

# A Super-resolution Algorithm for Remote Sensing Images based on Data Enhancement and Generative Adversarial Networks

Wei Chen, Haiwei Shen \*

College of Applied Science and Technology of Beijing Union University, Beijing, China

\* Corresponding author: Haiwei Shen

---

**Abstract:** In this paper, a new algorithm is proposed to address the problem of excessive errors in current super-resolution algorithms for remote sensing images, which optimizes the processing effect of generative adversarial networks by increasing the type and size of training data with data enhancement techniques. First, the algorithm constructs a more reasonable degradation model for the problem of lacking data sets of paired images, and randomly mixes and washes the degraded prior knowledge (such as blur, noise, downsampling, etc.) in the imaging process to simulate the generation process of low-resolution remote sensing images in natural scenes, and generates realistic low-resolution images for training; meanwhile, the algorithm uses ResNet34 and CNN as the basis to enhance the texture details of remote sensing images, thus richer features can be extracted from the images. The results of experiments conducted on the Alsat2 B real dataset show that the method reduces the time and cost required for sample data acquisition, optimizes the processing speed, enhances the temporal and spatial resolution of remote sensing images, and improves the effect of super-resolution processing.

**Keywords:** Data Enhancement; Super-resolution; Remote Sensing Images; Generative Adversarial Networks.

---

## 1. Introduction

With the development of remote sensing satellite technology, the current high-resolution satellites of various countries have reached the centimeter level. In the field of remote sensing satellite images, the remote sensing images acquired by users are map products distributed by satellite centers after imaging processing, so the resolution of images can no longer be improved by adjusting the hardware. If the resolution of the finished satellite images is to be improved, the images can only be processed by software.

Image super-resolution reconstruction techniques can reconstruct the corresponding high-resolution images based on low-resolution images. After years of development, the research focus has shifted from traditional interpolation and reconstruction methods to methods based on deep learning. Dong et al.[1] first combined convolutional neural networks with super-resolution image reconstruction techniques and proposed the SRCNN algorithm, and then a large number of super-resolution reconstruction methods based on deep learning began to appear. Kim et al.[2] also introduced the VGG network structure to increase the number of network layers and use different sizes of perceptual fields to comprehensively extract image detail information in the shallow, intermediate, and high layers, solving the problem that the SRCNN algorithm relies on feature information of small image regions. He et al.[3] stacked multiple residual blocks to form a residual network (ResNet) to solve the problem of network degradation due to the deep structure of the convolutional network, and used the jump connection between residual blocks to enhance the image feature information transfer on different layers and alleviate the model gradient disappearance problem. After Goodfellow et al. [4] proposed GAN, many super-resolutions image reconstruction algorithms based on GAN have emerged, which have shown good results in terms of image

reconstruction effect, network computation, and computation speed. The generator of SRGAN is a ResNet34 architecture that extracts features from LR images. For upsampling, a subpixel convolutional layer is used. the SRGAN model was proposed by Ledig et al. [5] using ResNet and VGG as the generator and discriminator architectures, respectively. It uses the generator to generate HR images, the discriminator to discriminate between the reconstructed HR images and the original HR images, and the backward optimization of the generator and discriminator networks, while the traditional MSE loss function is replaced by the "perceptual loss" to enhance the image detail information recovery. Vu et al. [6] used relative generative adversarial network to replace the generative adversarial network in SRGAN, which makes the extraction and fusion of image details more reasonable and reduces the influence of noise and blur.

Compared with natural images, remotely sensed images contain rich topographical features such as canyons, mountains, and cities within a smaller pixel size, and thus detail loss is more serious. In addition, remote sensing images are also affected by atmospheric disturbance, cloud obscuration, noise and motion blur during the acquisition process, which makes super-resolution reconstruction more difficult. Due to the difficulty of remote sensing data set collection and the lack of corresponding high-resolution and low-resolution remote sensing image pairs for algorithm training, the effect of downstream tasks such as remote sensing image classification and target recognition can be effectively improved after super-resolution processing of remote sensing images. Aiming at the problems of insufficient sample data of remote sensing images and poor recovery effect of super-resolution reconstruction, this paper proposes a super-resolution reconstruction algorithm for remote sensing images based on data enhancement and generative adversarial network. Data augmentation expands the training set by systematically generating more training samples,

increases the amount of data for training, and improves the generalization ability and robustness of the model, which has achieved good results in the fields of machine learning, image recognition and image restoration; generative adversarial networks have also proved to be effective in the field of image processing and are widely used. The algorithm in this paper consists of a degradation model and a super-resolution reconstruction model. The degradation model uses randomly mixing and washing degradation factors (such as blur, downsampling, noise, etc.) to simulate the real imaging process to generate realistic low-resolution images combined with the original images in pairs for training; meanwhile, a new image processing architecture based on generative adversarial networks is proposed to enhance image details and preserve image features through ResNet34 and mini-CNN networks, and the method can better enhance the effect of super-resolution reconstruction of remote sensing images.

