

# Remote Sensing Image Super-Resolution Reconstruction based on Generative Adversarial Network

Aili Wang<sup>a,\*</sup>, Ying Wang<sup>a</sup>, Xiaoying Song<sup>a</sup>, and Yuji Iwahori<sup>b</sup>

<sup>a</sup>*Higher Education Key Lab for Measuring and Control Technology and Instrumentations of Heilongjiang*

*Harbin University of Science and Technology, Harbin, 150001, China*

<sup>b</sup>*Computer Science, Chubu University, Aichi, 487-8501, Japan*

---

## Abstract

The super-resolution reconstruction algorithm based on generative adversarial network (GAN) can generate realistic texture in the super-resolution process of a single remote sensing image. In order to further improve the visual quality of the reconstructed image, this paper will improve the generation network, discrimination network, and perceptual loss of the generated confrontation network. Firstly, the batch normalization layer is removed and dense connections are used in the residual blocks, which effectively improves the performance of the generated network. Then, we use the relative discriminant network to learn more detailed texture. Finally, we obtain the perception loss before the activation function to maintain the consistency of brightness. In addition, transfer learning is used to solve the problem of insufficient remote sensing data. The experimental results show that the proposed algorithm has superiority in the super-resolution reconstruction of remote sensing images and can obtain better subjective visual effects.

**Keywords:** generative adversarial network; remote sensing image super-resolution reconstruction; residual dense block; transfer learning

(Submitted on December 10, 2018; Revised on January 12, 2019; Accepted on February 10, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

With the continuous development of image processing technology, the application of digital images is increasingly extensive, and remote sensing images play an important role in industry, agriculture, military, and other fields [1-2]. In industrial and agricultural areas, remote sensing images are widely used in meteorological monitoring, geological surveys, natural disaster prevention, and so on. In the military field, remote sensing images are helpful for the precise guidance of weapons and consolidation of national defence security. High-resolution (HR) remote sensing images provide higher pixel density and include more discernible targets to provide more accurate information. However, it is difficult to obtain HR remote sensing images directly because of the limitation of the camera's resolution and atmospheric interference. Therefore, this paper uses the Super-Resolution (SR) reconstruction algorithm to improve the resolution of remote sensing images.

Single image super-resolution (SISR) aims to reconstruct HR images from a single low-resolution (LR) image. As a basic low-level vision problem, SISR has attracted more and more attention in academic research and artificial intelligence companies. In 2014, Chao applied Convolutional Neural Networks (CNN) in deep learning to image SR reconstruction firstly. Super-Resolution Convolutional Neural Networks (SRCNN) utilized the relationship between deep learning and traditional sparse coding through image block extraction, nonlinear mapping, and final reconstruction [3-4], significantly improving the reconstruction effect. In 2016, in view of the successful experience of deep networks [5-6] on ImageNet, Kim et al. proposed a very deep network (VDSR) and a deep recursive convolutional neural network (DRCN), which superimposed 20 convolution layers from the perspective of training efficiency and storage respectively [7-8]. Compared with the shallow model, these two models achieved better reconstruction results by increasing the network depth. However, the peak signal-to-noise ratio (PSNR) of these methods tended to produce excessively smooth results without sufficient high frequency details.

---

\* Corresponding author.

E-mail address: [aيلي925@hrbust.edu.cn](mailto:aيلي925@hrbust.edu.cn)

At present, scholars have proposed several perceptual driving methods to improve the visual quality of SR results. For example, Li et al. proposed perceptual loss to optimize the SR model in the feature space instead of the original pixel space one [9-10]. In 2017, Generative Adversarial Networks (GAN) was introduced into SR by Christian [11-13]. SR based on GAN (SRGAN) generated more natural-looking images by training networks to reconstruct more realistic textures. SRGAN was one of the milestones in the pursuit of visual subjective evaluation results. Its basic model was constructed by residual blocks [14]. Meanwhile, it applied perceptual loss to optimize the GAN framework. Through these techniques, SRGAN significantly improved the overall visual quality of the reconstruction. In the same year, Lim et al. extended the model size by removing the unnecessary batch normalization layer (BN layer) in the residual network, and they proposed the EDSR model [15]. The reconstruction results were significantly improved. In 2018, Xu et al. introduced dense connections in residual blocks based on residual networks in SR, which greatly reduced the number of parameters. The results showed that dense residual blocks achieved better SR performance in PSNR and SSIM [16].

This paper proposed an improved SR algorithm remote sensing images based on GAN. Firstly, the BN layer in residual blocks is removed; then, dense connection is adopted to achieve higher network capacity; and finally, the problem of insufficient remote sensing data is solved by transfer learning. The experimental results show that the subjective quality of the reconstructed image can be effectively improved by introducing the improved GAN into the super-resolution reconstruction of remote sensing images.

