Research on Key Techniques for Super-resolution Reconstruction of Satellite Remote Sensing Images of Transmission Lines

Weiguo Si¹, Jiajia Han², Zhi Yang³, Xiao Ma³, Yuanjing Deng³, Meng Luo⁴*, Chuang Li³, Binbin Zhao³, Jian Wang³ and Bin Zhao³

¹State Grid Zhejiang Electric Power Co., Ltd. Hangzhou Power Supply Company, Hangzhou, Zhejiang, 310000, China
²State Grid Zhejiang Electric Power Co., Ltd., Research Institute, Hangzhou, Zhejiang, 310000, China
³Power Transmission and Transformation Engineering Research Institute, State Grid China Electric Power Research Institute Co., Ltd., Beijing, 100000, China
⁴School of Surveying Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong, 266000, China

*Corresponding author’s e-mail: yangzhi0713@foxmail.com

Abstract. The data point target on the satellite remote sensing image of the transmission line is restricted by satellite remote sensing imaging equipment and transmission conditions, which makes it difficult to guarantee the clarity of the transmission line. Image super-resolution technology aims to recover high-resolution images from low-resolution images to improve the detailed information of the transmission line itself, which is of great significance for the intelligent inspection of transmission line satellite remote sensing and hidden danger monitoring. Aiming at the problems of traditional methods relying on multi-frame image sequences and the reconstruction results are too smooth, this paper proposes a single-frame remote sensing image super-resolution method based on the boundary balance generation confrontation network. Experimental results show that this method can provide more high-frequency information, and the reconstruction result is closest to the real image. Compared with neighbor interpolation, bicubic interpolation and other methods based on deep convolutional neural networks, the PSNR of the results in this paper is significantly improved, and effectively enhances the detailed information of the transmission line and the surrounding environment.

1. Introduction
After more than half a century of development, remote sensing image processing technology has been widely used in resource exploration, environmental monitoring, military reconnaissance and other fields [1-3]. The resolution of remote sensing images has an important impact on the quality of image interpretation. Compared with low-resolution images, high-resolution images have a higher pixel density and can provide more tones, shapes and texture information. Hardware methods to improve image resolution often require high cost and high technology, and are difficult to effectively apply in large-scale deployments and in real-time environments. Therefore, a more feasible method is to use...
digital image processing technology to obtain high-resolution images from low-resolution images. This method is called image super-resolution (SR).

Image super-resolution is a typical ill-conditioned inverse problem. For this problem, scholars in the fields of signal processing and computer vision have proposed a series of solutions. Su and Sun et al. [4-5] reviewed and analyzed this. Existing remote sensing image super-resolution methods can generally be divided into single-frame image super-resolution and multi-frame image super-resolution according to different types of input and output. Among them, due to the lack of correlation information between multiple frames of images, a single frame image is difficult to obtain prior knowledge of image degradation, which has become the focus and difficulty of the image super-resolution method.

In recent years, the method based on example learning [6] has gradually become a research hotspot. This type of method learns the parameter representation of features from the training sample set, and predicts the missing information in the reconstructed image [4]. Specifically, this type of method is based on deep learning technology to establish an end-to-end convolutional neural network structure to learn end-to-end mapping between low-resolution images. Dong et al. [7] introduced the convolutional neural network to the image super-resolution problem for the first time, and interpreted the three convolutional layers in the network as three steps of image block extraction, non-linear mapping and image reconstruction. Ledig et al. [8] believe that the image reconstruction results should be as close as possible to both the low-level pixel values and the high-level overall style, and the peak signal to noise ratio (PSNR) is not suitable as the evaluation criterion for reconstruction results. So a new perceptual loss function is proposed, and the Generative Adversarial Network (GAN) is introduced into the image super-resolution task for the first time. Berthelot et al. [9] overcome the shortcomings of GAN training that are difficult to converge, and propose a Boundary Equilibrium Generative Adversarial Network (BEGAN).

