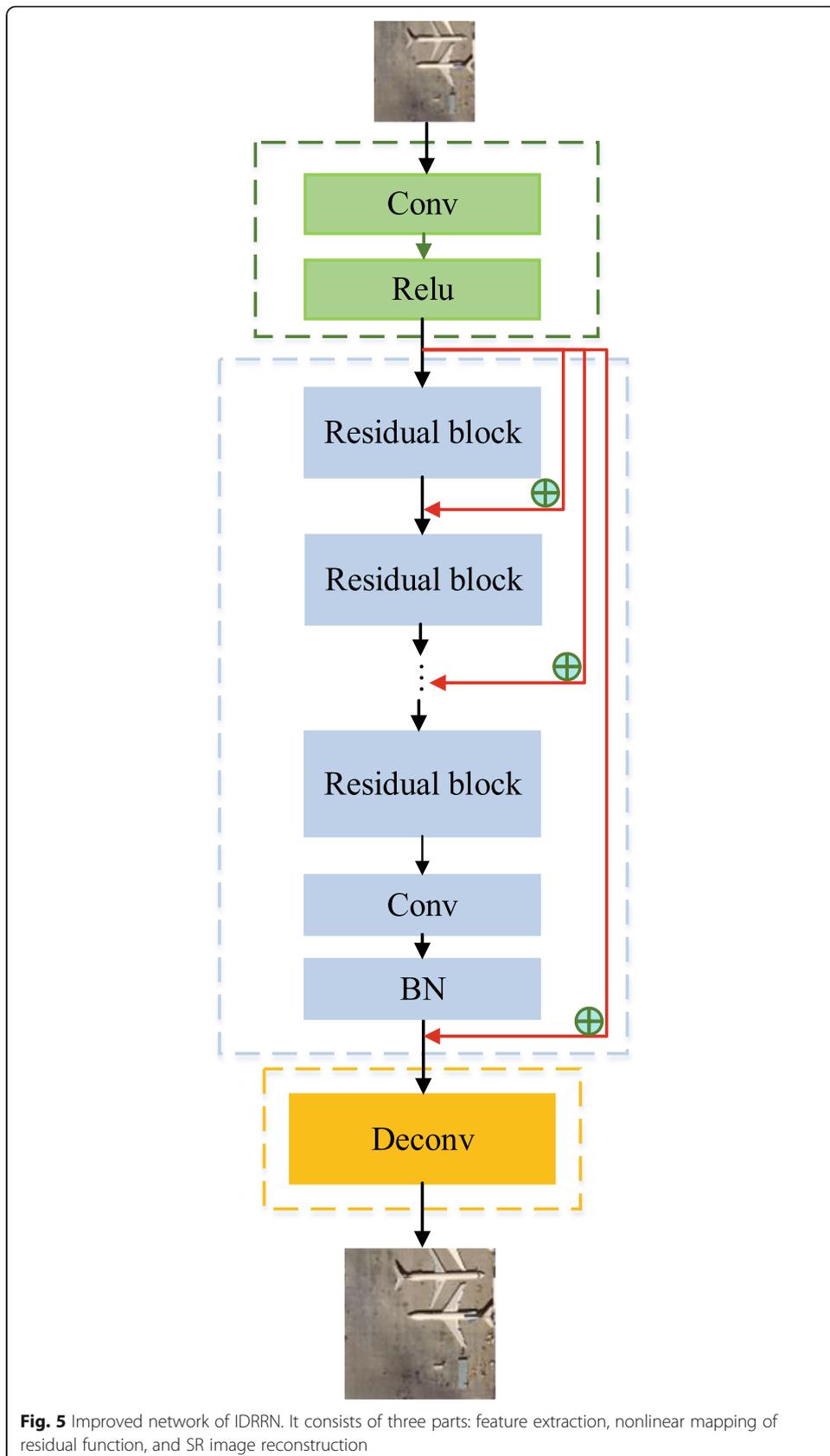
### 1.5.2 Network structure optimization

The algorithm introduces local residual learning to reduce the difficulty of training the deep network. First, the high-frequency features of the LR input image are extracted through the convolution layer, and then after each two-layer convolution layer, the feature image extracted by the first convolution layer is added. That is, the inputs of all identity branches in the residual block remain the same. In this way, more image information can be transmitted to the deeper layer of the network, and its identity branch also helps the back propagation of gradients during training, avoiding the overfitting phenomenon [36]. The improved residual block structure consists of two convolutional layers and two Relu layers. The residual block structure is shown in Fig. 4.

Recursive structure is introduced in the residual block. The parameters are reduced, which is more helpful for information transmission and gradient flow. LR input image goes through a convolutional layer and a Relu layer, extracts features, and then inputs the extracted features into several residual blocks, and recursively learn the residual mapping function. Finally, at the end of the network, a deconvolution layer is used to directly upsample the learned residual image and restore SR output image. The optimized network structure is shown in Fig. 5.

It can be seen from the figure the number of convolutional layers in each residual network unit. In the improved network, there are more layers of residual network units at the front part of the network and fewer layers of the residual network units at the later part. This design can make the entire network contain deeper network branches while using the same number of parameters, thereby improving the quality of the generated images. The deep branches of the adjusted network increase, so that the optimized network can work more efficiently. At the same time, in order to avoid gradient dispersion and overfitting in deep networks, a pooling layer is added to the branches with deeper network layers, that is, residual network units near the output end.