## 2. GAN Model and Improvements

SRGAN provides a powerful framework for generating natural images with high perceptual quality. There are two neural networks in the GAN framework: a generator and a discriminator, which are trained simultaneously. The discriminator network is trained to distinguish natural and synthetically generated images, while the generator learns to generate images that are indistinguishable from natural images by the best discriminator. The SRGAN network model is shown in Figure 1.

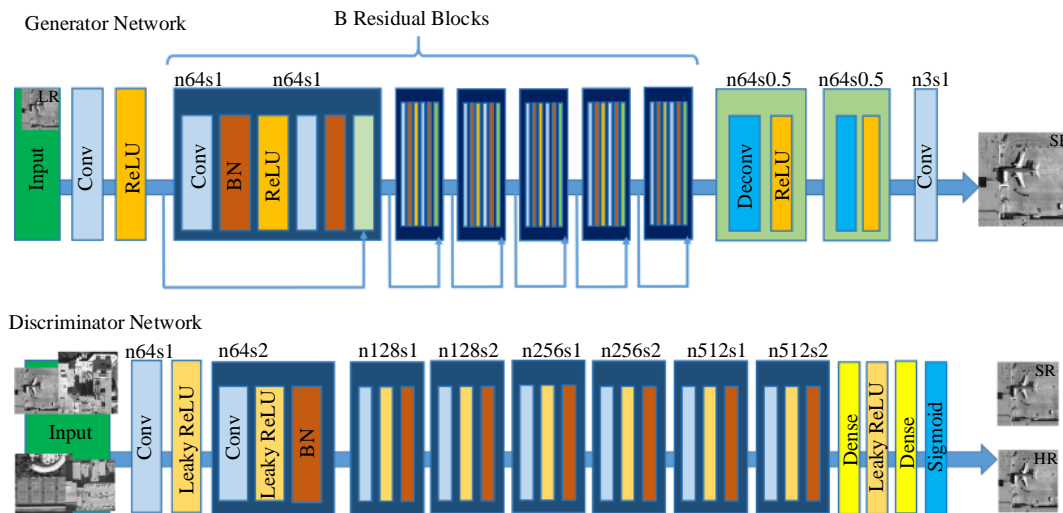


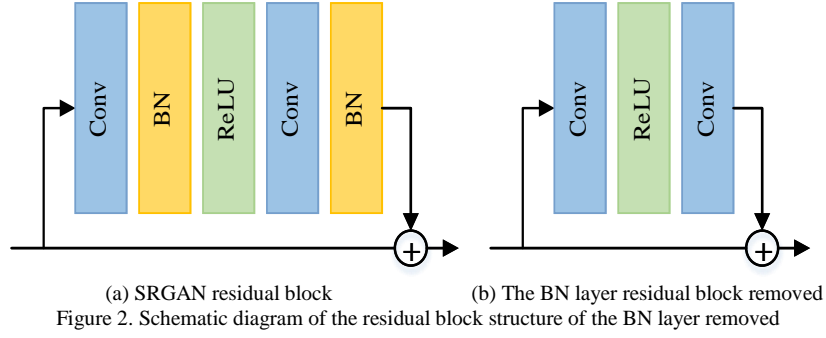
Figure 1. Architecture of Generative Adversarial Network (GAN)

In this paper, SRGAN is introduced into the process of remote sensing images super-resolution reconstruction, and four improvements are made: 1) remove the BN layer in the generator network model and introduce dense connections in the residual block; 2) use relativistic discriminator networks instead of the standard discriminator network; 3) change the way to obtain the perceptual loss after the activation function, and adjust it to obtain before the activation function; and 4) train the EDSR model with the candidate data set, and transfer the model parameters to the improved GAN model to make up for the shortage of the remote sensing data set.

