Image Super Resolution Reconstruction Algorithm Based on Convolution Neural Network

Feng Zhao and Zhifeng Hu
Modern Education Technology Center, Wuhan Business University, Wuhan, Hubei, China
Email: zhaofeng2015@wbu.edu.cn; hzf@wbu.edu.cn

Abstract. Aiming at the characteristics of large amount of remote sensing image data, large terrain fluctuations and wide coverage, this paper proposes a method for super-resolution reconstruction of remote sensing image. This method combines dense network and deep back projection network to form dense projection. The unit forms a deep dense projection network, which solves the problems of insufficient texture, loss of image details and training difficulty in the super-resolution reconstruction of remote sensing image. The experimental results show that on multiple remote sensing image datasets, compared other traditional methods, the PSNR and SSIM are significantly improved, and reconstructed remote sensing image texture signs and details are more abundant.

Keywords. Deep back projection network; super-resolution reconstruction; convolutional neural network.

1. Introduction
At present, the research hotspots in the field of image processing mainly focus on super-resolution reconstruction technology [1, 2]. High-resolution remote sensing images can better perform image feature extraction, image segmentation, and target tracking research, improving the quality of research. The super-resolution technology to reconstruct images can increase the resolution of remote sensing images without increasing the cost, thereby obtaining the required information.

At this stage, the super-resolution reconstruction technology is mainly achieved through three mainstream methods: interpolation method, reconstruction method and learning method. The classical interpolation methods include nearest neighbor interpolation, bilinear interpolation [3] and bicubic interpolation [4]. These methods are relatively simple, but the details of reconstruction are limited. The classical reconstruction methods include iterative back projection method [5] and convex set projection method [6]. This method not only ensures the accuracy of image reconstruction, but also enhances the application range. However, the reconstructed image is easy to produce smooth, fuzzy and other problems. With the optimization of parallel computing, data processing is more rapid in the field of artificial intelligence, and deep learning benefits a lot. In 2017, Lim et al. [7] proposed an enhanced depth super-resolution network EDSR, which expanded the size of the model to improve the quality of the results, but its model complexity is too high, which increases the difficulty of training.

Remote sensing images are compared with ordinary images, which have the characteristics of large amount of data, large topographic relief and wide coverage [8-10]. The above methods have achieved good reconstruction results, but there are still some problems, such as insufficient texture representation and training difficulties. To solve the problem, designed a super-resolution reconstruction method based on DenseNet [11] and depth back projection network [12] that greatly
reduced the use of parameters, solved the training difficulties, overcome the problem of insufficient
details extraction of remote sensing image, and finally improved the training and reconstruction
accuracy.

2. Related Theories

2.1. Deep Back Projection Network
The deep back projection network includes two core modules: an upper projection unit and a lower
projection unit.

2.1.1. Upper Projection Unit. Upper projection unit: As shown in figure 1, firstly deconvolve the input
low-resolution image can get the high-resolution feature map, then convolve a and then down-sample
to obtain low-resolution feature map Then, perform a residual operation on and the input image, add
the residual value and perform deconvolution to obtain a high-resolution feature map, and finally add
and to obtain the final high-resolution feature map.

2.1.2. Down Projection Unit. Down projection unit: As shown in figure 2, first input the high-
resolution image A to convolve to obtain the feature map B, then deconvolve B to obtain the feature
map C, and then perform residual operation on C and the input image A Then, the residual value D is
added and convolution is performed to obtain a feature map E. Finally, F and E are added to obtain a
final low-resolution feature map. The projection unit can be used to provide the projection error to the
sampling layer, and iteratively produce better solutions, which can extract more clear and rich details,
but the sampling method may lose some image information.

3. Methodology

3.1. Network Model
Aiming at the problems of texture characterization is not detailed enough and missing details, and
difficult training in the current algorithm in the field of remote sensing image reconstruction, this
paper combines dense network [11] and deep back-projection network [12] to design a remote sensing
network, which is based on convolutional neural. Image super-resolution reconstruction method. The
network model of this method includes three stages of feature extraction, back projection and
reconstruction, as shown in figure 3.
Feature extraction: using two $3 \times 3$ and $1 \times 1$ convolutional layers, first use the $3 \times 3$ convolutional layer to build the initial low-resolution feature map, and then implement the dimension reduction through the $1 \times 1$ convolutional layer.

Back projection: As shown in figure 4, in order to avoid the loss of image information, the improved projection unit used in this paper uses a $1 \times 1$ convolution layer for feature pooling. Assuming T stages, there are T improved upper projection units and T-1 improved lower projection units. After the projection units, the high-resolution feature map generated by each improved upper projection unit is processed, cascade.

Reconstruction: reconstruct and enlarge the cascaded high-resolution feature map, and then use a $3 \times 3$ convolution to reconstruct the feature map, and finally obtain a high-resolution image.

