SAR IMAGE SUPER-RESOLUTION BASED ON NOISE-FREE GENERATIVE ADVERSARIAL NETWORK

Feng Gu[1,2], Hong Zhang[1], Chao Wang[1,2], Fan Wu[1]*

1. Key Laboratory of Digital Earth Science, of Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing, China
2. University of Chinese Academy of Science, Beijing, China
*Correspondence:zhanghong@radi.ac.cn(H.Z.);wangchao@radi.ac.cn(C.W.); Tel.: +86-10-8217-8186

ABSTRACT

Deep learning has been successfully applied to the ordinary image super-resolution (SR). However, since the synthetic aperture radar (SAR) images are often disturbed by multiplicative noise known as speckle and more blurry than ordinary images, there are few deep learning methods for the SAR image SR. In this paper, a deep generative adversarial network (DGAN) is proposed to reconstruct the pseudo high-resolution (HR) SAR images. First, a generator network is constructed to remove the noise of low-resolution SAR image and generate HR SAR image. Second, a discriminator network is used to differentiate between the pseudo super-resolution images and the realistic HR images. The adversarial objective function is introduced to make the pseudo HR SAR images closer to real SAR images. The experimental results show that our method can maintain the SAR image content with high-level noise suppression. The performance evaluation based on peak signal-to-noise-ratio and structural similarity index shows the superiority of the proposed method to the conventional CNN baselines.

Index Terms—Synthetic aperture radar, super-resolution, generative adversarial network

1. INTRODUCTION

The need for high resolution (HR) synthetic aperture radar (SAR) image is urgent in performing target recognition, object tracking, and image classification tasks. However, due to the interference of speckle noise inherent in SAR images and the loss of high-frequency information, SAR image super-resolution (SR) has always been a hot but challenging problem.

Estimating an HR image from its low-resolution (LR) counterpart is referred to as SR which is an ill-posed problem. The traditional linear interpolation methods, such as bilinear and spline interpolation, are the most direct and simple tools to improve the resolution but perform poor in edge preserving. Therefore, some derivative nonlinear interpolation techniques combined with statistic model are proposed to maintain edge details [1], [2]. Reconstruction-based methods are another branch to reconstruct realistic texture detail while avoiding edge artifacts. For example, Merino et al. [3] applied variable-pixel linear reconstruction (VPLR) to evaluate the statistical significance of the pixels and remove the effect of geometric distortion on both shape and photometry. In [4], a multi-scale dictionary is introduced to produce desired sharp details by sparse representation. In addition, since the Bayesian estimation and the maximum likelihood came into being, the learning-based reconstruction approaches, such as gradient profile prior [5] and TV regularization [6], have been applied to stabilize the inversion of SR by taking the prior knowledge into consideration.

Currently, convolutional neural network (CNN) based SR algorithms have shown outstanding performance on ordinary images [7]. But the CNN baselines so far have not considered the presence of multiplicative noise; therefore it is hard for them to maintain high-frequency details of the HR SAR images. In order to solve the above problems, also inspired by the concept of the generative adversarial network [8], we design a network capable of simultaneously denoising and improving resolution. The network is divided into two components, namely a noise-free generator and a discriminator with the adversarial objective function. The former produces pseudo high-resolution SAR images and the latter tries to identify them from real HR images. Such a strategy can encourage the reconstructions to move towards the manifold of real HR SAR images with high probability. Therefore, the reconstructions look more naturally to recover clear edges and finer texture details. The adopted qualitatively and quantitative metrics demonstrate that the proposed network can achieve the state-of-the-art performances.

2. METHOD

We define low-resolution SAR image as I^{LR}, which is the low-resolution version of its high-resolution counterpart I^{HR} by bicubic interpolation. The pseudo super-resolution image is denoted as I^{SR}. For an intensity image of HH or VV channel, we describe I^{LR} by a matrix of size $W \times H \times 1$ and I^{HR}, I^{SR} by $rW \times rH \times 1$, where r is the downsampling factor. In this paper, $r = 2$.

2.1. Noise-Free Generative Adversarial Network

The proposed Noise-Free Generative Adversarial Network (NF-GAN) for SAR image super-resolution contains two sub-networks, as illustrated in Fig. 1. The generator network takes I^{LR} as input and outputs I^{SR}. The discriminator network takes in both I^{SR} and I^{HR} and returns probability, a number between 0 and 1, with 1 representing a prediction of authenticity and 0 representing fake. The aim of the generator network is to produce believable I^{SR} to let the discriminator network make mistakes where the aim of the discriminator network is to identify images coming from the generator as fake.