**Fig. 5** Improved network of IDRRN. It consists of three parts: feature extraction, nonlinear mapping of residual function, and SR image reconstruction

The whole network is composed of three parts: feature extraction, nonlinear mapping of residual function, and SR image reconstruction. The LR input image passes through a convolutional layer and a Relu layer to extract features, and then, the extracted features are input into several residual blocks, and the residual mapping function is learned recursively. Finally, at the end of the network, a deconvolution layer is used to directly upsample the learned residual image to reconstruct the SR output image.

### 1.6 Evaluation criteria

Objectively, the deviation error between the restored image and the original image is generally used to evaluate the quality of the image restoration. In this paper, peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) are used as reference evaluation indicators for image quality.

The larger the PSNR value, the smaller the difference between the reconstruction result and the original image, the better the reconstruction effect. The calculation formula is as follows:

$$PSNR = 10 * \log_{10} \left\{ \frac{[\max(X_i)]^2 * MN}{\|X_i - Y_i\|^2} \right\} \quad (10)$$

Among them,  $X_i$  is the high-resolution image of the original reference,  $Y_i$  is the reconstructed image,  $M$  and  $N$  are the height and width of the image, and generally, the maximum value of  $\max(X_i)$  is 255, which can be directly substituted in the formula.

PSNR is mainly based on the comparison between pixels, and the evaluation of the local structure of the image is relatively weak. Sometimes the PSNR values of the two images are close, but the visual effects of the images are very different. Images generally have their own structures, and there is more or less correlation between adjacent pixels. SSIM is a structural parameter between the reconstruction result and the reference high-resolution image. The calculation formula is as follows:

$$SSIM(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (11)$$

Among them,  $X$  and  $Y$  are the reference HR image and the restored result image, respectively.  $\mu_X$  and  $\mu_Y$  represent the average pixel value of two image pairs, which are defined as follows:

$$\mu_X = \frac{1}{N} \sum_{i=1}^N X(i) \quad (12)$$

$$\mu_Y = \frac{1}{N} \sum_{i=1}^N Y(i) \quad (13)$$

$N$  is the number of dimensions to expand the image by column.  $\sigma_X$  and  $\sigma_Y$  are the corresponding variance, defined as follows:

$$\sigma_X = \left( \frac{1}{N-1} \sum_{i=1}^N (X(i) - \mu_X)^2 \right)^{\frac{1}{2}} \quad (14)$$

$$\sigma_Y = \left( \frac{1}{N-1} \sum_{i=1}^N (Y(i) - \mu_Y)^2 \right)^{\frac{1}{2}} \quad (15)$$

$\sigma_{XY}$  is the covariance, which is defined as:

$$\sigma_{XY} = \frac{1}{N-1} \sum_{i=1}^N (X(i) - \mu_X)(Y(i) - \mu_Y) \quad (16)$$

$C_1$  and  $C_2$  are normal number whose denominator is not zero. The value of SSIM ranges from 0 to 1. The closer the value is to 1, the more similar the two images are, and the better the reconstruction result is.

ERGAS is a quality evaluation method proposed for image fusion research, which reflects the degree of spectral distortion between the restored image and the reference image. It is also commonly used in the super-resolution restoration quality evaluation of images. The calculation formula is as follows:

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{K} \sum_{k=1}^K \left( \frac{RMSE(k)}{\mu(k)} \right)^2} \quad (17)$$

$l$  and  $h$  represent the resolution before and after image reconstruction,  $K$  represents the number of bands,  $\mu(k)$  represents the average of  $k$  band, and  $RMSE$  represents the root mean square error of the image. The ideal value of ERGAS is 0.