## 2. Methodology of this Paper

For the super-resolution reconstruction of remote sensing images that lack paired datasets for training, this paper proposes a method consisting of a degradation model and a reconstruction model. The overall network framework is shown in Figure 1, which is divided into two stages: training and testing.

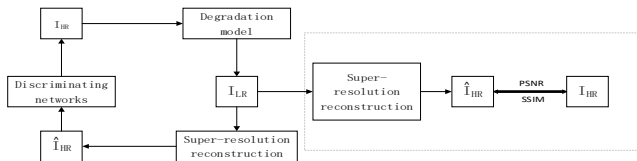


Figure 1. Overall Network Framework

In the training phase, the degradation model is used to process the high-resolution image  $I_{HR}$  in the UC Merced dataset to generate synthetic images with feature distributions approximating the real low-resolution image  $I_{LR}$ , and the composed images are trained on  $\hat{I}_{LR} \sim I_{HR}$ , and the  $I_{LR}$  is super-resolution reconstructed to generate high-resolution images  $\hat{I}_{HR}$ , and the discriminative network optimizes the training process by reducing the loss between  $\hat{I}_{HR} \sim I_{HR}$ ; in the testing phase, the AI-sat2B dataset is input to the trained super-resolution reconstruction model to obtain  $\hat{I}_{HR}$ , and then compare the subjective visual effect with the real  $I_{HR}$  and calculate PSNR/SSIM to qualitatively and quantitatively evaluate the reconstruction effect.

### 2.1. Degradation Model

In the super-resolution reconstruction of remote sensing images with unknown degradation model, how to accurately model the degradation model using remote sensing a priori knowledge is the key to enhance the reconstruction effect. In this paper, we adopt a random blending strategy to incorporate possible degradation factors (such as blurring, downsampling, noise, etc.) in the field of remote sensing.[7]

It is added to the generation process of low-resolution images. And according to the characteristics of remote sensing image imaging links, some degradation factors (such as isotropic Gaussian blur kernel, camera sensor noise and JPEG compression noise) are excluded, and possible

degradation factors (such as motion blur, additive noise and multiplicative noise) in remote sensing image imaging are added, so as to ensure the diversity of the synthesized low-resolution images and reflect the distribution of real image features as much as possible while avoiding synthesizing low-resolution images that do not match the natural low resolution images that do not match at all in the scene, reducing the training difficulty and speeding up the fitting process. The degradation process modeled in this paper consists of three factors: blur, downsampling, and noise, and in the actual imaging process. The degradation factors are disrupted by using a random mixing and washing strategy, so the order of the degradation factors is uncertain. Specifically, the fuzzy kernel is randomly selected from the anisotropic Gaussian fuzzy kernel, the downsampling is randomly selected from the nearest neighbor sampling, bilinear sampling, and bicubic sampling, and the noise is randomly selected from the additive noise and multiplicative noise, and then these three degradation factors are randomly disordered to obtain the final degradation model, as shown in Figure 2.

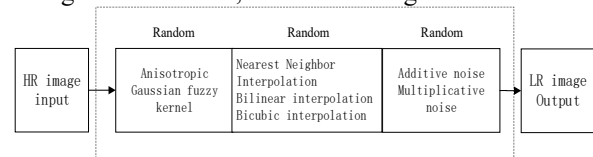


Figure 2. Degradation model

#### 2.1.1. Blur

The remote sensing image imaging process mainly contains motion blur and optical blur. Motion blur refers to the blurring effect caused by the relative motion between the sensor and the ground feature during the satellite motion and sensor scanning process, which is influenced by factors such as angle and movement length. When the remote sensing satellite is imaging, the sensor will receive a variety of radiation such as radiation reflected directly from the surface features, radiation reflected secondarily by the surface after downward scattering from the atmosphere and the part of upward scattering from the solar radiation, which eventually leads to the scattered in the actual imaging process, both motion blur and optical blur may occur, acting together to form image blur.