Inspired by the above work, this paper proposes an improved single-frame remote sensing image super-resolution reconstruction method. The main work and innovations can be summarized as follows: First of all, as far as known, BEGAN has been adjusted and improved for the first time in super-resolution reconstruction of remote sensing images. Secondly, an end-to-end self-encoder structure network is designed. Both the generation network and the discrimination network use this type of structure, which simplifies the network structure design and facilitates code reuse. Finally, unlike traditional GANs that use random vectors as the input of the generation network, the generation network in this paper uses low-resolution images as input to achieve the super-resolution of single-frame remote sensing images with 4 times upsampling.

2. Boundary balance generation confrontation network

2.1. Generating a confrontation network

The generative confrontation network was proposed by Ian Goodfellow and his students. It is a novel generative model, as shown in Figure 1. GAN is inspired by the two-person zero-sum game in game theory. The two parties of the game in GAN are respectively acted by the Generator Network(G) and Discriminator Network(D). Specifically, the generator G learns the distribution of the training set sample data, and converts the noise vector into a sample similar to the real training set data. The higher the similarity, the better. On the contrary, the discriminator D is a two-class model to estimate the probability that a sample belongs to the real training set sample. When a real sample, D outputs a larger probability, and when G generates a fake sample, D outputs a smaller probability. The training process adopts an unsupervised method, in which G and D are trained against each other to improve their own capabilities. G tries to maximize the probability of D's misjudgment, and D tries to minimize the probability of making mistakes, until the sample generated by G, D can no longer judge Its authenticity (output probability is 1/2), and the training ends at this time.
2.2. Boundary balance generation confrontation network
In theory, through GAN's confrontation training, the generated samples can be infinitely close to the real samples and reach the Nash balance [10]. However, imbalances often occur during network training. When the discriminator learns too well, the generator gradient disappears seriously. With this in mind, the Boundary Balanced Generative Adversarial Network (BEGAN) introduces a hyperparameter $k$, which can balance $D$ and $G$ to stabilize the training process. In addition, unlike GAN, BEGAN does not use the method of estimating the similarity between two distributions. Instead, it estimates the difference between the error distribution of the distribution. If the error of the distribution is small, the distribution should be close. Specifically, BEGAN belongs to the GAN of the self-encoder structure type. It designs the generator as a decoder, and the discriminator is no longer a classification network, but as an encoder, the loss function is based on the reconstruction error of the discriminator.

3. Methods of this paper
Deep learning performs well on single-frame image super-resolution tasks. Most of the proposed methods use deep convolutional neural networks as the network structure and the mean square error as the loss function to be optimized. There are two main problems when they are applied to the super-resolution of single-frame remote sensing images. One is that the network is difficult to train, and the problem of gradient disappearance often occurs. The other is that the reconstruction result is too smooth. To this end, this paper proposes a single-frame remote sensing image super-resolution method based on the boundary balance generation confrontation network. The following is a detailed introduction from the network structure and the objective function setting.

3.1. Network structure and parameter setting
Due to the diversity and complexity of remote sensing images, this paper designs a special super-resolution neural network structure as shown in Figure 2. In general, this network is a conditional generation confrontation network, in which high-resolution remote sensing images are used as conditional variables to guide the low-resolution image reconstruction process. Both the generator and the discriminator are the same type of end-to-end autoencoder type network. Unlike the generator input in the traditional generative confrontation network which is an arbitrary random vector, the input of the network generator in this paper is a low-resolution remote sensing image magnified 4 times by the bilinear cubic interpolation method. In order to facilitate the use of pre-trained network weights, the autoencoder type network structure design in this paper refers to SegNet [11]. Compared with the original BEGAN, the network structure designed in this paper has more convolutional groups and feature maps. Compared with the original BEGAN, the network structure designed in this paper has more convolutional groups and feature maps. In order to enhance the quality of the generated image and accelerate the network convergence, a "jump connection" method is used between each convolution group and its corresponding deconvolution group. In the setting of other hyperparameters, batch_size is 16, epochs is 100, and the initial learning rate is 0.0001. The learning rate decays every 2000 iterations to 0.95, and the initial value of $\gamma$ is set to 0.75.
3.2. Objective function

In the network training, this paper uses a loss function based on the L1 norm and the reconstruction error of the discriminator. This loss function can not only balance the competitiveness of the generator and the discriminator, but also enhance the visual quality of the generated image.