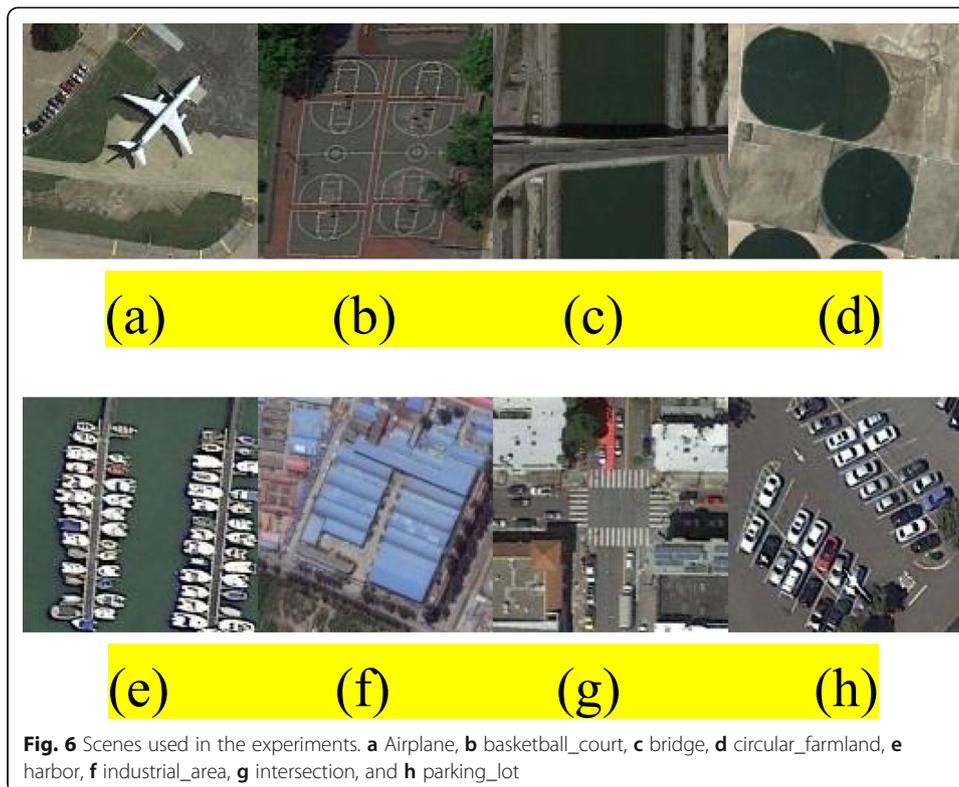
## 2 Results and discussion

### 2.1 Experimental environment and settings

The experimental software environment uses Ubuntu 14.04, Python 2.7, TensorFlow 1.4; the hardware environment is Intel Core i7-6700K, RAM 16GB, and the GPU is NVIDIA GTX1080. We use remote sensing image scene classification data set NWPU-RESIS45 [37] created by Northwestern Polytechnical University. Data set includes 45 scenes, each scene has 700 images, and each image size is 256×256, ensuring the authenticity and diversity of experimental data.

From each type of remote sensing image, 100 images with obvious features are selected, with a total of 4500 images. These images constitute a training data set to train the algorithm model. In addition, a total of 450 images of each type are chosen as test data sets, and different SR algorithms (SRCNN, FSRCNN, DRCN, VDSR, EDSR, and IDRRN) are used to simulate the test results. There are some of the training images as shown in Fig. 6, comprising the following scenes: airplane, basketball\_court, bridge, circular\_farmland, harbor, industrial\_area, intersection, and parking\_lot.

For input images, first use the magnification factor  $n$  to downsample the original training image, and it becomes an LR image. Then crop the LR image into a set of sub-images with stride  $s$  and  $size_{sub} \times size_{sub}$  pixel and crop the corresponding size from the corresponding real image to  $(n \cdot size_{sub})^2$  pixel HR sub-images. These LR/HR sub-image pairs are training samples. To ensure that the image size does not change during the mapping process, the convolutional layers are filled with "0." When training IDRRN, the deconvolution filter will generate a size of  $(n \cdot size_{sub} - n + 1)^2$  output image. Therefore, we need to crop the  $n - 1$  pixel boundaries of HR sub-image.



## 2.2 Quantitative results of SR methods

The network depth of the IDRRN algorithm proposed in this paper has 12 layers. The filter size should be odd so that it has a center, such as  $3 \times 3$ ,  $5 \times 5$ , or  $7 \times 7$ . The use of smaller convolution kernels is one of the current trends to reduce parameters while ensuring network accuracy.

The parameter setting of the convolution layer is the same as VDSR [24]. All convolutional layer filters are  $3 \times 3$  in size and the number of filters is 64. The deconvolution uses the mean value of 0, the standard deviation is 0.001 random initialization of Gaussian distribution, and take Relu function as activation function. The size of the filter refers to the DRCN algorithm [35], which is  $5 \times 5$ . The step is equal to the amplification factor  $n$ . During training, the size of the image batch is 128, the momentum is 0.9, and the weight attenuation parameter is 0.0001. The initial learning rate is set to 0.1, then the learning rate is halved every 15 generations; the learning stops after 120 generations, and the loss function is the MSE (mean square error) function.

The performance of the proposed approach has been compared with the results obtained by six different SR methods available in the literature (Bicubic, SRCNN [22], FSRCNN [28], DRCN [35], VDSR [24], and EDSR [30]). Three different scaling factors,  $\times 2$ ,  $\times 3$ , and  $\times 4$ , have been tested over the considered image data set (*airplane*, *bridge*, *harbor*, *intersection*, and *parking\_lot*). All the tested methods have been used considering the default settings suggested by the methods' authors for each particular scaling ratio. Table 1 provides a brief PSNR/SSIM description of the SR techniques.

