# Edge Loss for Remote Sensing Image Super-Resolution

Jiaoyue LI [a], Weifeng LIU [b], Kai ZHANG [c] and Baodi LIU [b,1]

[a] *College of Oceanography and Space Informatics, China University of Petroleum, China*

[b] *College of Control Science and Engineering, China University of Petroleum, China*

[c] *School of Petroleum Engineering, China University of Petroleum, China*

**Abstract.** Remote sensing image super-resolution (SR) plays an essential role in many remote sensing applications. Recently, remote sensing image super-resolution methods based on deep learning have shown remarkable performance. However, directly utilizing the deep learning methods becomes helpless to recover the remote sensing images with a large number of complex objectives or scene. So we propose an edge-based dense connection generative adversarial network (SREDGAN), which minimizes the edge differences between the generated image and its corresponding ground truth. Experimental results on NWPU-VHR-10 and UCAS-AOD datasets demonstrate that our method improves 1.92 and 0.045 in PSNR and SSIM compared with SRGAN, respectively.

**Keywords.** Image super-resolution, remote sensing image, edge loss, generative adversarial network

## 1. Introduction

Image SR is to repair a high resolution (HR) image from its low resolution (LR) counterpart. Remote sensing images often have a low spatial resolution but have high requirements for the clarity of small objects in the vision, so using image super-resolution to process remote sensing images has immense application value.

In the past, quantities dictionary-based algorithms had been applied to super-resolution of remote sensing images, which was usually troublesome to capture sufficient deep features to restore satisfactory images.

The high-frequency information is very beneficial for restoring high-quality images. Fortunately, the appearance of residual structure makes it possible to obtain deep features using deep learning-based approaches. Various deep learning-based methods have emerged at the historic moment, and they can successfully seize much high-frequency information to repair the image. In particular, the Generative Adversarial Network [1] is a very successful deep network model and it is trained through adversarial methods to learn the distribution of real data, and the training purpose of generator and discriminator is exactly the opposite. The generated image confirms that a good perception effect can

---

[1] Corresponding Author: E-mail: thu.liubaodi@gmail.com.
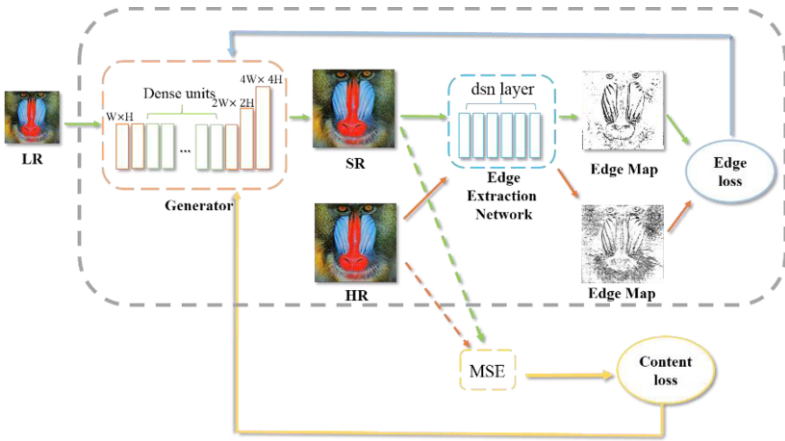
**Figure 1.** Overview of SREDGAN. First, the images generated by the generator and the ground truth images are input into an edge extraction network. Second, the edge maps created by the two kinds of images are differentiated to obtain edge loss. Finally, we propagate the edge loss to the generator and minimize the loss function to make it show solicitude for edges.

be obtained by using a generative confrontation network, but there is often a common phenomenon that the edges are blurred and the artifacts are in evidence.

Superb high-resolution images with sharp edges and realistic shapes, moreover, the edge information contained in remote sensing images is more abundant, and it becomes more important to increase the attention to the edge information acquisition during the training process. Therefore, we pose a densely connected edge-based generative adversarial network for image super-resolution (SREDGAN). The generator of SREDGAN draws support from a densely connected network to extract multiple features, and the network's training is adjusted continuously by a hybrid loss.

