Multi-channel Feature Extraction and Super-Resolution Reconstruction of Remote Sensing Images

Zheng Zhang\textsuperscript{1,a}, Jiabin Zhang\textsuperscript{2,b}, and Chang’an Liu\textsuperscript{1,*}
\textsuperscript{1} North China University of Technology, Beijing, 100144, China
\textsuperscript{2} North China University of Technology, Beijing, 100144, China
\textsuperscript{a}email: zhangzheng@ncut.edu.cn, \textsuperscript{b}email: Jabari.z@outlook.com
\textsuperscript{*}Corresponding author’s e-mail: 2019312120119@mail.ncut.edu.cn

Abstract: Super-resolution reconstruction is an imaging method to improve image resolution. It refers to reconstruct a clear high-resolution image from a low-resolution image. High-resolution remote sensing images can provide more detailed information and higher density, but in the field of remote sensing, because of the limitation of the hardware and vast distances, the remote sensing images are fuzzy sometimes. To facilitate subsequent tasks, this paper proposes a multi-channel feature extraction generative adversarial remote sensing image reconstruction method. According to the characteristics of remote sensing image, a generator is designed, which adds Laplace operator to enhance the edge information of the image, and uses multi-channel feature extraction, which not only enhances the ability of feature extraction but also reduces the number of parameters. In this paper, the super-resolution reconstruction task is carried out based on the 2X magnification factor, and the experimental results are evaluated on SET5/14 and NWPU-RESISC45 dataset. The experimental results show that the images generated by this method have a higher detailed texture and better super-resolution reconstruction effect of remote sensing images.

1. Introduction
With recent advancements in space technology, now remote sensing technology has been used all over the place in our side. The resolution of remote sensing images is the ultimate basic attribute of image. It is important in analyzing and utilization image content information. High-resolution images have more texture information and higher density information compared with low-resolution images. Super-resolution reconstruction, which is a basic visual task, refers to the process of recovering high-resolution (HR) images from low-resolution (LR) images \cite{1}. Generally, improving resolution method is that, one is the physical method, which improves the various parameters of the photographic equipment, the other one is a software approach, which super-resolution reconstruction. In the field of remote sensing due to the limitations of hardware and ultra-long distance, the remote sensing images obtained are slightly blurry. The most common way to improve the resolution size is to use best imaging sensors to achieve it. However, it is difficult to update the remote sensing imaging equipment on the satellite. Therefore, using the existing low resolution to obtain more high-frequency information makes it possible to improve the resolution of observation images without changing the satellite detection system itself. High-resolution remote sensing image compared with low-resolution remote sensing image, the image detail is better, and has more advantageous to the analysis of the remote sensing data in the future\cite{2}. Therefore, image super-resolution reconstruction to improve remote sensing image has great potential
and application value in the remote sensing information domain.

Super-resolution reconstruction can be generally divided into three methods. They are interpolation method, reconstruction method and learning method. However, there are the following problems: 1) the edges of the reconstructed image are blurred, and it is difficult to identify small objects; 2) with the increase of magnification, the image details and reality are lost; 3) higher training requirements, it requires prior information. Spatial information of remote sensing images was reflected through image pixels values on the space changes. It included a practical significance on the point, line, surface. They belonged to spatial information and texture information, which in the image edge due to it contains abundant information. Therefore, it is an essential attribute for image characteristics in image recognition. The brightness distribution of remote sensing images is more uniform, among which there are more edge breakpoints and more edge information loss of high frequency. The image details and edge gain benefit after sharpening. Edge information can improve the performance of objection detection, image segmentation, edge extraction and other tasks[19].

Aiming at the drawback of current super-resolution methods in remote sensing images, such as single background texture, fuzzy edge, and network generalization performance. A new SR network was proposed in this paper according to the inherent characteristics of remote sensing images. The main contributions of this paper are as follows: an improved super-resolution reconstruction network of remote sensing images based on GANs network [4] is proposed. The Laplacian operator is added to the generator to enhance the sensitivity of target edge information in the image. The multi-channel branch structure is used to add residual connection as feature extraction module, followed by feature optimization module to extract and purify the image. The CGAN [5] training mode is introduced, the low-resolution image is used as the input of the generator, and WGAN is used to optimize the network to solve the original GAN training problem.

2. Materials and Methods
The main purpose of this paper is to improve the edge information of remote sensing images in the reconstruction process, make the feature information more obvious, and better perform the detection and classification tasks. The goal of SISR is to estimate the SR image based on the input LR image. LR is obtained by down-sampling the HR. The final goal of this article is to obtain a mapping f(x) to estimate the corresponding LR input image between HR images. In order to achieve the goal, in this paper, a generative adversarial network structure including generator and discriminator was designed. In this section, firstly, we describe the network structure, and then discuss the generator and loss function.