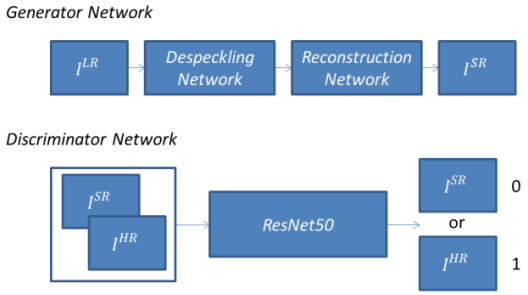

Fig. 1. The diagrammatic sketch of NF-GAN

The despeckling network and the reconstruction network in generator share one structure shown in Figure 2. We follow the architectural guidelines summarized in [9] and lightweight the whole network. Note that the dense block stacks the output of each layer in a densely connected manner. Therefore, the number of output channels of each dense block is 64 after 4 times of stacking. We use Resnet50 [10] as the backbone of the discriminator.

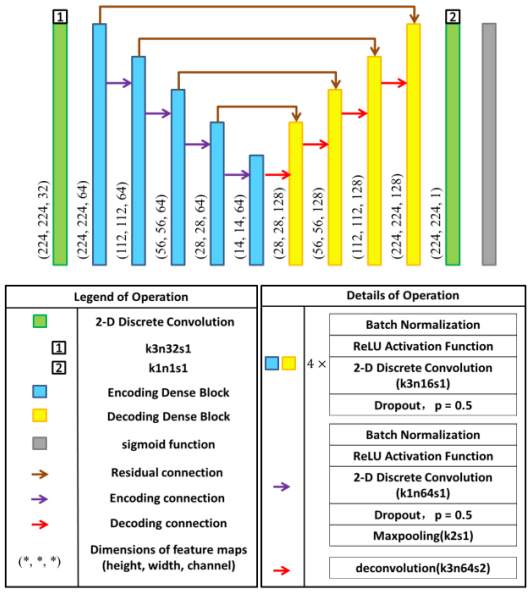

Fig. 2. The architecture of the despeckling and the reconstruction network. k denotes kernel size, n denotes the number of output channel and s denotes stride.

The goal of the despeckling network is to first remove the multiplicative noise in I^{LR}. The optical photographs are used to pre-train the despeckling network because there is no clean SAR image in the real word. We add speckle noise to the clean optical photographs according to the K-distribution noise degradation model as follows:

$$Y = gI \tag{1}$$

$$p_g(g) = \frac{1}{g\Gamma(L)}(Lg)^L exp(-Lg)\,, g \geq 0 \tag{2}$$

where **Y** is the noise image and **I** is the clean optical photograph. ***L*** denotes the equivalent number of looks and $\Gamma(\cdot)$ is the gamma function. $\{\mathbf{I}, \mathbf{Y}\}$ denotes the training pair to force the despeckling network to learn noise pattern and pre-update its own weights. ***L*** is randomly selected for each training pair in the range from 1 to 4, which basically covers all possible equivalent number of looks. The rest of the training details of the despeckling network, such as the loss function, optimization and initialization algorithm can be found in [11]. With a reconstruction network followed at the end, the whole generator can generate noise-free high-resolution SAR image end-to-end.

2.2. Objective function and training details

The objective function consists of an adversarial item and a pixel-wise item. The adversarial item is first proposed in [8], and in this paper, defined as follows:

$$\min_{\theta_G}\max_{\theta_D} \boldsymbol{E}_{I^{HR}}\left[log N_{\theta_D}(\mathrm{I}^{\mathrm{HR}})\right] + \boldsymbol{E}_{I^{LR}}[log(-N_{\theta_D}(N_{\theta_G}(I^{LR})))] \tag{3}$$

where N_{θ_D} is the discriminator network parameterized by θ_D and N_{θ_G} is the generator network parameterized by θ_G. For $\mathrm{N}_{\theta_D}(\mathrm{I}^{\mathrm{HR}})$, it should be as close as possible to 1 and $\mathrm{N}_{\theta_D}(\mathrm{N}_{\theta_G}(\mathrm{I}^{\mathrm{LR}}))$ to 0. On the generator side, $\mathrm{N}_{\theta_D}(\mathrm{N}_{\theta_G}(\mathrm{I}^{\mathrm{LR}}))$ should output the highest possible value to fool the discriminator. The above min-max optimization problem encourages perceptually better reconstructions which are mixed within the subspace of real HR SAR images.