In general, the main contributions of our work are summarized in the following three points:

- Take edge maps into account, which is of great help to the sharpness of images, and replacing traditional edge extraction with pre-trained edge extraction networks.
- Designed an end-to-end network SREDGAN that is easy to train.
- Prove that our approach has achieved excellent results by comparing it with other advanced SR methods.

## 2. Approaches

We advance the SREDGAN model to generate a high-quality image with distinct edges and neural contents. First of all, input the LR image $I^{LR}$ into generator G and output the SR image $I^{SR}$. Then, the $I^{SR}$ and high-resolution images $I^{HR}$ are sent to the discriminator D to distinguish the similarity between $I^{SR}$ and $I^{HR}$. Our purpose is to find the appropriate loss function to make the generated images real. An overview of the generator of the proposed model is showcased in Figure 1.

For the most SR task, minimize the pixel-level and perceptual-level differences between $I^{SR}$ and $I^{HR}$ is widely adopted. However, it leads to uneven edges inevitably. In
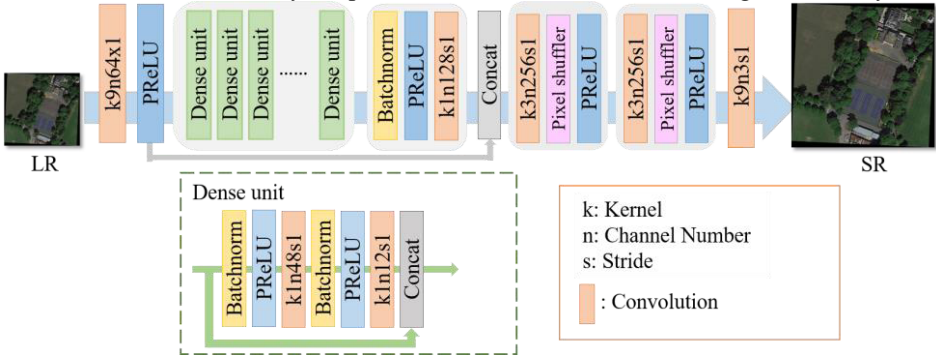


**Figure 2.** Structure of the generator which has five parts: shallow feature extraction, deep feature extraction, parameter transition, upsampling, image reconstruction.

this paper, we intend to add an edge loss to relieve the problem. The whole model is explained in Section 2.1, Section 2.2 plains describes our proposed edge loss and other functions of SREDGAN.

## 2.1. Structure of SREDGAN

Our generator is divided into five parts, the input image goes through a convolutional layer to get shallow features, and then the deep feature is extracted by sixteen dense units with a growth rate at 12. After several dense units, the parameters increases rapidly, which would bring substantial computational costs to the subsequent processing. Therefore, a $1 \times 1$ convolution is added to decrease parameters. To help gradient propagate, a long skip connection joins shallow features to deep features. Sub-pixel convolution [2] is a method to enlarge image size. It combines individual pixels on a multi-channel feature into a unit on a feature; the pixels on each feature are equivalent to sub-pixels on the reconstructed feature. We use two sub-pixel convolution layers to upscale the length and width of the input images by four times in our network. At last, a convolution implements the reconstruction of HR images. Our generator network is pictured in Figure 2.

Discriminator aims to identify which image is true and which one is false; it guides the generator to show solicitude for one direction for training, so a suitable discriminator is also very essential. SR intends to restore realist and high-resolution images. Therefore, it requires more high-frequency features. VGG [3] can extract deep features, and its structure is simple, so it is a pretty choice to use the VGG network as the discriminator framework, which is the similar to the discriminator of SRGAN [4].

## 2.2. Design of Loss Functions

To solve the situation that the image edges are indistinct, and the contents do not conform to the real, we construct an edge loss. The purpose of edge loss is to decrease the edge discrepancies between SR images and HR images. Although training with L1 or L2 loss will get a pleasant effect under objective criteria, visual results obtained are not ideal. The network does not put more emphasis on learning edge distribution, so the unreasonable contents and artificial scene frequently arise, and the edges are irrational

and unequivocal. Edge loss harmer at making object edges authentic and distinguishable. It makes the edges been paid close attention to network training.