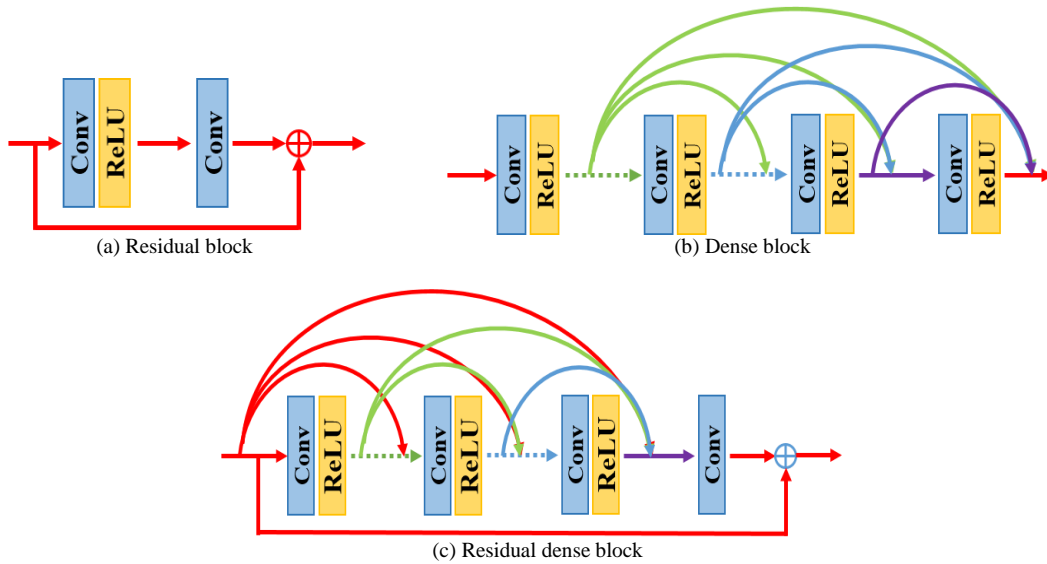
### 2.1. Residual Dense Block

The generating network in SRGAN has  $B$  residual blocks with the same layout. Firstly, the network extracts 64 feature maps using two convolution layers with a  $3 \times 3$  kernel function, followed by the BN layer, and it uses ReLU as the activation. Additionally, the generating network uses two deconvolution layers to increase the resolution of the input image. The BN layer uses the mean and variance of batches to normalize the features during training and uses the mean and variance of the entire training data set during the test. When the statistical data of training data sets and test data sets are very different, the

BN layer often introduces artifacts and limits the generalization ability of the network. Lim and other scholars have found that when deeper network models are trained, the BN layer is more likely to bring artifacts, and these artifacts will appear in different iterations, which affects the training stability of the network model. In addition, the BN layer will occupy the same size of memory space as the convolution layer. Compared with SRResNet, the EDSR model without the BN layer will reduce the memory usage by about 40% during the training process. Therefore, this paper will delete the BN layer, build a deeper model, and perform stable training in the case of limited computing resources, so as to improve the perception performance of the network. The comparison of residual blocks before and after improvement is shown in Figure 2. Furthermore, removing the BN layer also helps improve generalization capabilities while reducing the computational complexity of the network.



With the deepening of network model layers, the features in each convolutional layer will have different levels of receptive fields, and the objects in the image have different sizes, viewing angles, and aspect ratios. The layering feature of a very deep network will provide more details information for SR reconstruction. However, most of the current deep learning-based methods neglect to make full use of the hierarchical features of each convolutional layer in reconstruction. Thus, this paper will use the residual dense block instead of the original residual block in SRGAN. As shown in Figure 3, all hierarchical features of the original LR image are fully utilized by the residual dense block. It is difficult for a very deep network to extract the output of each convolutional layer in the LR space directly, while the residual dense block contains a dense connections layer and adaptively preserves the accumulated features by local feature fusion with local residual learning. The output of one residual dense block can be directly used as input to each layer of the next residual dense block, so that the extracted features are continuously transmitted.



## 2.2. Relativistic Discriminator

The standard discriminator network in SRGAN estimates the probability that an input image  $x$  is a real image, while the relativistic discriminator network estimates the probability that the real image  $x_r$  is more real than the generated false image  $x_f$ , as shown in Figure 4.

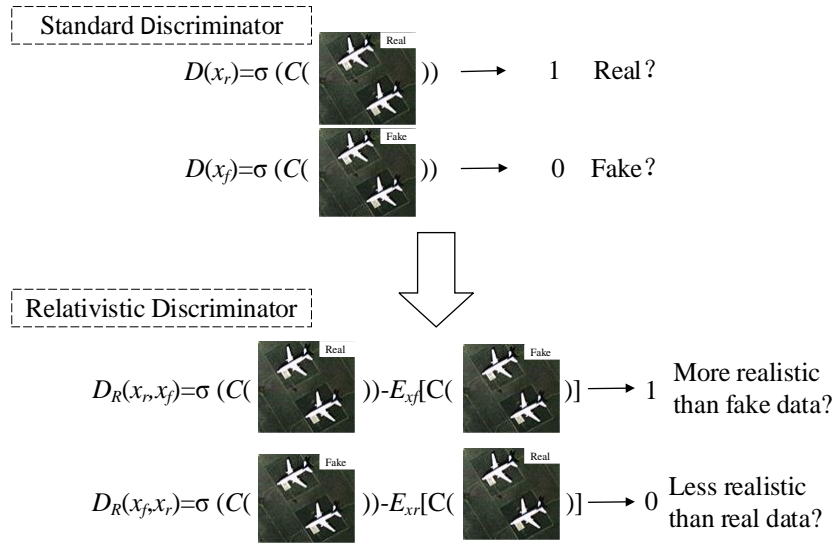


Figure 4. Comparison of the difference between standard discriminator network and relativistic discriminator network

