SUPER-RESOLUTION RESEARCH ON REMOTE SENSING IMAGES IN THE MEGACITY BASED ON IMPROVED SRGAN

Liang Xin [1,2,*], Zhipeng Li [1,2], Sijian Wang [1,2]

1. Shanghai Institute of Surveying and Mapping, Shanghai,,No.419 of Wuning Road, Putuo District of Shanghai, rserlzp@aliyun.com;

2. Key Laboratory of Spatial-temporal Big Data Analysis and Application of Natural Resources in Megacities,MNR

KEY WORDS: High resolution satellite imagery, Deeplearning, Super resolution, Urban Remote Sensing, SRGAN

ABSTRACT:

Remote sensing images of Earth observation with high spatial resolution and high temporal resolution are critical for the application of remote sensing technology in Megacities.With the development of Smart City,more demands which are still difficult to be perfectly satisfied on the spatial resolution and temporal resolution of remote sensing images have been put forward.This paper studies the use of SRGAN which means Super-Resolution using a Generative Adversarial Network (a network structure that uses the loss function considering the perceptual loss and the adversarial loss to improve the spatial resolution of remote sensing images) for super-resolution reconstruction of single remote sensing image.It is able to enhance the spatial resolution of remote sensing images and improve the depth and breadth of remote sensing images.We adjust the reasonable parameters and network structure for our research by analysing the SRGAN in the network architecture, the perceptual loss and the adversarial loss.A super-resolution model is obtained by training with aerial photogrammetry images whose spatial resolution are 0.1 meter in Shanghai.We find the improved SRGAN has a good performance in in remote sensing image super-resolution by comparing the super-resoved images with real high-resolution images in visual perception, spatial position mapping accuracy and chromaticity spatial information. In addition, it is proved that the trained model is also effective to deal with Worldview-2 and SuperView-1 satellite images whose spatial resolution are 0.5 m. Our research shows that our method which can effectively realize the super-resolution of remote sensing images has great potential in the application of remote sensing technology such as urban mapping and changes monitoring.

1. INTRODUCTION

High resolution remote sensing images play an important role in agriculture and forestry monitoring, urban planning, military reconnaissance and urban mapping, etc. The spatial resolution of remote sensing image represents the actual size of the unit pixel, which is one of the key indexes of evaluating the image quality. However, due to the high cost and long period of researching and developing, how to acquire high-quality images is always one huge challenge of the field of our domestic remote sensing. Super resolution (SR) provides us several ideas to solve this problem. SR is firstly proposed by Harris et al (Harris et al, 1964) in the 1960s. It starts from the image information itself, which can be divided into single-frame reconstruction and multiple reconstruction according to the number of low-resolution images required for reconstruction.

There are three categories of SR in the current, method of interpolation, reconstruction and learning.Method of interpolation (Tsai R et al ,1984) is the earliest method of SR, which uses the gray value of the adjacent pixel to generate the gray value of the pixel to be interpolated. Among all the SR methods, it has the lowest complexity and most real-time. Yet the fringe effect is obvious and details are poorly restored in the result.Method of reconstruction models the image imaging process. It provides a priori information for high resolution image reconstruction based on low resolution image sequences, and simulates high-resolution images combined different information in the scene, which is a method for changing spatial resolution in time resolution. This method usually needs pre-georeferencing, which demands large amount of computation.

Moreover, the precision cannot be guaranteed and is less efficient.Method of learning develops rapidly in recent years and become the mainstream direction of SR method. It overcomes the limitation of the reconstruction method that is difficult to determine the increasement multiple of resolution, and can deal with the single image.Relying on constructing high- and low-resolution image libraries, this method obtains the intrinsic correspondence between the two through sample learning.