2.1 Network Introduce
A generator network and a discriminator network are defined according to the architectural method of the generative adversarial network. The overall idea is that it trains a generator (G) which the purpose of the generative model is to deceive the discriminator (D) with discriminative ability. The discriminator is trained to distinguish super-resolution images from real images. In this way, the G can learn to generate a solution space that is highly similar to the real image, so it is difficult to be classified by D. This encourages perceptually better solutions located in the subspace of the natural image in the manifold. There is an obviously contrast to the SR solution get by minimizing the value of pixel-by-pixel error, such as MSE. Because the generative adversarial network is extremely uncontrollable, it is difficult to control the output. To solve this problem, CGAN[5](conditional confrontation generation) is introduced.

2.2 Network Architecture
As shown in Figure 1. the generator network consists of three parts, a feature extraction block, a reconstruction block, and an image optimization block. The design of the generator has the following points: a) use Laplace transforms to process the input image, b) short-connect to the reconstruction block, c) remove all BN layers, d) use the multi-path feature extraction, and e) add the image optimization block.
In this paper, the feature extraction block is designed according to the idea of residual connection, and the regularization Batch Normalization layer is removed. The BN layer usually reduces the computational complexity, but the effect is different in different visual tasks. The BN layer is similar to a kind of contrast stretching for the image. After any image passes through the BN layer, its color distribution will be normalized, and the original contrast information of the image will be destroyed, at the same time will affect the quality of the image, so the Batch Normalization layer was removed in the super-resolution reconstruction process [17]. The following Figure 1 shows the structure of the feature extraction block.

![Network Architecture](image)

**Figure 1. Network Architecture (The top is generator, the bottom is discrimination)**

The feature extraction block plays an important role in the image super-resolution reconstruction model. The more features were extracted, the better the reconstruction effect the model got. As shown in Figure 2, the generator stacks 16 feature extraction block. Each feature extraction block contains three branches, that is a short connection, a combination of 1x3 and 3x1 convolutions, and a 1x1 combination. Extraction of images deep feature information of the given image is the feature extraction block’s purpose. The use of 1x1 convolution can effectively reduce the number of parameters and can flow information across channels. The most important thing is to extract effective information. Using a combination of 1x3 and 3x1 convolution is equivalent to 3x3 convolution, but after splitting, it is conducive to the extraction of side information. Short connections can make rational use of information at all levels. Using linear interpolation can turn a low-resolution image into a high-resolution image. The resulting image is the same size as the one obtained after a super-resolution network is trained, but this is an operation that consumes computing, power and memory. In order to solve this problem, people began to study up-sample technology. FSRCNN[18] uses the transposed convolution layer as the last layer of the neural network to reconstruct the image. Although this method can improve the efficiency of training, its performance is not well in SISR. In this article, the up-sampling uses Shuffle Pixel[18]. Finally, an image optimization block was added to optimize the generated image to reduce the difference between it and the real image. It contains five cascaded convolutional layers, and each convolutional layer is activated by the function. After this block image’s characteristic information was strengthened.

The network receives low-resolution as input and passes through three blocks to high-resolution.

\[
FM = FEB(LR) \\
out = ConstructB(FM)
\]
\[ FRB = \text{conv}(out) \]  
\[ HR = (FRB(FRB(FRB(FRB(out)))))) \]  

The main function of discrimination is to judge the quality of the generated model. The output of the generator is the reconstructed remote sensing image. The output and the original image are used as input and enter the discriminator at the same time. After the discriminator network, the similarity of the two pictures is finally predicted, the result is fed back to the generator. And then the generator adjusts the parameter.

2.3 Loss Function

For the super-resolution reconstruction of remote sensing images, image characteristics and the network structure designed in this paper are under consideration. In the initial training generator stage, Mean Square Error (MSE) was used as the loss function.

\[ l(\theta) = \frac{1}{2N} \sum_{j=1}^{N} \| f(I_{LR}^j) - I_{HR}^j \|_2^2 \]  

In order to reduce the visual difference between the real image and the super-resolution image, we have added an adversarial loss function to the objective loss function, which will force the generator to generate a super-resolution image that is sufficiently similar to the real image to deceive the discriminator. It is defined as follows:

\[ L_{adv} = \frac{1}{N} \sum_{j=1}^{N} \log \left( 1 - D \left( G(I_{LR}^j) \right) \right) + \log D(I_{HR}^j) \]  

3. Experiment & Discussion

3.1 Dataset

The dataset uses the DIV2K dataset [18] and NWPU-RESISC45 as the training data set of the model in this paper. The DIV2K dataset contains 800 training images, 100 verification images, and 100 test images at different scales. Each LR image is crop into a 96x96 image block from the HR image and they were sent to training by horizontal flip, vertical flip, rotation, and other data augmentation methods under transform to obtain more training data. NWPU-RESISC45 is a high-resolution remote sensing map dataset, including 45 scene categories such as airports, overpasses, railway stations, residential areas, etc. There are 700 images in every category. To ensuring the authenticity and diversity of experimental data, 10 images of each category in the original data set were selected, composing a total of 450 images form the experimental dataset.