In this paper, the relativistic discriminator network is used instead of the standard discriminator network and is expressed as  $D_{Ra}$ . The standard discriminator in SRGAN can be expressed as Equation (1):

$$D(x) = \sigma(C(x)) \quad (1)$$

Where  $\sigma$  is the sigmoid function and  $C(x)$  represents the output of the untransformed discriminator network. Then, denote  $D_{Ra}$  as Equation (2):

$$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}_{x_f}[C(x_f)]) \quad (2)$$

Where  $\mathbb{E}_{x_f}[\cdot]$  represents the operation of averaging all generated data in a mini batch, and then the loss of the discriminator network is defined as Equation (3):

$$L_D^{Ra} = -\mathbb{E}_{x_r}[\log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[1 - \log(D_{Ra}(x_f, x_r))] \quad (3)$$

The adversarial loss of the generator network is symmetrical, expressed as Equation (4):

$$L_G^{Ra} = -\mathbb{E}_{x_r}[1 - \log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(D_{Ra}(x_f, x_r))] \quad (4)$$

Where  $x_f = G(x_r)$  and  $x_r$  represents the input LR image. It can be seen from Equation (4) that the adversarial loss of the generator network includes two parts,  $x_r$  and  $x_f$ . Therefore, the process of generating network reconstruction images in this paper benefits from the gradient between the data generated in the training and the real data. In SRGAN, only the generated part is effective. Experiments show that the application of the relativistic discriminator network can help learn clearer edges and more detailed textures.

### 2.3. Improvement of Perception Loss

In SRGAN, the perceptual loss function proposed by Johnson et al. is applied. The loss function is defined on the active layer of the pre-trained VGG network. The Euclidean distance between feature maps is minimized to solve the smoothing problem of reconstructed results caused by pixel loss functions such as MSE. However, there are two disadvantages in dealing with features after activating the function: 1) the activated features are very sparse, especially after extremely deep networks, resulting in poor network performance; 2) the use of activated features will also lead to brightness inconsistency

between the reconstructed image and the real image, thus affecting the reconstructed results. Therefore, in this paper, feature extraction and processing are performed before the activation function, and then the total loss of the generator network is expressed as Equation (5):

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \quad (5)$$

Where  $L_1 = \mathbb{E}\|G(x_i) - x\|_1$  is the content loss that evaluates the 1-norm distance between the generated image  $G(x_i)$  and the real image  $x$ , and  $\lambda$  and  $\eta$  are coefficients to balance different loss terms.

#### 2.4. Transfer Learning

The existing excellent algorithms such as DRCN, SRResNet, and other deep networks have great advantages in SR reconstruction and have achieved good reconstruction results. However, training these deep networks is time-consuming and requires a high level of computer hardware configuration. At the same time, the deep network model needs to learn a large number of training data sets and obtains optimal network parameters through random gradient descent and back propagation. In this paper, SR reconstruction is performed on remote sensing data sets. However, in the experimental environment, the number of remote sensing data sets is limited, only a few thousand, which is difficult to meet the requirements of large data sets for training deep networks. Therefore, this paper introduces the method of transfer learning. Firstly, the EDSR model is pre-trained by ImageNet. The training data set contains 1.5 million training samples and 1,000 different target classes. Then, using the similarity between the EDSR model and the residual dense blocks in this paper, the initialization parameters are obtained by transfer learning. Finally, the improved GAN model is fine-tuned, that is, re-trained using remote sensing data sets, and then the final improved GAN-based SR reconstruction mode of the remote sensing image is obtained. The training process of improved GAN by using the transfer learning method is shown in Figure 5.