2. TECHNIQUE METHODS

2.1 Technical process

SISR(single image super resolution) aims at obtaining high resolution images from input low resolution images.Deep learning has been proved to be applicable to the super-resolution reconstruction of images.Deep neural network for image super-resolution reconstruction can be roughly divided into two categories according to their characteristics.One is aimed at optimizing distortion values, such as peak signal-to-noise ratio (PSNR) and mean square error (MSE), to reconstruct higher resolution images.The network based on the true loss function of pixels belongs to this category. The other is to produce more realistic images, which have better perception quality.

Usually, additional perceptual loss is considered. The proposed SRGAN combines perceptual loss with countermeasure loss to generate realistic texture in SR. There are three key parts of the algorithm: network architecture, confrontation loss and

*Zhipeng Li, E-mail: rserlzp@aliyun.com

perception loss. The following figure shows an overall technical process.
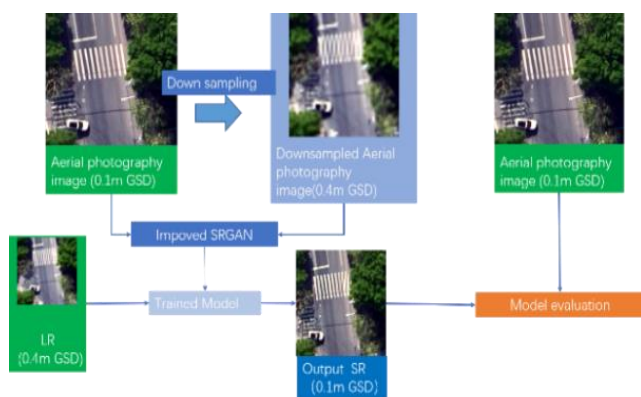

Figure 1. The flowchart of this paper

2.2 Network architecture

2.2.1 Generator network

Generator networks are designed to train a generating function G to estimate the relationship between low-resolution images(LR) and corresponding high-resolution images(HR).The generator network is a feedforward CNN G_{θ_G} which is determined by θ_G. θ_G represents the weight and offset of the L-layer depth network, which is determined by optimizing the specific super-resolution loss function l^{SR}.The relationship between image data set I_n^{HR} and corresponding I_n^{LR} is trained by the following formulas:

$$\hat{\theta}_G = argmin_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{SR}\left(G_{\theta_G}(I_n^{LR}), I_n^{LR}\right) \qquad (1)$$

θ_G represents the weight and offset of the L-layer depth network, l^{SR} means the loss function,

Referring to X. Wang et.al(X. Wang et al, 2018) , this paper makes two modifications to the generator: 1) removing all BN layers; 2) replacing the original basic block connection with residual dense blocks, which combines multi-level residual network and density connection.
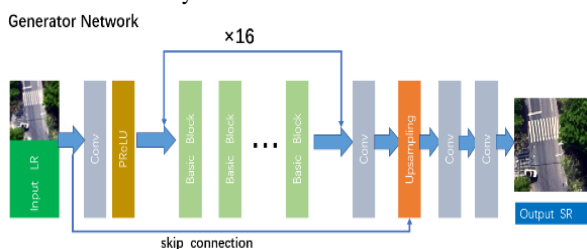

Figure 2. Structure of Generator Network

The above structure follows the criteria proposed by Radford. There are 16 hidden layers in the network; in each layer, we use 3*3 convolution with 64 channels. Batch standardization is added after convolution. After using Parametric ReLU as an activation function, the function map is activated. After activation is another convolution layer and batch standardization layer. This skips the start and end of the connection connection residual block.

2.2.2 Discriminator network

The discriminator network D_{θ_G} is used to optimize the generator network to solve the problem of confrontation loss of super-resolution images.

$$min_{\theta_G} max_{\theta_D} \; \mathbb{E}_{I^{HR} \sim p_{train(I^{HR})}}\left[\log D_{\theta_D}(I^{HR})\right] + \mathbb{E}_{I^{LR} \sim p_{G(I^{HR})}}\left[\log(1 - D_{\theta_D}\left(D_{\theta_G}(I^{LR})\right))\right] \qquad (2)$$

The core idea of this formula is to train the generator to obtain the best performance, so that the difference between the model super-resolution image and the real image can be "deceived" discriminator as successful as possible.