As shown in Table 1, the average PSNR and SSIM values of the images generated by the method in this paper are higher than other current mainstream SISR algorithm. The PSNR values are optimal in 5 types of scenarios. The maximum boost value is

**Table 1** PSNR/SSIM values of state-of-the-art SR methods. The bold number indicates the best performance

Scene	Scale	Bicubic	SRCNN	FSRCNN	DRCN	VDSR	EDSR	IDRRN
Airplane	x2	31.43/0.8999	33.78/0.9270	34.05/0.9287	34.19/0.9322	34.78/0.9347	35.01/0.9354	<b>35.20/0.9493</b>
	x3	29.12/0.8211	30.53/0.8439	31.12/0.8502	31.25/0.8667	31.41/0.8961	31.37/0.8863	<b>31.78/0.8997</b>
	x4	27.03/0.7529	27.99/0.7718	28.24/0.7945	28.97/0.7940	29.26/0.8242	29.39/0.8271	<b>29.42/0.8288</b>
Bridge	x2	27.94/0.8813	28.50/0.9003	28.89/0.9039	29.02/0.9024	29.50/0.9168	29.42/0.9076	<b>30.03/0.9194</b>
	x3	25.89/0.7714	27.01/0.7948	27.52/0.8109	27.99/0.8072	28.33/0.8222	28.48/0.8115	<b>28.49/0.8258</b>
	x4	24.33/0.7253	25.12/0.7501	25.49/0.7592	25.71/0.7663	26.05/0.7620	25.47/0.7708	<b>26.44/0.7854</b>
Harbor	x2	21.16/0.8944	23.63/0.9058	23.97/0.9139	24.09/0.9243	26.05/0.9440	26.31/0.9502	<b>26.35/0.9593</b>
	x3	18.49/0.7467	20.27/0.7580	20.81/0.7789	21.13/0.7941	22.16/0.8102	22.39/0.8003	<b>22.48/0.8115</b>
	x4	16.93/0.5465	18.50/0.6019	18.84/0.6463	19.01/0.6740	19.36/ <b>0.7509</b>	19.33/0.7482	<b>19.42/0.7485</b>
Inter-section	x2	19.84/0.7444	20.95/0.8039	21.28/0.8176	21.13/0.7916	23.27/0.8513	22.19/0.8240	<b>23.86/0.8532</b>
	x3	17.49/0.5894	18.30/0.6189	19.52/0.6540	20.01/0.6693	20.12/0.7578	19.64/0.7602	<b>20.16/0.7733</b>
	x4	14.47/0.4413	15.12/0.4914	15.36/0.4805	14.83/0.5301	15.02/0.5016	15.96/ <b>0.5633</b>	<b>16.43/0.5605</b>
Parking_lot	x2	24.91/0.8081	26.02/0.8192	26.99/0.8206	27.38/0.8243	28.01/0.8444	27.94/0.8135	<b>28.03/0.8487</b>
	x3	22.47/0.6176	22.96/0.6895	23.53/0.6543	23.97/0.7014	24.81/0.7288	25.02/0.7355	<b>25.28/0.7356</b>
	x4	21.04/0.5543	21.77/0.5720	22.84/0.5782	23.15/0.5924	22.98/0.6013	23.55/0.5947	<b>23.78/0.6172</b>

5.19dB, when under  $\times 2$  magnification, the maximum boost value is 3.99dB when under  $\times 3$  magnification, and the maximum boost value is 2.74dB when under  $\times 4$  magnifications. In terms of value, except for the  $\times 4$  magnification conditions of *harbor* and *inter-section*, the rest are optimal. The algorithm in this paper reaches maximum boost value 0.1088 at  $\times 2$  magnification, maximum boost value 0.1839 at  $\times 3$  magnification, and maximum boost value 0.0759 at  $\times 4$  magnification.

Because of the particularity of remote sensing images, this paper uses ERGAS value in Formula (17) to compare the SR effect in order to further verify the effectiveness of the improved algorithm. From Table 2, we can get that among the 15 ERGAS data results, the IDRRN algorithm obtained 11 optimal values.

By analyzing and comparing the SR results of Tables 1 and 2, we find that the recursive residual learning can transfer more effective image information to the depth of the network, learn more image features, and make the image restoration quality improve greatly.

Furthermore, the proposed IDRRN approach from inherent parameter sharing obtains higher parameter efficiency compared to other learning-based methods. In Fig. 7, we illustrate the parameters-to-PSNR relationship of our model and several state-of-the-art methods, including SRCNN, FSRCNN, DRCN, VDSR, and EDSR. Our method represents a favorable trade-off between model size and SR performance and has modest processing time.

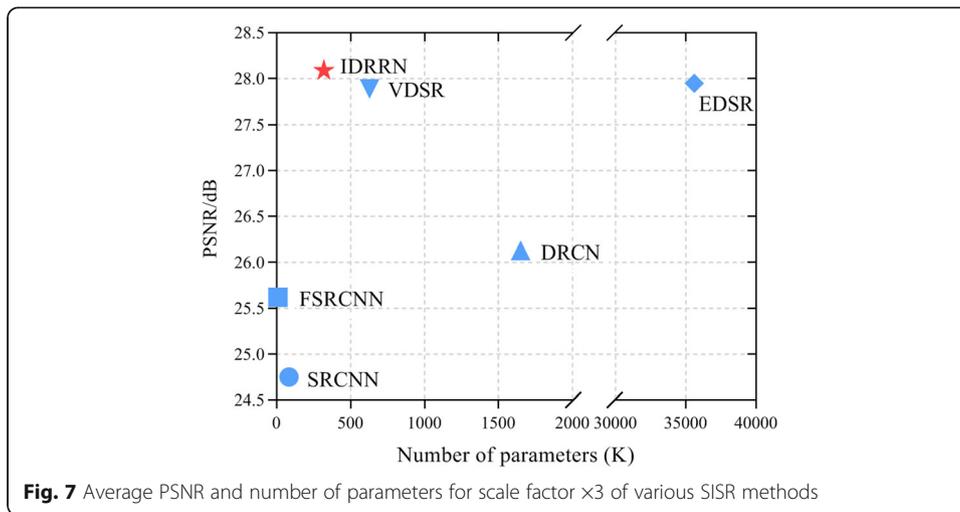
The addition of improved recursive structure does not need to increase the number of parameters. In addition, it improves the restoration quality of the image. The network structure is more compact and the objective performance is better.