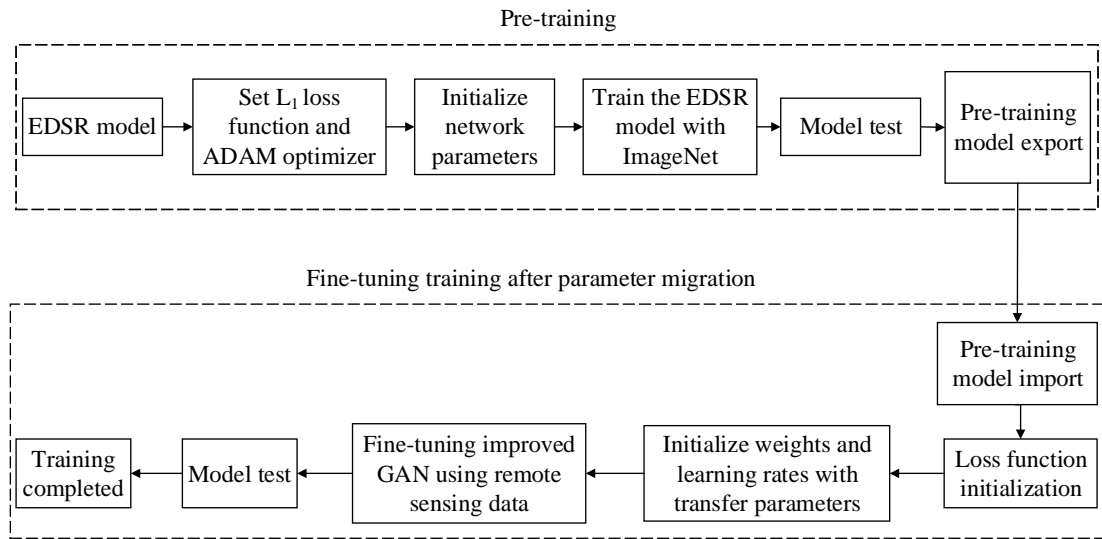


Figure 5. Improved GAN training process based on transfer learning method

The training process of the improved GAN model is divided into two parts: 1) the pre-training EDSR model is dominated by PSNR using the  $L_1$  loss function and ADAM optimizer, and the network is optimized by  $\text{beta}_1 = 0.9$  and  $\text{beta}_2 = 0.999$ . At the same time, the learning rate is initialized to  $10^{-4}$ , and the learning rate of every  $2 \times 10^5$  minibatch updates is reduced by twice. 2) The pre-training network model is used to generate the parameters of the network. Initialization and the total loss function are used to re-train the generating network.  $\lambda$  in Equation (5) is set to  $5 \times 10^{-3}$  and  $\eta$  is  $1 \times 10^{-2}$ . In particular, the loss function at the pixel level in the pre-training is more conducive to improve the visual effect of the reconstructed image, because it can avoid the local optimum of the generating network. After the pre-training, the input images of the discriminant network are relatively good SR images rather than extreme generated images, such as black or noise images, which can promote the recognition of more detailed textures by the discriminant network.

### 3. Experimental Results and Analysis

In this paper, the remote sensing image data collection collected by the U.S. Geological Survey from large images of

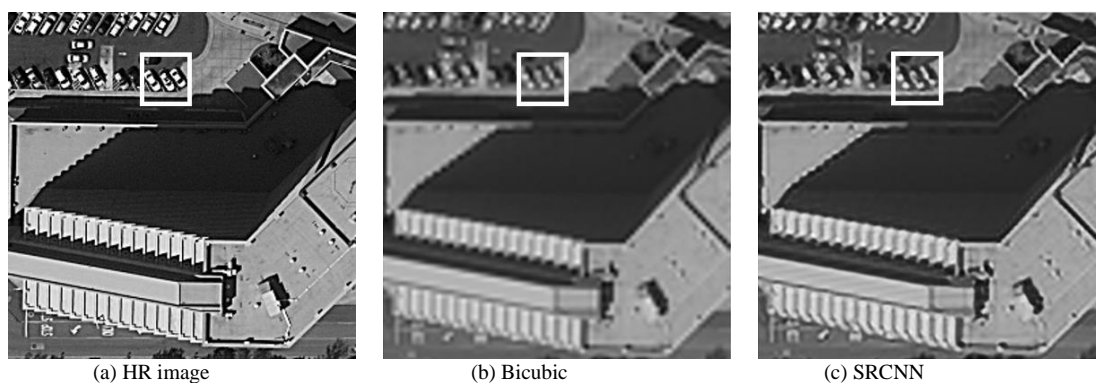
various urban areas includes 21 types of images, such as Boston airport, Dallas buildings, Venlatu tennis court, and Miami overpass. The size of each remote sensing image is  $256 \times 256$  pixels. Image examples of the data set are shown in Figure 6.



Figure 6. Image samples

The experimental steps are as follows: the first step is to pre-train the EDSR model in the ImageNet large-scale database to obtain the optimal network model parameters; the second step is to initialize the residual dense blocks in the improved GAN model using the optimal parameters obtained by transfer learning; and the third step is to select 50 images with obvious features from each kind of remote sensing image, totalling 1050 images, which constitute the training. The data set fine-tunes the GAN model. In addition, a total of 420 pieces of each type are selected as the test data set, and the reconstruction results are obtained using MATLAB 2014a for simulation. In terms of hardware, the experimental environment is a Dell-Inspiron-5420 notebook computer with NVIDIA-GT630M independent graphics card. Under the Ubuntu 14.04 system, the deep learning framework Caffe platform and the development package CUDA7.5 for calling GPU are installed and compiled.