This paper enhances the discriminator based on the idea of relative GAN.(Jolicoeur-Martineau et al, 2018) A relative discriminator constantly strives to predict the possibility that real images are more realistic than forged ones. The modification of this discriminator helps to learn more sharp edges and more detailed textures.

2.2.3 Perceptual loss function

Perceptual loss function l^{SR} is very important for the performance of generator networks. l^{SR} is usually modeling based on MSE.(C. Dong et al, 2016) This method is prone to excessive smoothing of images, resulting in poor perception quality. E. Denton et al (E. Denton et al.2015) uses the method of generating antagonism to solve this problem (GAN).

Referring to Ledig, C et al(Ledig, C et al, 2018) this paper defines the perceptual loss function by weighting the loss of content and the loss of confrontation.

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \qquad (3)$$

Content loss：Traditional pixel-by-pixel MSE loss is widely used in image optimization, but this method is prone to lack of high-frequency information, resulting in smooth texture, obscure edges and poor visual perception.

The loss function based on perceptual similarity can overcome this disadvantage.The VGG loss is defined based on the ReLU activation layer of the pre-training 19-layer VGG network.

Adversarial loss：According to the structure of GAN, the loss of resistance is added into the loss function, which can continuously enhance the performance of the generator in the iteration with the discriminator. The confrontation loss l_{Gen}^{SR} is defined according to the probability that the discriminator covers all samples.

$$l_{Gen}^{SR} = \sum_{n=1}^{N} -log D_{\theta_D}(G_{\theta_G}(I^{LR})) \qquad (4)$$

$D_{\theta_D}(G_{\theta_G}(I^{LR}))$ refers to the probability that the super-resolution reconstructed image $G_{\theta_G}(I^{LR}$ coincides with the original high-resolution image.

2.2.4 Training details and parameters

According to the 2040*1024 pixel frame (each pixel represents 0.1m*0.1m aerial image), 1000 0.1m spatial resolution aerial images of Shanghai in 2017 were cut. IDL Biline algorithm is

used to downsample the original clipped image 4*4 to obtain the corresponding low-resolution image.The data are trained in NVIDIA Quadro P6000 GPU and the model is built using PyTorch framework, with the minimum batch set to 32 and the size of the cut HR block 128*128.

The training is divided into two stages: first, a PSNR-oriented model is trained with L1 loss. The learning rate is initialized to $2*10^{-4}$, and the minimum batch update scale is reduced by two times every $2*10^5$. Then a trained PSNR-oriented model is used as the initialization of the generator to avoid the occurrence of erroneous local optima. The loss function is used to train the generator. The learning rate is set to $1*10^{-4}$ and the number of iterations is 10^6.The use of pre-training model helps to avoid the occurrence of local optimum errors, shorten the training time and make the discriminator focus on texture discrimination more quickly.To optimize, we alternately update the generator and discriminator networks until the model converges.

3. DATA

The remote sensing images utilized in this article are the images acquired in Shanghai, China whose ground sampling distance is below 1 meter. There are satellite images: SuperView-1 and Worldview-2; airborne images obtained by Leica DMC III.
SuperView-1 is the first business satellite launched by China Aerospace Science and Technology Corporation. This constellation has four satellites in orbit, which can revisit anywhere in the world. Worldview-2 is launched by Digital globe Corporation in America, the average satellite revisit period is 1.1 days. Table 1 demonstrates several parameters of the satellite.
The aerial images of Shanghai are acquired by the airborne Leica DMC III. DMC III is the first camera with large array frame CMOS sensor, whose ground sampling distance of image can reach 0.1 meter. Table 2 shows the partial parameters of the camera.