3.2. Loss Function
In the current super-resolution reconstruction researches, most of them use the Mean Square Error (MSE) method as the loss function for network training. This paper still chooses it as a loss function. For a series of remote sensing images $F^i(Y; \theta)$ and the original series of high-resolution remote sensing images $X^i$, in order to minimize the mean square error, the optimal parameter $\theta$ is obtained. The mean square error formula is shown in equation (1):

$$L(\theta) = \frac{1}{n} \sum_{i}^{n} \| F^i(Y; \theta) - X^i \|^2$$

(1)

The fitting effect of the model parameters depends on the size of the $L(\theta)$ value, the closer to 0 the better the effect, where n is the number of samples.
4. Experiments and Results

4.1. Data Set
We divided the high-resolution remote sensing images of a city in southern China in 2015 provided by the CCF Big Data Competition into a training set and test set 1 of this article according to 7:3 ratio. This paper also uses two actual captured remote sensing image data sets as test Set 2 and Test Set 3. Test Set 2 and Test Set 3 were obtained by the Beijing Key Laboratory of Digital Plants, and the multi-rotor drone DJI M600Pro was deployed at the corn breeding base of Beijing Academy of Agriculture and Forestry, Nanfan Base, Sanzhou Yazhou District Independent research and development of high-throughput 3D laser scanning imaging phenotypic measurement system Croplidar (integrated sensors include: visible light camera Alpha 5100 APS-C format, 24.3 million effective pixels; Micasense multispectral camera RedEdge-MX; 3D lidar Velodyne VLP-16) Orthophotos, point clouds, and canopy spectra of the corn population at the jointing stage. The data set before partial cropping is shown in figure 5.

![Figure 5](image)

Figure 5. Two times larger rendering on test 1 with different super-resolution reconstruction algorithms.

4.2. Evaluation Index
At present, subjective evaluation and objective evaluation are the main evaluation standards for the existing image super-resolution reconstruction. The subjective evaluation is the evaluation of the image by the human eye. At present, the objective quantification methods mainly include: peak signal-to-noise ratio (PSNR) and mean square error MSE.

\[
MSE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i, j) - Y(i, j))^2
\]

\[
PSNR = 10 \log_{10} \left( \frac{2^n - 1}{MSE} \right)
\]

Among them, MSE is equal to the mean square difference between image X and contrast image Y; \(X(i, j)\) represents the pixel value of the current image, \(Y(i, j)\) represents the pixel value of the contrast image, H represents the height of the image, and W represents the width of the image. The higher the PSNR, the closer the quality of the reconstructed image to the quality of the high-resolution image, the better the reconstruction effect.

Structural similarity (SSIM) is another image quality evaluation index. It is mainly to compare the similarity of the image, and take values from the aspects of contrast and structure. The distortion of the image depends on the value of the value, and the range is generally [0,1]. The definition is as follows:
\[ SSIM(X,Y) = \frac{l(X,Y)c(X,Y)s(X,Y)}{l(X,Y)c(X,Y)s(X,Y)} \]  

(4)

\( l(X,Y) \) represents the brightness of the image, \( c(X,Y) \) represents the image contrast, and \( s(X,Y) \) represents the structural image similarity. The closer the SSIM value is to 1, the higher the similarity between the two pictures, the better the reconstruction effect.

4.3. Comparative Analysis

In order to verify the effect of the method proposed in this article on the remote sensing image reconstruction, we include traditional super-resolution methods by comparing algorithms (A + [11], Bicubic [8]) and learning-based super-resolution methods (SRCNN [13], DRCN [14], EDSR [7]). These methods were used to reconstruct remote sensing images with scale factors of 2x, 3x and 4x, respectively. The experiments were performed in the same experimental environment. Comparative experiments were performed on the three test sets Test1, Test2, and Test3. The experimental results are shown in Table 1.

**Table 1.** Average PSNR/SSIM of test sets Test1, Test2, and Test3 under different super-resolution reconstruction algorithms.

| Data set | Scale factor | Bicubic   | A+        | SRCNN    | DRCN     | EDSR    | Our method         |
|----------|--------------|-----------|-----------|----------|----------|---------|-------------------|
| Test1    | 2            | 37.69/0.896 | 38.26/0.925 | 39.31/0.928 | 40.29/0.951 | 40.55/0.961 | 40.75/0.964 |
|          | 3            | 34.26/0.853 | 34.67/0.881 | 35.78/0.887 | 36.81/0.918 | 37.05/0.929 | 37.26/0.930 |
|          | 4            | 33.17/0.815 | 33.54/0.847 | 34.69/0.853 | 35.61/0.876 | 35.98/0.884 | 36.17/0.897 |
| Test2    | 2            | 20.68/0.735 | 21.41/0.773 | 22.30/0.777 | 23.20/0.798 | 23.45/0.809 | 23.68/0.813 |
|          | 3            | 19.97/0.708 | 20.54/0.741 | 21.10/0.748 | 22.13/0.764 | 22.31/0.775 | 22.56/0.781 |
|          | 4            | 21.67/0.703 | 23.01/0.710 | 23.86/0.714 | 24.79/0.736 | 25.02/0.743 | 25.21/0.752 |
| Test3    | 2            | 19.05/0.662 | 20.75/0.678 | 21.25/0.681 | 22.01/0.701 | 22.23/0.713 | 22.41/0.726 |
|          | 3            | 18.65/0.624 | 19.89/0.637 | 20.43/0.648 | 21.43/0.668 | 21.65/0.674 | 21.87/0.689 |