This paper evaluates the super-resolution reconstruction algorithm of remote sensing images based on the improved generation antagonism network from two aspects of subjective evaluation and objective evaluation. In subjective evaluation, the improved SRGAN (ISRGAN) algorithm is compared with the bicubic algorithm, SRCNN algorithm, and standard SRGAN algorithm. The visual effect of aircraft and building reconstruction is shown in Figures 7 and 8. It can be seen clearly that the edges of the reconstructed image can be relatively sharp, while the edges of the reconstructed image can be blurred by other algorithms.





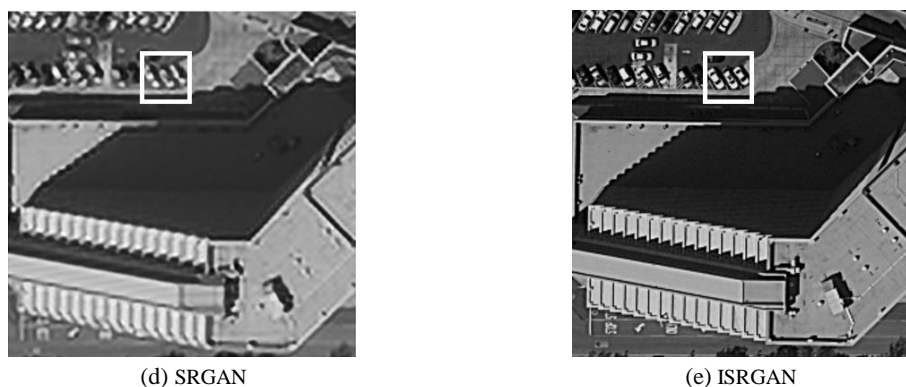


Figure 7. Reconstructed buildings by different SR algorithms

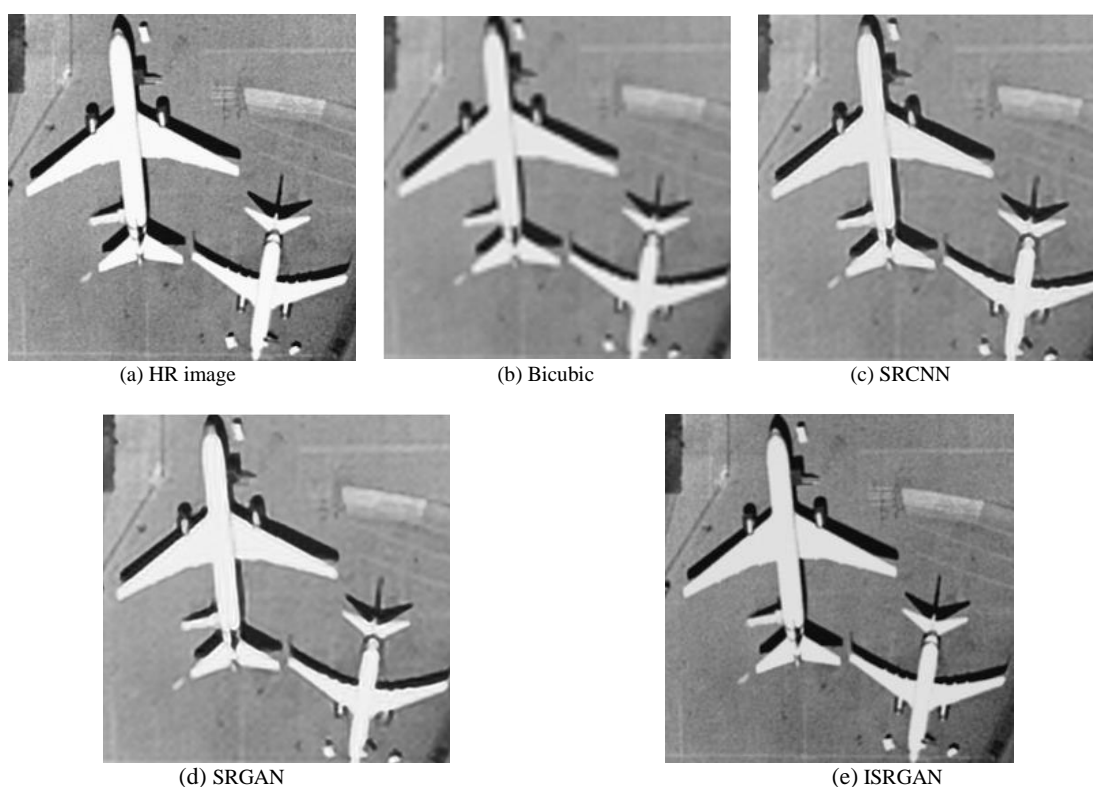


Figure 8. Reconstructed airplanes by different SR algorithms

The visual effect of reconstructed local details of buildings by different SR algorithms is shown in Figure 9. In Figure 9, the effect of the local detail image is more obvious. The edge of the reconstructed image is clearer by the SRCNN algorithm, and the texture details are not obvious. The SRGAN algorithm proposes a new perception loss function, adds the confrontational content loss function, and solves the problem that the SR reconstructed image is too smooth. Though the reconstructed image looks more real, there is still a certain gap with the HR image. In contrast, ISRGAN has higher image definition, enhances texture details, eliminates the influence of artifacts, and achieves the best visual effect.