DMC III	
pixel size	3.9μm CMOS
sensor size	26112*15000
focal length	92mm
flight altitude	2500m
Ground sampling distance	10cm

Table 2 Partial parameters of Leica DMC III

4. RESULTS AND ANALYSIS

In this paper,deep network model between images resolution of 0.4m and 0.1m has been trained and evaluated by using aerial photogrammetry images whose spatial resolution are 0.1m in Shanghai as training data.Images of low-resolution with downsampling factor r= 4,super-resolution and original high-resolution have been listed as followed,involing different kinds of land cover.
It can be found from visual perception that the improved SRGAN could effectively reconstruct spatial resolution of remote sensing images.We will continue to analyze and evaluate the accuracy of the survey and chromaticity of the images in the future.In this paper,the trained model has been applied to the satellite images of WorldView-2 and SuperView-1 whose spatial resolution are both 0.5 meter, and positive results have been achieved.
Certainly，considering the differences in imaging parameters and modes between satellite sensors and aerial photogrammetry cameras, the super-resolution models trained by aerial photogrammetry images applied to satellite images need further research and analysis.

Satellite sensors	PAN		MS				
	Spatial resolution	wavelength	Spatial resolution	wavelength			
				Blue	Green	Red	Nir
SuperView-1	0.5 m	450-890 nm	2.0 m	450-520 nm	520-590 nm	630-690 nm	770-890 nm
Worldview-2	0.5 m	450-800 nm	2.0 m	450-510 nm	510-580 nm	630-690 nm	770-895 nm

Table 1. Parameters of Satellite sensors

Figure 3. 50*50 LR \ 200*200 SR \ 200*200 HR   Road marking in the aerial image

Figure 4. 50*50 LR \ 200*200 SR \ 200*200 HR   Vehicles in the aerial image
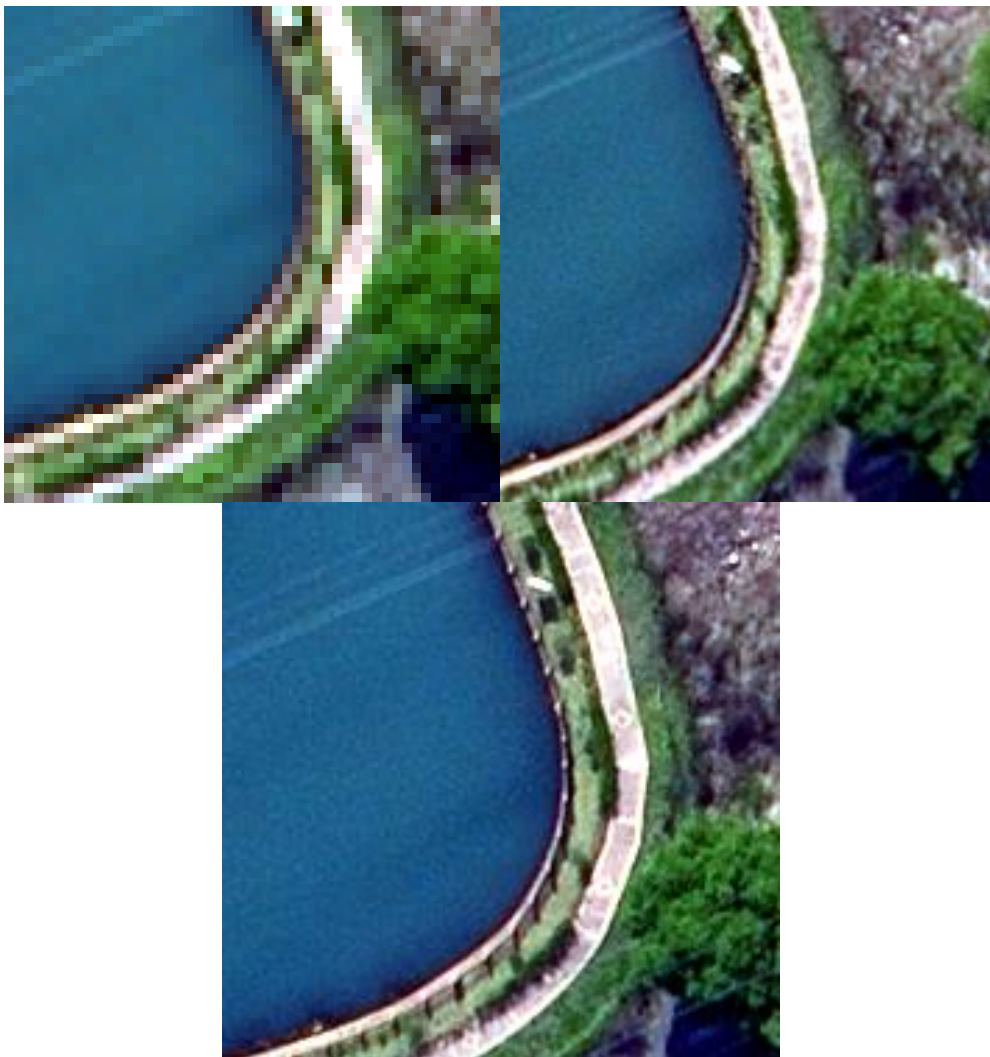