### 2.3 Visual results and discussion

In order to demonstrate the effectiveness of our approach more fully, we also show some of the visual comparisons on three scales  $\times 2$ ,  $\times 3$ , and  $\times 4$ . Figures 8, 9, and 10

**Table 2** ERGAS values of state-of-the-art SR methods. The bold number indicates the best performance

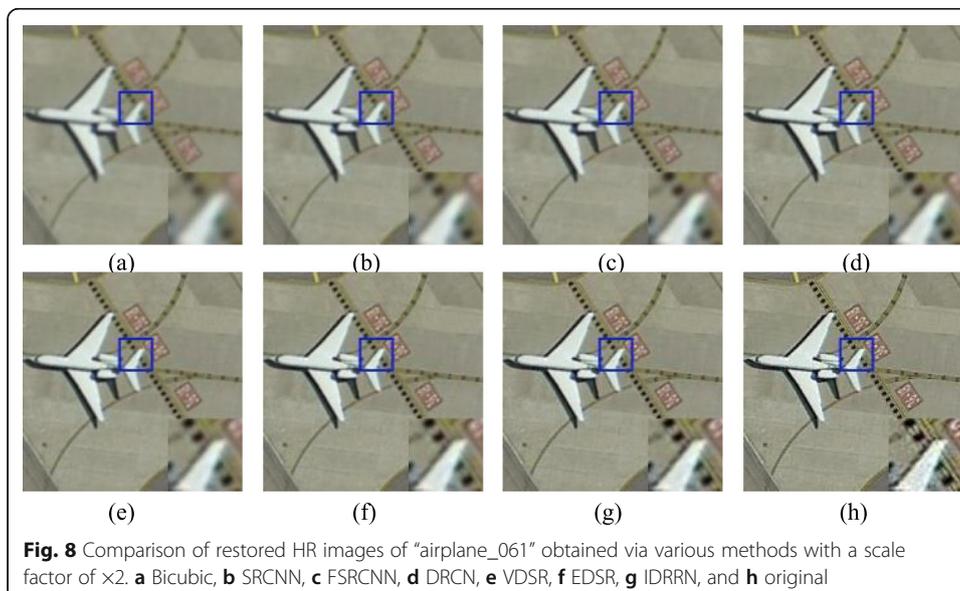
Scene	Scale	Bicubic	SRCNN	FSRCNN	DRCN	VDSR	EDSR	IDRRN
Airplane	$\times 2$	3.1462	2.6994	2.4515	2.3379	2.2613	2.2075	<b>2.1987</b>
	$\times 3$	2.9942	2.6364	2.3988	2.4103	2.3015	2.2889	<b>2.2044</b>
	$\times 4$	2.6828	2.4634	2.4077	2.3345	2.1613	2.0481	<b>1.9985</b>
Bridge	$\times 2$	6.5783	6.0234	5.9902	5.6359	5.2230	5.5834	<b>5.1681</b>
	$\times 3$	5.7809	5.5293	5.4437	5.2569	5.1571	<b>4.9849</b>	4.9901
	$\times 4$	4.8862	4.6293	4.6128	4.5153	4.3820	4.4121	<b>4.3815</b>
Harbor	$\times 2$	9.4446	8.1335	8.3982	9.0341	7.0818	7.9539	<b>6.3731</b>
	$\times 3$	11.2215	10.0231	9.9025	8.9336	<b>8.0004</b>	8.0349	8.1795
	$\times 4$	12.9885	11.3769	10.5539	10.2495	10.5594	10.5104	<b>9.5517</b>
Inter-section	$\times 2$	14.2147	11.2275	11.0060	12.4054	11.1685	11.7313	<b>10.8496</b>
	$\times 3$	13.4601	12.7299	12.0232	11.9441	11.5986	12.7035	<b>11.0287</b>
	$\times 4$	15.1112	13.0236	12.7750	13.6313	11.0426	<b>11.0405</b>	11.4467
Parking_lot	$\times 2$	7.7762	7.4731	7.5994	6.9825	6.9904	7.5398	<b>6.8247</b>
	$\times 3$	6.3143	6.1019	5.7282	5.5627	5.1651	<b>5.0154</b>	5.3081
	$\times 4$	5.8401	4.6099	4.5310	6.7772	5.5812	4.8773	<b>4.5238</b>