Assuming that the pixel-wise error of the image obeys the Gaussian distribution, the L1 norm loss function is:

\[ L(I) = |I^{HR} - G(I^{LR})| \] (1)

In the formula, \( I^{HR} \) represents the real high-resolution image, and \( I^{LR} \) represents the corresponding low-resolution image. The loss function to be optimized and update rules of the generator and discriminator are:

\[
\begin{align*}
L_D &= L_{D_{fr}} - k_t L_{D_{f}}, \quad \text{for } \theta_D \\
L_G &= L(G(z) - x), \quad \text{for } \theta_G \\
k_{t+1} &= k_t + \lambda_k (y L_{D_{fr}} - L_G)
\end{align*}
\] (2)

among them

\[
L_{D_{fr}} = L(D(x; \theta_D) - x) \quad \text{(3)}
\]

\[
L_{D_f} = L(D(y; \theta_D) - y) = L(D(G(z; \theta_G) - G(z; \theta_G)); \theta_D) \quad \text{(4)}
\]

\[
y = G(x; \theta_D) \quad \text{(5)}
\]

\[
y = \frac{E[L(G(z))]}{E[L(x)]} \quad \text{(6)}
\]

In equations (2) to (6), where \( x \) represents a real high-resolution image, \( y \) represents a high-resolution image, and \( z \) represents a corresponding low-resolution image. \( L_{D_{fr}} \) and \( L_{D_f} \) represent the loss of the discriminator with respect to \( x \) and \( y \), \( \theta_G \) and \( \theta_D \) represents the relevant parameters of the generator and the discriminator respectively. \( \lambda_k \) represents the proportional gain of \( k \), and \( k_t \) represents the \( t \)-th update value of \( k \). The parameter \( k \) can balance the competitiveness of the generator and the discriminator. Where \( \gamma \) is the ratio of the expected loss of the generated sample to the expected value of the true sample loss. This parameter can control the trade-off between the diversity of the generated image and the visual quality.

4. Experiment and analysis

The data set used in this experiment is derived from the NWPU-RESISC45 high-resolution remote sensing map data set. NWPU-RESISC45 includes 45 scene categories such as airports, overpasses,
railway stations, residential areas, etc. Each category has 700 images, ensuring the authenticity and diversity of experimental data. In this experiment, 10 images of each category in the original data set are selected, and a total of 450 images form the experimental data set. According to convention, the order of the image samples is randomly shuffled and divided according to the ratio of 8:2, which means that 80% of the samples are used for training and 20% of the samples are used for testing. The experiment in this article uses Python language, Tensorflow r1.4 development framework and computer hardware configuration mainly two Quadro M4000GPU (8GB), 16GB RAM. Before making a detailed quantitative evaluation of the super-resolution reconstruction results of remote sensing images, firstly show the comparison between other representative methods and the super-resolution reconstruction results of the method in this paper. The red rectangle in the figure indicates the region of interest, as shown in Figure 3 and Figure 4.

A total of 160,000 iterations of network training have been performed. Figures 5 and 6 respectively show the changing trends of the network convergence metric Mglobal and the image super-resolution evaluation indicator PSNR during the network training process. It can be seen that the training process is fast and stable. After 80,000 iterations, the PSNR gradually converges to 27.3, and Mglobal finally converges to 0.055. The parameter Mglobal represents the degree of network training. From equation (7), the smaller the Mglobal value, the better the network training.

\[ M_{\text{global}} = L(x) + |\gamma L(x) - L(G(Z_G))| \] (7)

The mean square error (MSE) is:

\[ \text{MSE} = \frac{1}{hw} \sum_{k=1}^{3} \sum_{i=1}^{h} \sum_{j=1}^{w} (x(i,j,k) - y(i,j,k))^2 \] (8)