In the field of super-resolution reconstruction, the fuzzy kernel is modeled by the zero-mean Gaussian probability distribution function  $N(0, \Sigma)$ , and the rotation angle  $\theta$  and the eigenvalues of the eigenvectors of the covariance matrix  $\Sigma$  jointly determine the shape of the fuzzy kernel. In this paper, the rotation angle  $\theta$  is taken in the range of  $[0, \pi]$  and the eigenvalues  $l_1, l_2$  of the eigenvectors. are taken in the range of  $[0.2, 8]$ , and when  $l_1 = l_2$ , it is an isotropic Gaussian fuzzy kernel. The joint effect of motion blur and optical blur in remote sensing imaging leads to an isotropic fuzzy kernel with very low probability of occurrence and is no longer added separately.

#### 2.1.2. Downsampling

In the field of super-resolution reconstruction of remote sensing images, the bicubic interpolation method is usually used to downsample the high-resolution images, which can simultaneously achieve the requirements of shortening the processing time and improving the image quality. In order to enrich the number of samples for downsampling, in addition to bicubic interpolation, nearest-neighbor interpolation and bilinear interpolation are also used in this paper, and one of these three interpolation methods will be randomly selected in the downsampling operation.

### 2.1.3. Noise

Noise in remote sensing images may be introduced by different processing processes at different stages, including acquisition signal, acquisition noise, and internal sensor circuit noise. The imaging noise varies with different sensor types. In this paper, we mainly consider two noise models: additive noise model in optical imaging system and multiplicative noise model in remote sensing imaging system.[8]

## 2.2. Generative Adversarial Networks

Remote sensing images contain rich information of landforms and store a large amount of information of cities, roads, bridges, etc. within smaller pixels, which require high recovery of texture and level details of images. In the traditional method of convolutional neural network CNN, the perceptual field of the convolutional neural network is determined by the size of the convolutional kernel. All the information within the convolutional kernel is equally attended to, making the loss in the recovery process smaller in planar regions, but the loss is larger in the recovery process of non-planar regions such as bridges and building clusters, which may lead to a huge difference between the recovered image and the original image, resulting in information misalignment. In this paper, an image super-resolution method based on SRGAN is proposed, and its generator architecture is modified in the proposed method.

### 2.2.1. Generate Network

In order to overcome the drawbacks of SRGAN and to improve its image recovery accuracy, a novel generator network with the following three main changes is proposed in this paper. A block diagram of the proposed architecture illustrating the changes is shown in Figure 4. It consists of two main parts: depth feature extraction with 2-fold amplification, enhanced feature extraction and 2-fold enhanced extraction.

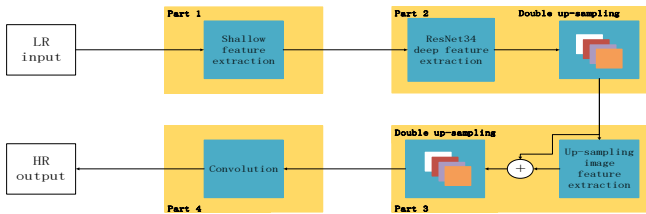


Figure 3. General flow chart of the algorithm

### 2.2.2. Shallow Feature Extraction

In this part of the architecture, the LR version of the original HR image is used as the input to the architecture. And basic or shallow features are computed at three different scales using kernels of sizes 3, 5 and 7, respectively. For feature extraction at each scale, two blocks are used. Each block consists of two convolutional layers, the first convolutional layer in each branch is followed by a batch normalization layer and ReLU activation, and after the second convolutional layer is a batch normalization layer with 64 channels per convolutional layer. a jump connection is used between the outputs of block 1 and block 2, and the features of block 1 will be superimposed feature by feature with those of block 2. After feature extraction at different scales, features from all three scales are concatenated (by channel) into a single feature vector to form a 192-channel wide feature vector. This feature vector contains the basic or shallow features of the LR image and serves as the input for the next step of the architecture.