In addition to the adversarial item, the mean squared error (MSE) between I^{HR} and I^{SR} is used as a pixel-wise objective function:

$$\min_{\theta_G}\left(\frac{1}{2WH}\left\|N_{\theta_G}(I^{LR}) - I^{HR}\right\|_2^2\right) \tag{4}$$

where $N_{\theta_G}(\mathrm{I}^{\mathrm{LR}})$ and I^{HR} are of size W×H×1.

Once both objective functions are defined, they are learned jointly by the alternating gradient descent. In other words, when training the N_{θ_D}, we hold the N_{θ_G} constant, and vice versa for training the N_{θ_G}. Note that both θ_G and θ_D are initialized randomly except the weights in the

despeckling network, which is initialized by pre-training. Stochastic gradient descent (SGD) is used to train the

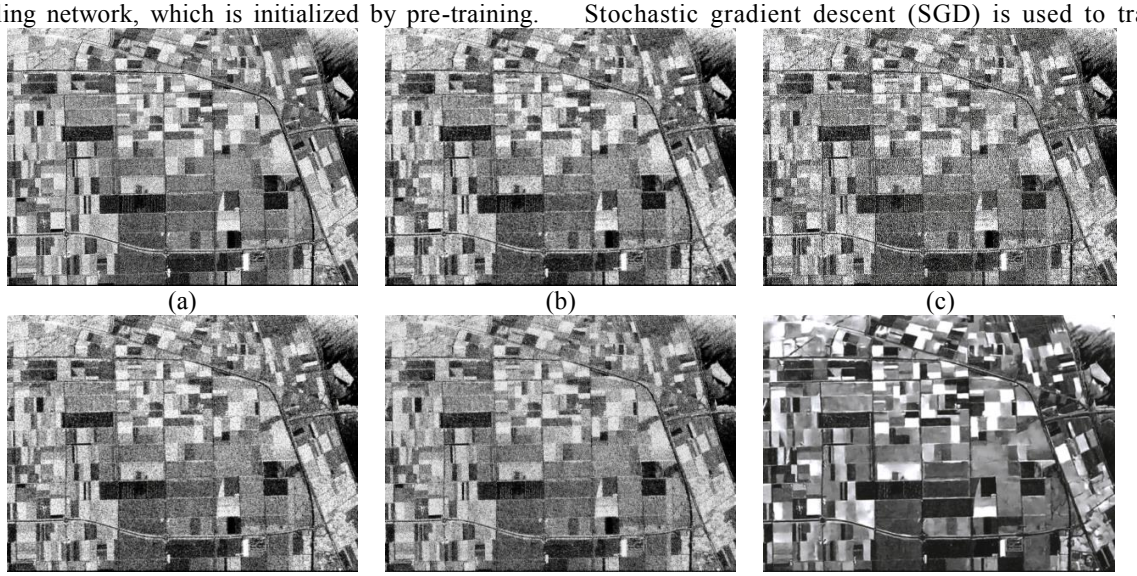

Fig. 3. Experimental results on Flevoland image. (a) Original SAR image. (b) Bicubic interpolation. (c) Pure-GAN. (d) SRCNN. (e) RDN. (f) NF-GAN.

network with weight decay of 0.0001, a momentum of 0.9 and a minibatch of 12. The training process of NF-GAN took 50 epochs (about 1500 iterations), and after every 10 epochs, the learning rate was reduced through being multiplied by a descending factor 0.5.

3. EXPERIMENTAL RESULTS

We choose three scenes of L-band SAR images of NASA/JPL AIRSAR as our experimental data. The HH channel is chosen as our experimental channel. The intensity images of HH channel are cropped into patches of size 224×224 for training and testing. Fig. 3(a) presents the testing image of Flevoland, Netherlands.