In objective evaluation, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used to measure the quality of super-resolution reconstruction of remote sensing images, as shown in Table 1. It can be found that the PSNR and SSIM of the ISRGAN algorithm for various remote sensing images are not as good as those of other excellent algorithms. These excellent algorithms are PSNR-oriented and tend to output smoother reconstruction results with few high-frequency details. PSNR indicators do not always coincide with the subjective visual evaluation of human observers. This paper focuses on the subjective evaluation index of the super-resolution reconstruction algorithm, which is more suitable for video surveillance and other areas requiring reconstructed image fidelity.

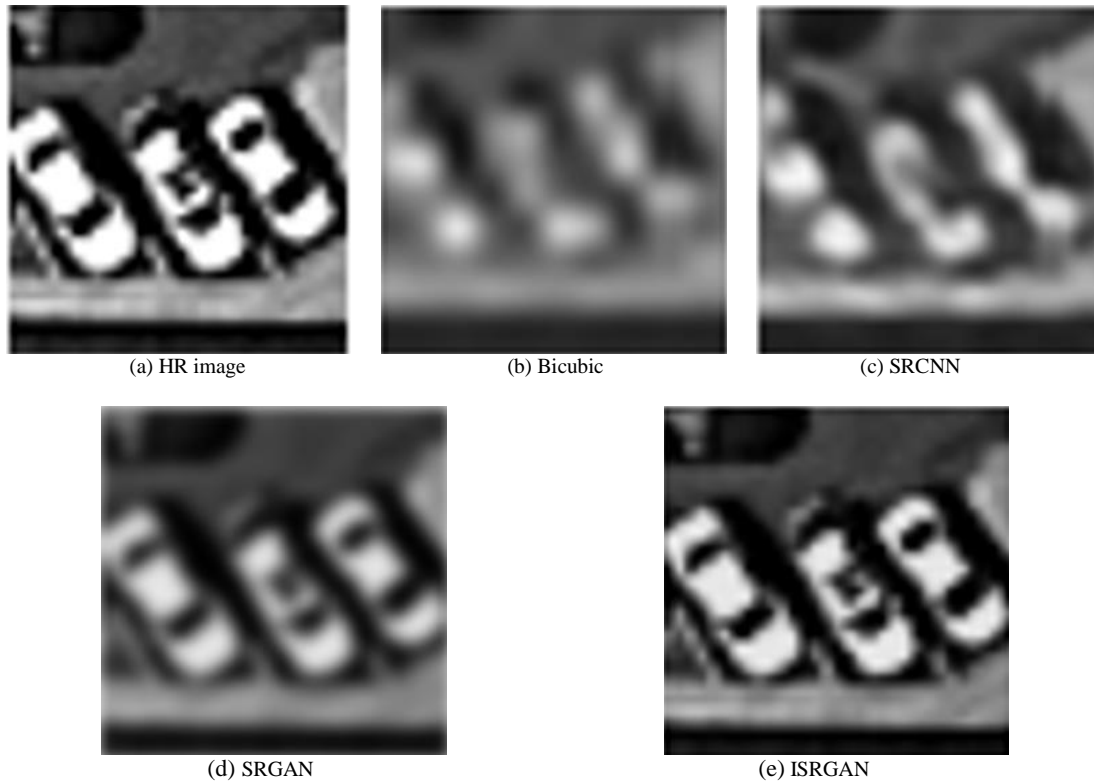


Figure 9. Reconstructed local details of buildings by different SR algorithms

Table 1. Comparison of PSNR (dB) and SSIM of different super-resolution reconstruction algorithms

Index	Method	Bicubic	SRCNN	SRGAN	ISRGAN
	Image				
PSNR (dB)	Viaduct	23.16	27.14	25.39	27.26
	Buildings	22.56	24.58	22.94	24.39
	Parking lot	23.39	28.02	26.07	28.23
	Tennis court	25.25	29.63	27.96	29.38
	Airplane	25.13	28.32	26.75	28.51
SSIM	Viaduct	0.8504	0.9088	0.8886	0.9091
	Buildings	0.8202	0.8452	0.8277	0.8442
	Parking lot	0.8413	0.9136	0.8934	0.9169
	Tennis court	0.7396	0.7691	0.7480	0.7678
	Airplane	0.7299	0.7943	0.7767	0.7961