**Table 1.** Results for test datasets, the data from label 1 to label 5 is the result of UCAS-AOD, the data from label 6 to label 10 is the result of NWPU-VHR-10, and the last column of the table represents the average.

| No. | Bicubic PSNR/SSIM | SRResNet PSNR/SSIM | SRGAN PSNR/SSIM | SRGAN2 PSNR/SSIM | SREDGAN PSNR/SSIM |
|---|---|---|---|---|---|
| (1) | 25.45/0.779 | 25.81/0.817 | 23.64/0.726 | 25.22/0.779 | **26.01/0.821** |
| (2) | 25.46/0.905 | 27.79/0.906 | 25.42/0.821 | 27.12/0.863 | **27.98/0.909** |
| (3) | 26.86/0.825 | 29.08/0.851 | 27.49/0.816 | 28.22/0.830 | **29.27/0.853** |
| (4) | 28.74/0.850 | 31.93/0.912 | 29.55/0.875 | 30.30/0.879 | **32.16/0.913** |
| (5) | 27.67/0.837 | 30.98/0.839 | 29.16/0.811 | 29.80/0.809 | **30.68/0.838** |
| (6) | 25.79/0.677 | 26.96/0.783 | 25.87/0.740 | 25.66/0.746 | **27.23/0.781** |
| (7) | 23.70/0.637 | 26.26/0.761 | 24.47/0.698 | 24.07/0.686 | **26.27/0.705** |
| (8) | 24.55/0.585 | 31.44/0.939 | 30.33/0.897 | 30.68/0.896 | **31.08/0.919** |
| (9) | 25.90/0.575 | 26.91/0.738 | 25.41/0.693 | 25.78/0.701 | **26.63/0.726** |
| (10) | 27.90/0.573 | 28.97/0.768 | 27.83/0.756 | 28.53/0.766 | **30.97/0.819** |
| AVE | 26.20/0.724 | 28.61/0.831 | 26.91/0.783 | 27.54/0.796 | **28.83/0.828** |

Nowadays, there are amounts of methods about edge extraction, although non-deep learning ways are easy, the kind of edge it extracts is too single. The salient disadvantage of the traditional methods is that their convolution kernels are unique and usually sensitive to a specific type of edge. HED [5] is a product of deep learning, and it has six outputs dsn1, dsn2, dsn3, dsn4, dsn5, and dsn-fuse, the first five represent edge detection maps of different feature layers respectively, and the last one indicates the fusion of the first five outputs. In our works, we have modified some convolution settings. We observe that the edges from the dsn1 layer is the most comprehensive. With the increase of network depth, the rougher the edge is, because HED [5] aims to detect the outline of an object, not the tiny edges. However, subtle edges are what we demand in our task, so we choose the dsn1 layer as a feature extraction layer. We input the generated images and HR counterparts into the HED network and minimize their feature discrepancies in the dsn1 layer. We named the error edge loss, and Eq. (1) shows the details.

$$L_{edge} = \frac{1}{WH} \sum_{i=1}^{C} \|h_i(I^{HR}) - h_i(I^{SR})\|_2 \tag{1}$$

Where $h_i(I^{HR})$ represents the dsn1 layer feature with the input of HR image, and $h_i(I^{SR})$ expresses the feature from SR image. Similar to the method of perceptual loss [6], the parameters of the HED network are trained in advance.

To ensure that the color of the generated image is as realistic as possible, we also take advantage of the common L2 loss and adversarial loss.

## 3. Experiments

### 3.1. Experimental Settings

We train our models on the RAISE dataset [7], which contains 8,156 common high-resolution images, and we get their low-resolution version pairs by Bicubic interpolation. Since large amount of training data is powerful to SR tasks, we flip images randomly and crop the size of HR images to 96×96 as well as low-resolution images to 24×24. We adopt remote sensing image datasets, NWPU-VHR-10 [8] and UCAS-AOD [9] as test datasets.