Table 1 shows the experimental comparison results under the three test sets, where the boldface represents the best experimental result. According to the results, the SSIM value of this method is higher than other methods, and the PSNR value is also higher than other methods. This shows that the image reconstruction result of the method in this paper is closer to the original image and achieves a better processing effect.

At the same time, this paper selects some reconstructed remote sensing images from all the comparison results for intuitive comparison. From the images shown in figures 6-8, we can see that the reconstruction effect of the method proposed in this paper is the best.
Figure 6. Two times larger rendering on test1 with different super-resolution reconstruction algorithms.

Figure 7. Different super-resolution reconstruction algorithms enlarge the renderings 3 times on test2.
5. Conclusion
This paper proposes a novel super-resolution reconstruction method for remote sensing images, which is based on DenseNet network convolutional neural network. Aiming at the problems of insufficient texture representation and insufficient detail extraction in other methods in the field of remote sensing images, this paper combines dense back projection network and dense network to connect densely. The idea is integrated into the upper and lower projection units of the deep back projection unit, which solves the problems of insufficient texture representation and insufficient detail extraction. The experimental results show that compared with other methods on the remote sensing image data set in this paper, subjectively, the remote sensing image reconstructed in this paper is clearer, objectively the PSNR and SSIM indicators have also been improved, and good results have been achieved in practical applications.

Acknowledgements
Wuhan education and science project ‘Research on the Deep Teaching Model Reform Based on Big Data in Applied-Oriented Universities Taking Wuhan Business University as an Example’ (2018C080) And Wuhan University Teaching Research Project ‘Research on the Reform and Innovation of SPOC Teaching Mode Based on Big Data-Taking Wuhan Business School as an Example’ (2019091)

References
[1] Zhang Y, Wu W, Dai Y, et al. 2013 Remote sensing images super-resolution based on sparse dictionaries and residual dictionaries IEEE International Conference on Dependable, Autonomic and Secure Computing pp 318-323.
[2] Wu W, Yang X, Liu K, et al. 2016 A new framework for remote sensing image super-resolution: Sparse representation-based method by processing dictionaries with multi-type features Journal of Systems Architecture 64 63-75.
[3] Zhou R G, Xin G and Luo L J 2017 Quantum circuit realization of the bilinear interpolation method for GQIR International Journal of Theoretical Physics 56 (9).
[4] Keys R 1981 Cubic convolution interpolation for digital image processing IEEE Transactions on Acoustics Speech &Signal Processing 29 (6) 1153-1160.
[5] Irani M and Peleg S 1991 Improving resolution by image registration CVGIP: Graphical Models and Image Processing 53 (3) 231-239.
[6] Stark H and Oskoui P 1989 High-resolution image recovery from image-plane arrays, using convex projection JOSAA 6 (11) 1715-1726.

[7] Dong C, Loy C C, He K, et al. 2014 Learning a deep convolutional network for image super resolution European Conference on Computer Vision (Springer, Cham) pp 184-199.

[8] Alvarado S T, Fornazari T, Cóstola A, et al. 2017 Drivers of fire occurrence in a mountainous Brazilian cerrado savanna: tracking long-term fire regimes using remote sensing Ecological Indicators 78 270-281.

[9] Gong C, Yu S, Joesting H, et al. 2013 Determining socioeconomic drivers of urban forest fragmentation with historical remote sensing images Landscape & Urban Planning 117 (9) 57-65.

[10] He H, Lin Y, Chen F, et al. 2017 Inshore ship detection in remote sensing images via weighted pose voting IEEE Transactions on Geoscience & Remote Sensing 55 (6) 3091-3107.

[11] Huang G, et al. 2017 Densely connected convolutional networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[12] Haris M, Shakhnarovich G and Ukita N 2019 Deep back-projection networks for super-resolution Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[13] Zhang H, Fu Y, Feng L B, Zhang Y and Hua R 2019 Implementation of hybrid alignment algorithm for protein database search on the SW26010 many-core processor IEEE Access 7 128054-128063.

[14] Zuo L Q, Sun H M, Mao Q C, Qi R and Jia R S 2019 Natural scene text recognition based on encoder-decoder framework IEEE Access 7 62616-62623.