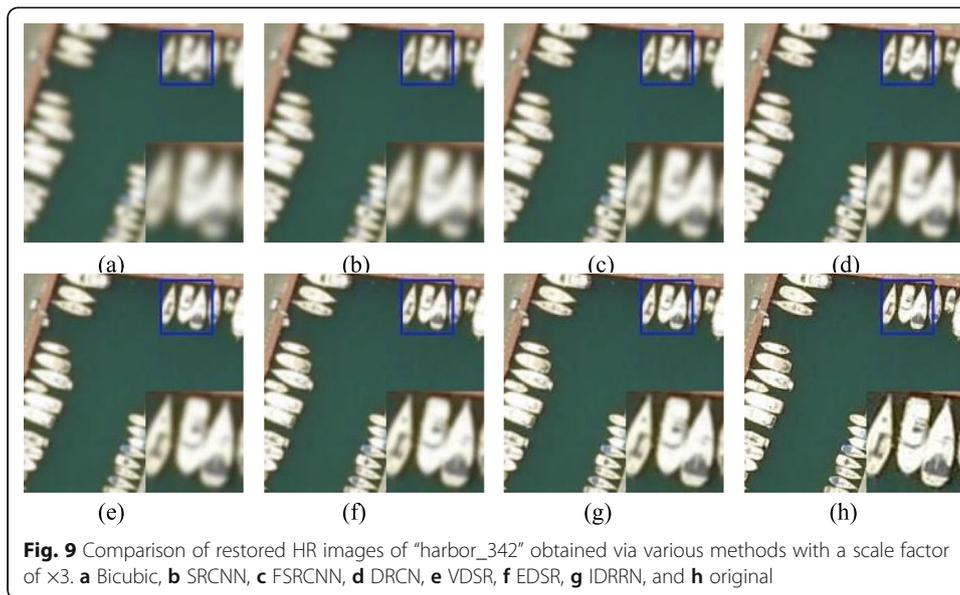


show the qualitative evaluation results of various algorithms. By enlarging the details of the image, the quality of image restoration of several SISR methods can be intuitively evaluated from the visual effect.

It can be seen from the figures that our method has a significant improvement in both image sharpness and clarity. After image processing, it is easier to identify multiple image categories in the remote sensing image. IDRRN overcomes the shortcomings of the overall smooth reconstruction result of the traditional method, and the reconstruction result restores more high-frequency details.

In addition, from the comparison of the enlarged parts of the tail of the aircraft in Fig. 8, the ships in the port in Fig. 9, and the vehicles on the bridge in Fig. 10, it can be seen that the image after the SR reconstruction by the IDRRN generation network is sharper compared with other mainstream algorithms. It has a better performance in the restoration of remote sensing image details, and it is more effective in repairing complex textures in damaged images. After repairing, the details in the image are richer

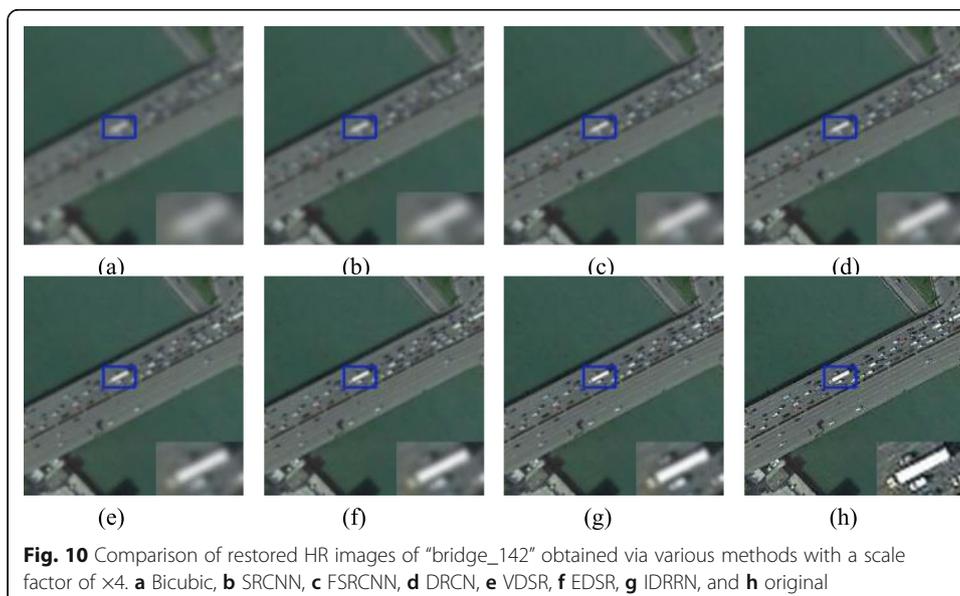




and more consistent with the visual characteristics of the human eye. With the SR restoration of the remote sensing image, the texture and edges are clearer, and the objects in the output image are easier to recognize.

### 3 Conclusion

In this paper, we propose a new type of residual network that introduces an improved recursive structure in the residual block. The jump connection and recursive structure can effectively reduce the burden of carrying characteristic information on the network, achieving high-quality SR remote sensing image recovery. Experiments were performed using the NWPU-RESISC45 remote sensing image data set, and PSNR, SSIM, and ERGAS are the objective quality evaluation index of image SR. Experimental results show that compared with other super-resolution methods based on CNN, the method



in this paper has more compact network structure and fewer model parameters, and the reconstruction details are more abundant. Moreover, the restoration results have better visual effects and are more conducive to further remote sensing image analysis.