Figure 5. 50*50 LR \ 200*200 SR \ 200*200 HR   River edge in the aerial image
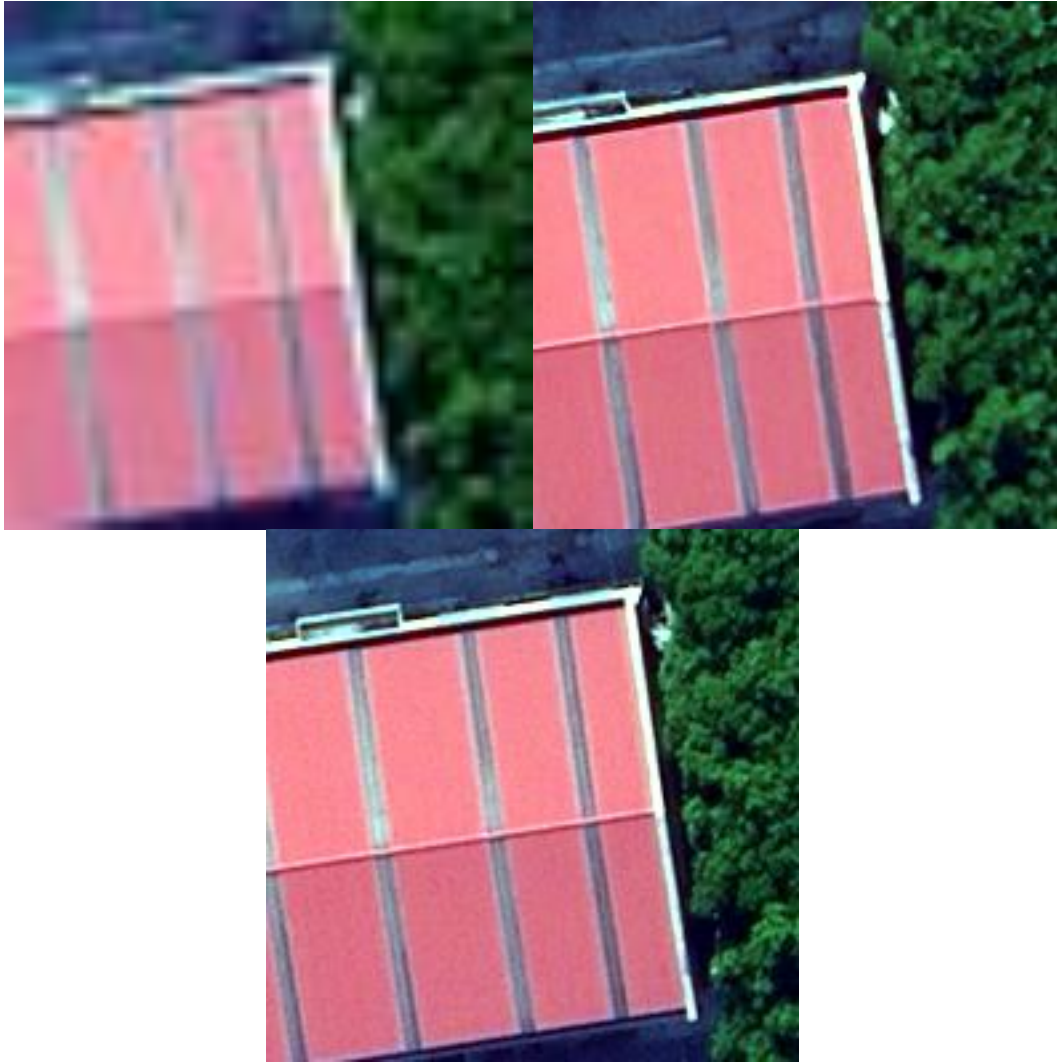
Figure 6. 50*50 LR \ 200*200 SR \ 200*200 HR   House in the aerial image
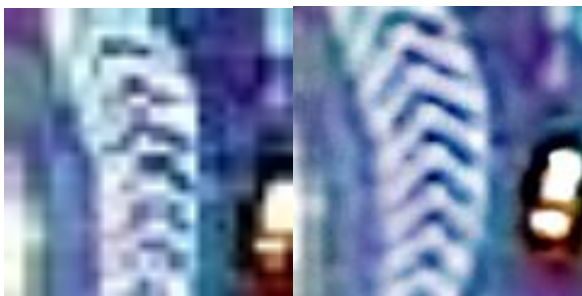