|  Bicubic | SRResnet | SRGAN | SRGAN2 | SREDGAN | HR |
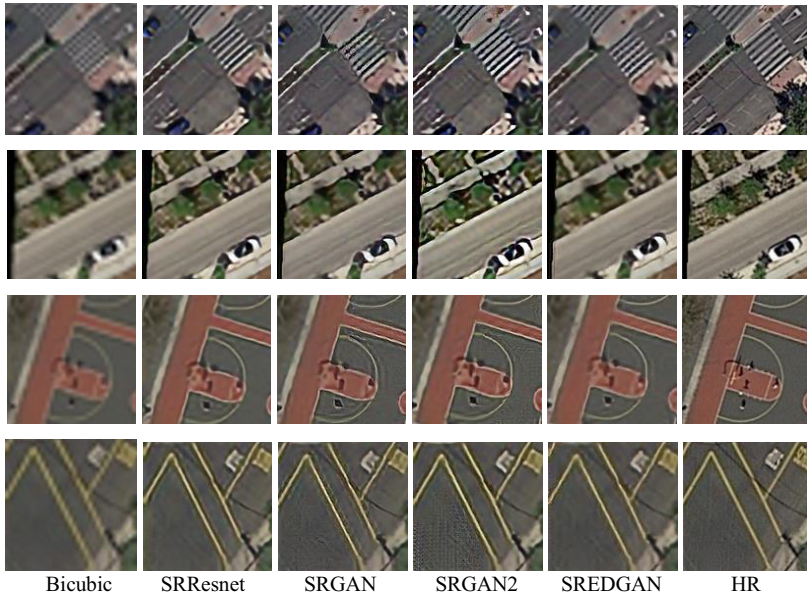
**Figure 3.** Result of the different methods on NWPU VHR-10.

Adam is chosen for optimizing training, and we set $\beta_1$ to 0.9 and set $\beta_2$ to 0.99. For optimum result, we assign the weight of content loss (L2 loss), adversarial loss and edge loss to 1, 0.0001 and 0.001, the weighted sum of these three losses constructs the total generator loss. The learning rate for the first 100000 iterations is set to $10^{-4}$, and its remaining 100000 iterations are tenfold. We employ Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) as the criterion for testing.

## 3.2. Experimental Settings

We compare our method with Bicubic, SRResnet [4], SRGAN [4], SRGAN2 and our SREDGAN, and SRGAN2 is SRGAN with our edge loss. Table 1 and Figure 3 show the evaluation results and generated images of the two test datasets.

We can find the images generated by Bicubic are greatly blurry. Although the images generated by SRResnet have achieved excellent results in Table 1, the image is still not clear enough. Under the evaluation criteria of PSNR and SSIM, the results of SRGAN and SRGAN2 are not very satisfactory, but they have a good observation effect. Synthetically, the images generated by the generative adversarial network have higher resolution than the non-GAN network, but their results on PSNR and SSIM are slightly lower than those of non-GAN network.

In order to verify that increasing the edge loss helps the network to restore the image, we compared SRGAN and SRGAN2. The experimental results show that the results of SRGAN2 are better than those of SRGAN, whether in objective indicators or direct observation. The experimental comparison results of SRGAN and SREDGAN show that our generator model is more suitable for image feature extraction.

Comparing SRResnet and SREDGAN, we found that under the action of network structure and loss function, the PSNR and SSIM values of SREDGAN are close to the results of SRResnet, which shows that extracting rich features and paying more attention to the edges of the image will double the clarity and accuracy of the image.

## 4. Conclusion

In this paper, we propose a new model SREDGAN that pays more attention to edge information, and we adopt advanced edge extraction method to embed into an end-to-end network. With edge loss, models can generate an image whose edge is clear and close to the ground truth. Extensive experimental results demonstrate the effectiveness of our proposed methods for remote sensing super-resolution.

## Acknowledgements

## References

[1]  Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. Proceedings of the Advances in Neural Information Processing System;

[2]  Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016 1874-1883.

[3]  Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[4]  Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017 4681-4690.

[5]  Xie S, Tu Z. Holistically-nested edge detection. Proceedings of the IEEE international conference on computer vision; 2015 1395-1403.

[6]  Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. European conference on computer vision. 2016; 694-711.

[7]  Dang-Nguyen DT, Pasquini C, Conotter V, Boato G. Raise: A raw images dataset for digital image forensics. Proceedings of the 6th ACM multimedia systems conference; 2015: 219-224.

[8]  Cheng G, Zhou P, Han J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing. 2016; 54(12): 7405-7415.

[9]  Xia GS, Bai X, Ding J, Zhu Z, Belongie S, Luo J, Datcu M, Pelillo M, Zhang L. DOTA: A large-scale dataset for object detection in aerial images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 3974-3983.