#### 4. Conclusions

In this paper, the SRGAN model is improved. Compared with other excellent algorithms, the improved model achieves better reconstruction results. This method firstly removes the BN layer and at the same time uses a dense connection in the residual blocks, which increases the capacity of the network model, reduces the computational complexity, and enhances the stability of the network training. Secondly, the transfer learning technology is used to promote the training of the depth model, which solves the problem of insufficient remote sensing data. This paper also uses a relative discriminator instead of a standard discriminator, which can judge whether an image is more real than another one and guide the generation of more detailed texture for network reconstruction. In addition, perceptual loss is enhanced by using pre-activation features to reconstruct more accurate brightness and realistic texture.

#### References

1. Y. Y. Zhang, W. Wu, Y. Dai, X. M. Yang, B. Y. Yan, and W. Lu, "Remote Sensing Images Super-Resolution based on Sparse Dictionaries and Residual Dictionaries," in *Proceedings of IEEE International Conference on Dependable, Autonomic and Secure Computing*, pp. 318-323, 2013
2. W. Wu, X. M. Yang, K. Liu, Y. G. Liu, B. Y. Yan, and H. Hua, "A New Framework for Remote Sensing Image Super-Resolution: Sparse Representation-based Method by Processing Dictionaries with Multi-Type Features," *Journal of Systems Architecture*, Vol. 64, pp. 63-75, 2016



3. C. Dong, C. L. Chen, K. M. He, and X. O. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in *Proceedings of European Conference on Computer Vision*, pp. 184-199, 2014
4. C. Dong, C. C. Loy, K. He, and X. O. Tang, "Image Super-Resolution using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 38, No. 2, pp. 295-307, 2016
5. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proceedings of International Conference on Learning Representations*, pp. 473-491, 2015
6. K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016
7. J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution using Very Deep Convolutional Networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646-1654, 2016
8. J. Kim, J. K. Lee, and K. M. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1637-1645, 2016
9. J. Johnson, A. Alahi, and F. F. Li, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," in *Proceedings of European Conference on Computer Vision*, pp. 1-18, 2016
10. J. Bruna, P. Sprechmann, and Y. L. Cun, "Super-Resolution with Deep Convolutional Sufficient Statistics," in *Proceedings of International Conference of Learning Representation*, pp. 2-6, 2015
11. I. J. Goodfellow, J. P. Abadie, and M. Mirza, "Generative Adversarial Networks," in *Proceedings of Conference and Workshop on Neural Information Processing Systems*, pp. 4-7, 2014
12. C. Ledig, L. Theis, and F. Huszar, "Photo-Realistic Single Image Super-Resolution using a Generative Adversarial Network," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 105-114, 2017
13. S. M. Mehdi, B. Schölkopf, and M. Hirsch, "EnhanceNet: Single Image Super-Resolution through Automated Texture Synthesis," in *Proceedings of IEEE International Conference on Computer Vision*, pp. 4501- 4510, 2017
14. K. M. He, X. Y. Zhang, and S. Q. Re, "Deep Residual Learning for Image Recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016
15. B. Lim and S. Son, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1132-1140, 2017
16. J. Xu, Y. Chae, and B. Stenger, "Residual Dense Network for Image Super Resolution," in *Proceedings of IEEE International Conference on Image Processing*, pp. 71-75, 2018

**Aili Wang** received her B.S. degree from the Department of Electronic and Communication Engineering at Harbin Institute of Technology in 2002 and her M.S. degree and Ph.D. from the Department of Information and Communication Engineering at Harbin Institute of Technology in 2004 and 2008, respectively. She joined Harbin University of Science and Technology as an assistant in 2004 and became an associate professor in 2010. She was a visiting professor at Chubu University in 2014. Her research interests include image super-resolution, image fusion, and object tracking.

**Ying Wang** is a master's student at Harbin University of Science and Technology. His research interests include image classification.

**Xiaoying Song** is a master's student at Harbin University of Science and Technology. His research interests include image super-resolution.

**Yuji Iwahori** received his B.S. degree from Nagoya Institute of Technology in 1983 and his M.S. degree and Ph.D. from Tokyo Institute of Technology in 1985 and 1988, respectively. He joined Nagoya Institute of Technology in 1988 and became a professor there in 2002. He has been a professor at Chubu University since 2004. Meanwhile, he has also collaborated with UBC since 1991, IIT Guwahati since 2010, and Chulalongkorn University since 2014.