### 2.2.3. Deep Feature Extraction and 2X Zoom

In this paper, 192 channels are used in each convolutional layer. The modified ResNet34 architecture consists of 17 blocks, each block containing two convolutional layers. The first convolutional layer is followed by a batch normalization layer and ReLU activation, and the batch normalization layer is followed by a second convolutional layer, where the features of each block are added one by one with the features of the previous block. In the last layer of the ResNet34 architecture, the size of the kernel is set to 3 to overcome the blurring effect and to retain more features. The last layer is added with the feature maps of the first part as elements. After the depth feature extraction, the extracted features are boosted by using a sub-pixel convolutional layer, and after obtaining the 2-fold zoomed LR image features, they are used as the input of this architecture in step 3.

### 2.2.4. Enhanced Feature Extraction and 2X Upgrade Enhancement

In this step, a 2x magnified LR image is obtained, after which it is necessary to calculate the features of the 2x magnified version. First, a mini-network including three residual blocks is used, as shown in Figure 6. Each residual block has two convolutional layers, the first convolutional layer is followed by the batch normalization layer and ReLU activation, and the batch normalization layer is followed by the second convolutional layer. At the end of the mini-residual network, the batch normalization layer is used with a 192-channel-wide convolutional layer. In a subsequent step, the feature map obtained after the 2-fold zoom is summed with the feature map elements extracted using the mini network, and this feature map is then zoomed by a factor of 2 using a sub-pixel convolutional layer. After obtaining the 4x magnified feature map of the LR image, the 4x magnified features are mapped back to the HR image using a convolutional reconstruction layer.

### 2.2.5. Discriminator Network

Since the generator model in this paper is a deep network, important modifications were made to the traditional discriminator SRGAN. First, the number of kernels in each layer is increased to 2048, and then the number of kernels is gradually reduced to 512, thus adding more convolutional layers to the discriminator. Then, three additional layers with the number of kernels equal to 128, 256 and 512 were added. After that the outputs were added before these three layers, with the final 512 kernel layer outputs.

## 3. Experiment

Due to the lack of datasets of real image pairs, most of the current methods are trained using synthetic low-resolution images and then the trained models are used to validate the super-resolution reconstruction effect on the synthetic low-resolution images, but both only validate the final effect on very few real low-resolution, such as KernelGAN [9] and IKC [10] both validate the results on only two real LR. In the field of super-resolution reconstruction of remote sensing images, Zhang [11] et al. validated on two jilin-1 images and Zhu [12] et al. validated on 364 GeoEye-1 images. Since these real low-resolution images do not have corresponding high-resolution images for comparative validation, only subjective visual effects are used to evaluate the effect on real low-resolution remote sensing images. In this paper, training on the UC Merced dataset and super-resolution reconstruction on the real dataset Alsat2B with paired images can be used to

judge the reconstruction effect by subjective visual effect evaluation and objective index. Compared with the original method, the method in this paper qualitatively and quantitatively validates the effectiveness of the super-resolution reconstruction algorithm on a larger dataset.

### 3.1. Shallow Feature Extraction

The UC Merced dataset is a widely used dataset in the field of remote sensing images, which contains 21 types of landforms such as farmland, airports, beaches, harbors, etc., 100 images of each type of landforms, a total of 2100 images, which are randomly divided in the text according to 70% as training set, 20% as validation set, and 10% as test set. Alsat2B dataset is used by Achraf equal to 2021 The latest real remote sensing satellite dataset released in 2021. The dataset covers different vegetation types and is size-sliced in blocks of 256\*256 for high-resolution images and 64\*64 for low-resolution images to obtain a total of 2,759 pairs of real remote sensing super-resolution datasets with a magnification of 4, of which 2,182 pairs are used as the training set and 577 pairs are used as the test set. The validation set is divided into 3 subsets according to the types of features, specifically 239 pairs of cities, 56 pairs of farmlands and 282 pairs of special buildings (e.g., bridges, stadiums, etc.) for the test set.

### 3.2. Experimental Setup

In order to verify the effectiveness of the degradation model and reconstruction model in this paper, multiple sets of experiments are conducted on the UC Merced dataset and the Alsat2B dataset. Firstly, the degradation model is used to process the UC Merced dataset to obtain the corresponding low-resolution images; then the image pairs are composed to train the super-resolution reconstruction network; finally, the super-resolution reconstruction effect is verified on the UC Merced test set and the Alsat2B test set, respectively, with the human subjective visual effect as the qualitative evaluation index, and the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as objective evaluation indexes.