The peak signal to noise ratio (PSNR) is:

\[ \text{PSNR} = 10 \log_2 \left( \frac{(2^n-1)^2}{\text{MSE}} \right) \] (9)

The structural similarity (SSIM) is:

\[ \text{SSIM} = \xi(x,y) \cdot c(x,y) \cdot s(x,y) \] (10)

In formula (8), h, w, and k respectively represent the length, width and channel number of the image. In formula (9), n is equal to 8. In formula (10), \( l(x,y) \), \( c(x,y) \), \( s(x,y) \) represent the brightness, contrast, and structural similarity of the image, respectively. The value range of SSIM is [0,1], the larger the value the image distortion is smaller.

The three indicators in equations (8) to (10) are used to comprehensively evaluate the super-resolution reconstruction method proposed in this paper. Table 1 is the comparison of PSNR and SSIM of each method in Figure 3 with respect to the sample image, and Table 2 is the comparison of the reconstruction results of each method on the test set samples.

Figure 3, Figure 4 and Table 1 comprehensively compare the reconstruction results of the four images of transmission towers, reservoirs, roads, and buildings in this method and other methods. The results show that this method overcomes the shortcomings of the overall smoothness of the traditional reconstruction results, and the reconstruction results restore more high-frequency details. This is mainly due to the fact that the network in this paper adopts a framework based on a generative confrontation network and no longer uses the mean square error as the loss function. Inspired by BEGAN, the network in this paper uses the hyperparameter \( \gamma \) to balance the diversity and visual quality of the generated images, which achieves the best remote sensing image super-resolution results. Table 2 is the comparison between the method in this paper and other methods on the test set. Compared with the neighbor interpolation and bicubic interpolation methods, the PSNR is improved by about 2.70dB, and the SSIM is improved by about 0.06. Compared with other methods based on deep convolutional neural networks, the PSNR is improved by about 0.72dB, and the SSIM is improved by about 0.02.
Figure 3. Comparison of super-resolution reconstruction results of various methods.

Figure 4. Comparison of ROI super-resolution reconstruction results.

Figure 5. Network convergence metric Mglobal.

Figure 6. Peak signal-to-noise ratio PSNR.

Table 1. Peak signal-to-noise ratio/structure similarity of each super-resolution reconstruction method

| Image     | Nearest | Bicubic | FSRCN | ESPCN | DRCN | RED | SRDenseNet | SRGAN | Method of this article | HR  |
|-----------|---------|---------|-------|-------|------|-----|------------|-------|------------------------|-----|
| airplane  | 22.93   | 24.09/  | 25.92/| 25.63/| 27.01/| 25.97/| 27.05/0.93 | 26.47/ | 0.92                   | ∞/1 |
|          | /0.82   | 0.85    | 0.90  | 0.89  | 0.93  | 0.90 | 0.88       | 0.88  |                        |     |
| lake_690 | 23.75   | 24.60/  | 27.18/| 26.74/| 28.19/| 27.23/| 28.11/0.97 | 26.68/ | 0.94                   | 27.72/0.96 | ∞/1 |
|          | /0.91   | 0.92    | 0.95  | 0.94  | 0.97  | 0.95 | 0.95       | 0.94  |                        |     |
| overpass | 23.16   | 24.16/  | 25.23/| 25.07/| 25.60/| 25.31/| 25.79/0.91 | 24.55/ | 0.85                   | 25.50/0.90 | ∞/1 |
| _405     | /0.82   | 0.85    | 0.88  | 0.87  | 0.91  | 0.88 | 0.88       | 0.85  |                        |     |
| ship_010 | 19.64   | 20.13/  | 21.37/| 20.89/| 22.03/| 21.46/| 22.32/0.93 | 21.86/ | 0.91                   | 21.97/0.92 | ∞/1 |
|          | /0.86   | 0.87    | 0.90  | 0.88  | 0.92  | 0.89 | 0.89       | 0.91  |                        |     |