Figure 7. 30*30 LR \ 120*120 SR   Road marking in the Worldview-2 satellite image
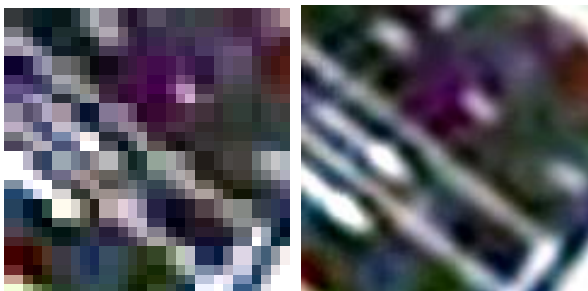

Figure 8. 30*30 LR \ 120*120 SR   Road marking in the Superview-1 satellite image

5. CONCLUSION

This paper researchs SRGAN and modifies the parameters and framework by analysing its network architecture,loss function and other parameters.It has been proved that our method has a good performance in spatial resolution reconstruction on remote sensing images by using aerial photogrammetry images in Shanghai as training data.In this paper, a super-resolution model of remote sensing images between the spatial resolution of 0.4 m and 0.1 m is established.Additionally,super-resolution remote sensing images whose spatial resolution is about 0.1 m are obtained by applying the trained model to WorldView-2 satellite images whose spatial resolution is 0.5 m.

Our method can effectively obtain remote sensing images of high spatial and temporal resolution remote sensing images by applying the model trained by aerial images with ultra high resolution to Satellite remote sensing images.It is significantly important to applications on urban remote sensing.

REFERENCES

Harris J L , 1964. Diffraction and resolving power. Journal of the Optical Society of America, 54 (7),pp. 931-933.

Tsai R，Huang T, 1984. Multiple frame image restoration and registration. Advances in Computer Vision & Image Processing, Greenwich，Ct: Jai Press Inc，pp: 317-339 .

X. Wang,et al, 2018. Esrgan: Enhanced super-resolution generative adversarial networks. 2018CVPR.

Ledig, C.et al , 2017. Photo-realistic single image superresolution using a generative adversarial network. 2017CVPR.
Jolicoeur-Martineau, 2018. The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734.

C. Dong, et al. 2016. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2),pp. 295–307.

E. Denton, et al. 2015. Deep generative image models using a laplacian pyramid of adversarial networks. InAdvances in Neural Information Processing Systems (NIPS), pp. 1486–1494.