In the next work, we will try to generalize the proposed IDRRN method to color images by designing a more compact network structure and improving the loss function of the model. In addition, we hope to further improve the details of super-resolution images and the repair effect of complex textures.

#### Abbreviations

SISR: Single-frame image super-resolution; IDRRN: Improved deep recursive residual network; HR: High resolution; LR: Low resolution; IBP: Iterative backward projection; POCS: Projection onto convex sets; MAP: Maximum a posterior; TV: Total variation; GTV: General total variation; CNN: Convolutional neural network; SRCNN: Super-resolution convolutional neural network; FSRCNN: Fast super-resolution convolutional neural network; VDSR: Very deep convolutional networks for image super-resolution; SRGAN: Super-resolution generative adversarial network; EEDS: End-to-end deep and shallow networks; EDSR: Enhanced deep super-resolution network; ResNet: Residual network; DRCN: Deeply-recursive convolutional network; PSNR: Peak signal-to-noise ratio; SSIM: Structural similarity; ERGA S: Erreur Relative Globale Adimensionnelle de Synthèse; MSE: Mean square error

#### Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

#### Authors' contributions

Jiali Tang and Jie Zhang designed the algorithm and revised the article content. Jiali Tang carried out the experiments and wrote the manuscript. Dan Chen and Najla Al-Nabhan gave the suggestions on the structure of the manuscript and participated in modifying it. Chenrong Huang approved the final manuscript and optimized the English language. The authors read and approved the final manuscript.

#### Authors' information

Jiali Tang received the Ph.D. degree in Mechatronic Engineering from Jiangsu University, China, in 2016. He is currently an Associate Professor at the College of Computer Engineering, Jiangsu University of Technology. His research interests include image processing, computer vision, and pattern recognition.

Jie Zhang received the M.S. degree in Computer Applications Technology from Yangzhou University, China, in 2012. He is currently an Associate Professor at the College of Computer Engineering, Jiangsu University of Technology. His research interests include image processing and intelligent information systems.

Dan Chen received the M.S. degree in Computer Application Technology from Nanjing University of Aeronautics and Astronautics, China, in 2009. She is currently a lecturer with the College of Computer Engineering, Jiangsu University of Technology. Her research interests include embedded hardware and intelligent information system.

Najla Al-Nabhan works as an Assistant Professor, at the Computer Science Department (CS), College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia. She received her Ph.D. degree in CS, from KSU. She was also a visiting Ph.D. student at George Washington University. Her research interest includes wireless sensor networks, the Internet of Things, networking and systems, communication modeling, mobile computing, cloud computing, emergency management, and UAVs.

Chenrong Huang is a professor at the Nanjing Institute of Technology. She received the M.S. and Ph.D. degree in Computer Application Technology from the Nanjing University of Science & Technology, China, in 1992 and 2005, respectively. Her current research interests include computer vision and image understanding.

#### Funding

This work was supported by the Opening Project of Jiangsu Key Laboratory of Advanced Numerical Control Technology (SYKJ201804), by the Project funded by Jiangsu Postdoctoral Science Foundation (2019K041), by Nanjing Subsidy Project of IUR Cooperation (201722043), and by Changzhou Sci&Tech Program (CE20195030). It was also supported in part by the Deanship of Scientific Research at King Saud University for funding this work through research group No. RG-1441-331.

#### Availability of data and materials

The supporting data involved in the current study are available from the corresponding author by reasonable request.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>College of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China. <sup>2</sup>Jiangsu Key Laboratory of Advanced Numerical Control Technology, Nanjing 211167, China. <sup>3</sup>Department of Computer Science, King Saud University, Riyadh 11362, Saudi Arabia. <sup>4</sup>School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China.