Table 2. The mean square error/peak signal-to-noise ratio/structure similarity of each super-resolution reconstruction method of the test set samples

| Index | Nearest | Bicubic | FSRCN | ESPCN | DRCN | RED | SRDenseNet | SRGAN | Method of this article | HR  |
|-------|---------|---------|-------|-------|------|-----|------------|-------|------------------------|-----|
| MSE   | 333.84  | 269.41  | 221.73| 209.32| 193.7/ 6 | 233.5/6 | 176.31 | 276.42 | 221.45 | 0               |

Table 3. The mean square error/peak signal-to-noise ratio/structure similarity of each super-resolution reconstruction method of the test set samples

| Index | Nearest | Bicubic | FSRCN | ESPCN | DRCN | RED | SRDenseNet | SRGAN | Method of this article | HR  |
|-------|---------|---------|-------|-------|------|-----|------------|-------|------------------------|-----|
| MSE   | 333.84  | 269.41  | 221.73| 209.32| 193.7/ 6 | 233.5/6 | 176.31 | 276.42 | 221.45 | 0               |
5. Conclusion
Generative confrontation network can generate realistic images. This article adjusts and improves the boundary balance generation confrontation network to achieve end-to-end single-frame remote sensing image super-resolution. The generator and discriminator of the network in this paper are both self-encoder structure with jump connection, and the generator input is low-resolution remote sensing image. When using the NWPU-RESISC45 dataset for experiments and using PSNR and SSIM as the image super-resolution reconstruction quality evaluation indicators, compared with other super-resolution methods based on convolutional neural networks, the method in this paper is more stable, faster, convergent, and more detailed in reconstruction. It proves the effectiveness and advancement of this method. But it is also noticed that the selection of hyperparameters in network training is still a difficult task. Future research work is mainly to do super-resolution reconstruction for specific types of remote sensing image data, and to combine deep learning-based methods with traditional methods to achieve better super-resolution reconstruction.

Acknowledgments
Thanks to the State Grid Corporation of China Science and Technology Program Project "Research on Satellite Remote Sensing and Large-scale Grid Remote Sensing and Fine Fusion Sensing Technology" fund (5700-202019162A-0-0-00), and thank colleagues who have studied with us.

References
[1] 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.
[2] 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.
[3] He H, Lin Y, Chen F, et al. (2017) Inshore ship detection in remote sensing images via weighted pose voting. IEEE Trans-actions on Geoscience & Remote Sensing, 55 (6) : 3091-3107.
[4] Su H, Zhou J, Zhang Z H. (2013) Review of super-resolution image reconstruction methods. Acta Automatica Sinica, 39(8): 1202-1213.
[5] Sun X, Li X G, Li J F, et al. (2017) Research progress of image super-resolution restoration based on deep learning . Acta Automatica Sinica, 43(5): 697-709.
[6] Timofte R, Rothe R, Gool L V. (2016) Seven ways to improve example- based single image super-resolution. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas. 1865-1873.
[7] Dong C, Chen C L, He K, et al. (2014) Learning a deep convo-lutional network for image super-resolution. European Conference on Computer Vision, 184-199.
[8] Ledig C, Wang Z, Shi W, et al. (2017) Photo-realistic single im-age super-resolution using a generative adversarial net-work. IEEE Conference on Computer Vision and Pat-tern Recognition, Hawaii. 105-114.
[9] Berthelot D, Schumm T, Metz L. (2017) BEGAN: boundary equi-librium generative adversarial networks. arXiv: 1703. 10717.
[10] Ratliff L J, Burden S A, Sastry S S. (2013) Characterization and computation of local Nash equilibria in continu-ous games. 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello. IL, 917-924.
[11] Badrinarayanan V, Handa A, Cipolla R. (2015) SegNet: a deep convolutional encoder- decoder architecture for robust semantic pixel-wise labelling. Computer Science.
[12] Kim J, Lee J K, Lee K M. (2016) Deeply-recursive convolutional network for image super-resolution. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas. 1637-1645.

[13] Tong T, Li G, Liu X, et al. (2017) Image super-resolution using dense skip connections. IEEE International Conference on Computer Vision, Italy. 4809-4817.