In the evolution phase, there are two common benchmark datasets in SR, such as Set5, Set14. We choose 2x magnification factor, and the LR images are generated in the same way as during training. And the remote sensing image on Google Earth was choose by us to test and evaluation in the experiment.

3.2 Result

The Loss and PSNR situation during training and verification is shown in Figure 3. To verify the performance of the proposed method, we compare our model’s performance with that of other method.
Table 1: Experimental results of the different super-resolution algorithms on the Set5 and Set14 dataset

| Method     | Set5 PSNR | Set5 SSIM | Set14 PSNR | Set14 SSIM |
|------------|-----------|-----------|------------|------------|
| nearest    | 26.26     | 0.7552    | 24.64      | 0.7100     |
| bicubic    | 28.43     | 0.8211    | 25.99      | 0.7486     |
| SRCNN      | 30.07     | 0.8627    | 27.18      | 0.7861     |
| SRGAN      | 29.40     | 0.8472    | 26.02      | 0.7397     |
| Ours       | 30.79     | 0.8631    | 27.79      | 0.8182     |

Table 2: Ablation study on the components of network. (Join Laplace operator)

| Component | PSNR | SSIM  |
|-----------|------|-------|
| None      | 30.79| 0.8631|
| Add Laplace| 31.64| 0.8712|

4. Conclusions

In this paper, a multi-channel feature extraction generator is designed to effectively extract the feature information of remote sensing images. Through the design of the generator network and adding the Laplacian operator, the super-resolution image by the method in this paper obtained a higher PSNR value and SSIM. The image details are closer to the original high-resolution image. Experiments show that the method achieves an excellent trade-off between visual effect and computational efficiency in super-resolution reconstruction remote sensing images. The super-resolution remote sensing image obtained can be used for other image tasks. It can improve the learning ability of the model and edge information of the feature space, so that can improve recognition accuracy and make best use of the image information.
Acknowledgments
Project fund: North China University of Technology Research Start-up Funds And The Fundamental Research Fund of Beijing Municipal Education Commission.

References
[1]. Wang Z, J Chen, Hoi S C H. (2019) Deep Learning for Image Super-resolution: A Survey.
[2]. Zou B , Xu X , Zhang L . (2020) Object-Based Classification of PolSAR Images Based on Spatial and Semantic Features[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, PP(99):1-1
[3]. Sun H H . (2020) The Analysis and Application of Image Edge Extraction in the Joint Transform Correlation Recognition[J]. Applied Physics, 2020, 10(1):8-14..
[4]. Goodfellow et al., (2014) Generative adversarial nets, in Proc. Adv. Neural Inf. Process. Syst., Montreal, QC, Canada, Dec, pp. 2672–2680.
[5]. Mehdi, M and Osindero, S.(2014) Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[6]. Huang T S . (1984) Multi-frame image restoration and registration[J]. Computer Vision & Image Processing, 1.
[7]. Shi W , Caballero J , F Huszár, et al. (2016) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[J]. IEEE..
[8]. Mao X J , Shen C , Yang Y B . (2016) Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections[J].
[9]. Tong T , Li G , Liu X , et al. (2017) Image Super-Resolution Using Dense Skip Connections[C]// IEEE International Conference on Computer Vision. IEEE Computer Society.
[10].Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. (2012) Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25: 1097-1105.
[11].Kim J , Lee J K , Lee K M . (2016) Deeply-Recursive Convolutional Network for Image Super-Resolution[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[12].He, K, X. Zhang, S. Ren, (2016) Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition.
[13].He K , Zhang X , Ren S , et al.(2016) Identity Mappings in Deep Residual Networks[C]// European Conference on Computer Vision. Springer, Cham.
[14].Radford A, Metz L, Chintala S .(2015) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer ence.
[15].Lim B, Son S , Kim H , et al.(2017) Enhanced Deep Residual Networks for Single Image Super-Resolution[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE.
[16].Nah S, Kim T H, (2016) Lee K M . Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring[C]// IEEE Computer Society. IEEE Computer Society.
[17].Agustsson E and Timofte R, (2017) Ntire 2017 challenge on single image super-resolution: Dataset and study. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.
[18].Shi Y and Caballero J, (2016) 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.
[19].Dong L , Xing L , Liu T , et al. (2019) Very High Resolution Remote Sensing Imagery Classification Using a Fusion of Random Forest and Deep Learning Technique—Subtropical Area for Example[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, PP(99):1-16..