The size of the convolution kernel of the residual module in the reconstruction network is 3\*3. The experiments are performed with Adam optimizer,  $\beta_1=0.9$ ,  $\beta_2=0.99$ ,  $\epsilon=10^{-8}$ , learning rate of 0.0001, batch size of 16, and 20,000 training rounds on a single 32 G NVIDIA Tesla V100 graphics card. After the training rounds reached 15,000, it gradually converged.

### 3.3. Experimental Results

The PSNR and SSIM are calculated on the UC Merced dataset, and the PSNR/SSJM of this paper improves 1.4071 dB/0.0672 compared with ESRGAN and 0.8211 dB/0.0235 compared with RCAN. The reconstructed image of RCAN method is too smooth and the texture effect is not good; while the method in this paper enhances the image details better and is visually better than other algorithms.

In order to further verify the effectiveness of the degradation model, this paper uses three different training sets trained by the reconstruction model for comparison experiments. One is the UC Merced dataset processed with dual triple interpolation; the second is the UC Merced dataset processed with the degradation model in this paper; and the third is trained directly with the training set in Alsat2B. The reconstruction models trained on the three different training sets are validated on the Alsat2B validation set respectively. Table 3 lists the quantitative evaluation metrics of the models

on the three different training sets on the Alsat2B test set in turn. From the results, it can be seen that the low-resolution images obtained by bi-trivial interpolation have huge differences in feature distributions from the real low-resolution images, resulting in their poorer results on the Alsat2B test set, while the UC Merced dataset processed with the degradation model of this paper has similar feature distributions to the Alsat2B training set, so it is closer in subjective visual effects and quantitative evaluation indexes, which also verifies the effectiveness of the degradation model is also verified, and the feature distribution of the generated low-resolution images is similar to that of the real low-resolution images.

## 4. Conclusion

The paper addresses the problem of poor recovery effect of current super-resolution reconstruction algorithms in the field of real remote sensing images, and adopts the method of randomly mixing and washing degradation factors to simulate the imaging process of real remote sensing images to generate low-resolution remote sensing images; meanwhile, a super-resolution reconstruction algorithm based on generative adversarial network is improved to enhance the texture details of remote sensing images with ResNet34 and CNN as the basis, so as to be able to extract more rich features from images to extract richer features, and the effects of the degradation model and super-resolution reconstruction model are verified on the dataset UC Merced and the real remote sensing dataset Alsat2B. By means of software processing, the resolution of the finished commercial remote sensing images is effectively enhanced, providing higher resolution images for tasks such as disaster prediction and target detection.

## Acknowledgments

This paper is supported by the Academic Research Projects of Beijing Union University (No. ZK10202205).

## References

- [1] DONG C, LOY C C, HE K, et al. Learning a Deep Convolutional Network for Image Super-Resolution [C]// LNCS 8692: Proceedings of the 13th European Conference on Computer Vision, Zurich, Sep 6-12, 2014. Cham: Springer, 2014: 184-199.
- [2] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 1646-1654.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 770-778.
- [4] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Networks[C]//Proceedings of the Annual Conference on Neural Information Processing Systems 2014, Montreal, Dec 8-13, 2014. Red Hook: Curran Associates, 2014: 2672-2680.
- [5] LEDIG C, THEIS L, HUSZAR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4681-4690.

- [6] VU T, LUU T M, YOO C D. Perception-enhanced image super-resolution via relativistic generative adversarial networks [C]// LNCS 11133: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 98-113.
- [7] ZHANG K, LIANG J Y, GOOL L V, et al. Blind Image Super-Resolution[EB/OL].(2021-03-25)[2021-07-06]. <https://arxiv.org/abs/2103.14006>.
- [8] GAO G, ZHANG J, LV X M, et al. Analysis of Multiplicative Noise Models in SAR Imagery[J]. Signal Processing, 2008, 24(2): 161-167.
- [9] SEFI B K, ASSAF S, MICHAL I. Blind super-resolution kernel estimation using an internal-gan[C]// Advances in Neural Information Processing Systems. 2019: 284-293.
- [10] GU J J, LU H N, ZUO W M, et al. Blind Super-Resolution With Iterative Kernel Correction[C]// Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2019
- [11] ZHANG N, WANG Y C, ZHANG X, et al. A Multi-Degradation Aided Method for Unsupervised Remote Sensing Image Super-Resolution With Convolution Neural Networks [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020.
- [12] ZHU X, TALEBI H, SHI X W, et al. Super-resolution commercial satellite imagery using realistic training data[C]// International Conference on Image Processing. IEEE, 2020: 498-502.