Received: 3 August 2020 Accepted: 28 April 2021

Published online: 24 May 2021

## References

1. H. Ghassemian, A review of remote sensing image fusion methods. *Inf. Fusion* **32**, 75–89 (2016)
2. J.C. White, N.C. Wulder, M. Vastaranta, T. Hilker, P. Tompalski, Remote sensing technologies for enhancing forest inventories: a review. *Can. J. Remote Sens.* **42**(5), 619–641 (2016)
3. J.M. Haut, R. Fernandez-Beltran, M.E. Paoletti, J. Plaza, A. Plaza, F. Pla, A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **56**(11), 6792–6810 (2018)
4. W. Ma, Z. Pan, J. Guo, B. Lei, Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net. *IEEE Trans. Geosci. Remote Sens.* **57**(6), 3512–3527 (2019)
5. J. Gu, X. Sun, Y. Zhang, K. Fu, and L. Wang, Deep residual squeeze and excitation network for remote sensing image super-resolution. *Remote Sens.* **11**(15), 1817 (2019).
6. S. Singh, M.K. Kalra, J. Hsieh, P.E. Licato, S. Do, H.H. Pien, M.A. Blake, Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. *Radiology* **257**(2), 373–383 (2010)
7. X. Li, W. Fu, Regularized super-resolution restoration algorithm for single medical image based on fuzzy similarity fusion. *J. Image Video Proc.* **2019**, 83 (2019)
8. H. Shen, L. Zhang, B. Huang, P. Li, A MAP approach for joint motion estimation, segmentation, and super resolution. *IEEE Trans. Image Process.* **16**(2), 479–490 (2007)
9. S. Huang, J. Sun, Y. Yang, Y. Fang, Y. Que, Robust single-image super-resolution based on adaptive edge-preserving smoothing regularization. *IEEE Trans. Image Process.* **27**(6), 2650–2663 (2018)
10. X. Li, Y. Hu, X. Gao, D. Tao, B. Ning, A multi-frame image super-resolution method. *Signal Process.* **90**(2), 405–414 (2010)
11. S. Farsiu, M.D. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super resolution. *IEEE Trans. Image Process.* **13**(10), 1327–1344 (2004)
12. N. Del Gallego, J. Ilao, Multiple-image super-resolution on mobile devices: an image warping approach. *J. Image Video Proc.* **2017**, 8 (2017)
13. L. Zhou, X. Lu, L. Yang, A local structure adaptive super-resolution reconstruction method based on BTV regularization. *Multimed. Tools Appl.* **71**(3), 1879–1892 (2014)
14. W.T. Freeman, E.C. Pasztor, O.T. Carmichael, Learning low-level vision. *Int. J. Comput. Vision* **40**(1), 25–47 (2000)
15. C. Liu, H.Y. Shum, W.T. Freeman, Hallucinating faces: theory and practice. *Int. J. Comput. Vision* **52**(4), 1289–1306 (2007)
16. J.J. Li, X.H. Li, Super-resolution reconstruction method for single frame image based on clustering. *Comput. Eng.* **39**(7), 284–287 (2013)
17. J. Gao, Y. Wang, M. Cai, Y. Pan, H. Xu, J. Jiang, H. Ji, H. Wang, Mechanistic insights into EGFR membrane clustering revealed by super-resolution imaging. *Nanoscale* **7**(6), 2511–2519 (2015)
18. Q. Dai, S. Yoo, A. Kappeler, A.K. Katsaggelos, Sparse representation-based multiple frame video super-resolution. *IEEE Trans. Image Process.* **26**(2), 765–781 (2017)
19. J.C. Ferreira, E. Vural, C. Guillemot, Geometry-aware neighborhood search for learning local models for image super-resolution. *IEEE Trans. Image Process.* **25**(3), 1354–1367 (2016)
20. Z. Zhu, F. Guo, H. Yu, C. Chen, Fast single image super-resolution via self-example learning and sparse representation. *IEEE Trans. Multimedia* **16**(8), 2178–2190 (2014)
21. H. Yin, S. Li, L. Fang, Simultaneous image fusion and super-resolution using sparse representation. *Inf. Fusion* **14**(3), 229–240 (2013)
22. C. Dong, C.C. Loy, K.M. He, X.O. Tang, Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016)
23. W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, S. Yan, Deep edge guided recurrent residual learning for image super-resolution. *IEEE Trans. Image Process.* **26**(12), 5895–5907 (2017)
24. J. Kim, J. K. Lee, and K. M. Lee, Accurate image super-resolution using very deep convolutional networks, in *Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Nov. 2016, pp. 1646–1654.
25. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in *Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Nov. 2016, pp. 1874–1883.
26. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, Photo-Realistic single image super-resolution using a generative adversarial network, in *Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4681–4690.
27. U. Mudenagudi, S. Banerjee, P.K. Kalra, Space-time super-resolution using graph-cut optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(5), 995–1008 (2011)
28. Dong C, Loy CC, Tang X. *Accelerating the super-resolution convolutional neural network*, in *Proc. European Conf. Comput. Vis. (ECCV)*. Amsterdam: Springer; 2016. pp. 391–407.
29. Y. Wang, L. Wang, H. Wang, P. Li, End-to-End image super-resolution via deep and shallow convolutional networks. *IEEE Access* **7**, 31959–31970 (2019)
30. B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, Enhanced deep residual networks for single image super-resolution, in *Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 136–144.
31. C. Yan, B. Gong, Y. Wei, Y. Gao, Deep multi-view enhancement hashing for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* (2020)
32. C. Yan, B. Shao, H. Zhao, R. Ning, Y. Zhang, F. Xu, 3D room layout estimation from a single RGB image. *IEEE Trans. Multimedia* (2020)
33. C. Yan, Z. Li, Y. Zhang, Y. Liu, X. Ji, Y. Zhang, Depth image denoising using nuclear norm and learning graph model. *ACM Trans. Multimedia Comput. Commun. Appl.* (2020)
34. K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Nov. 2016, pp. 770–778.

35. J. Kim, J. K. Lee, and K. M. Lee, Deeply-recursive convolutional network for image super-resolution, in Proc. IEEE Conf. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Nov. 2016, pp. 1637-1645.
36. Y. Tai, J. Yang, X. Liu, *Image super-resolution via deep recursive residual network*, in Proc. IEEE Conf. Vis (Pattern Recognit. (CVPR), Honolulu, 2017), pp. 3147–3155
37. G. Cheng, J. Han, X. Lu, Remote sensing image scene classification: benchmark and state of the art. P. IEEE **105**(10), 1865–1883 (2017)

### **Publisher's Note**

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

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

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

---

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

---