To verify the proposed method, we compared the NF-GAN with two mainstream super-resolution methods: SRCNN [12] and RDN [7]. The GAN without despeckling network is also used as a contrast method to perform the ablation experiment. In this letter, we denote it as pure-GAN. Fig. 3(b-f) shows the reconstruction results of different methods. From the results, we can see that our method produces perceptually convincing HR SAR images with finer details and better suppressed noise level. False texture arising in the result of pure-GAN confirms the necessity of the despeckling network. The results of SRCNN, RDN are greatly affected by noise and lose most of the high-frequency details.

In addition, peak signal-to-noise ratio (PSNR) (Equation 5) and structural similarity index (SSIM) (Equation 6) are used to assess the performance of different methods quantitatively. The higher value of PSNR and SSIM indicates better quality of the reconstructed.

$$PSNR = 10 \times log\left[\frac{255^2}{\frac{1}{WH}\left\|N_{\theta_G}(I^{LR}) - I^{HR}\right\|_2^2}\right] \quad (5)$$

$$SSIM = \frac{(2\mu_{I^{LR}}\mu_{I^{HR}} + c_1)(2\sigma_{I^{LR}I^{HR}} + c_2)}{(\mu_{I^{LR}}^2 + \mu_{I^{HR}}^2 + c_1)(\sigma_{I^{LR}}^2 + \sigma_{I^{HR}}^2 + c_2)} \quad (6)$$

where $\mu_{I^{LR}}$ and $\mu_{I^{HR}}$ are the averages of I^{LR} and I^{HR}. $\sigma_{I^{LR}}^2$ and $\sigma_{I^{HR}}^2$ are the variances of I^{LR} and I^{HR}. $\sigma_{I^{LR}I^{HR}}$ is the covariance between I^{LR} and I^{HR}. c_1 and c_2 are two variables to stabilize the division with the weak denominator. As can be seen from Table 1, our method is the best in both PSNR and SSIM indicators.

Table 1. Quantitative results on the testing image

	BICUBIC	SRCNN	RNSR	Pure-GAN	NF-GAN
PSNR	13.7103	15.5003	15.4829	12.6484	**16.2466**
SSIM	0.3443	0.3902	0.3565	0.2190	**0.4823**

4. CONCLUSION

We have proposed a novel deep learning approach for the SAR image super resolving task, learning an end-to-end mapping between the low-resolution and high-resolution SAR images. Differently from common convolutional operation, the presented approach is based on generative adversarial network, which can use the adversarial objective function to push the super-resolution SAR images closer to real high-resolution SAR images. Compared with the traditional CNN-based methods, the proposed NF-GAN shows a state-of-the-art performance in both quantitative analysis and visual examination. The better effect of noise suppression and SAR-realistic content retention are shown

in the contrast experiments. The NF-GAN can be extended in applying for other satellite configuration super-resolution tasks. In our future work, we will investigate more powerful learning models to deal with the complex real scenes in fully polarimetric SAR images.

5. ACKNOWLEDGMENT

This work was supported by National Key Research and Development Program of China under Grant 2016YFB0501501, the National Natural Science Foundation of China under Grants 41331176 and 41371352 and the National High-Tech Research and Development Program of China under Grant 2009AA12Z118.

6. REFERENCES

[1] L. Xin, M. T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing,* vol. 10, no. 10, pp. 1521-1527, 2001**.**

[2] Z. Lei, W. Xiaolin, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Transactions on Image Processing,* vol. 15, no. 8, pp. 2226-2238, 2006**.**

[3] M. T. Merino, J. Nunez, "Super-resolution of remotely sensed images with variable-pixel linear reconstruction," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 5, pp. 1446-1457, 2007**.**

[4] K. Zhang, X. Gao, D. Tao et al. "Multi-scale dictionary for single image super-resolution," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1114-1121, 2012.

[5] S. Jian, X. Zongben, S. Heung-Yeung "Image super-resolution using gradient profile prior," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[6] L. Liu, W. Huang, C. Wang et al. "Sar image super-resolution based on tv-regularization using gradient profile prior," in *2016 CIE International Conference on Radar (RADAR)*, pp. 1-4, 2016.

[7] Y. Zhang, Y. Tian, Y. Kong et al. "Residual dense network for image super-resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2472-2481, 2018.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al. "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672-2680, 2014.

[9] S. Jégou, M. Drozdzal, D. Vazquez et al. "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1175-1183, 2017.

[10] K. He, X. Zhang, S. Ren et al. "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[11] F. Gu, H. Zhang, C. Wang et al. "Residual encoder-decoder network introduced for multisource sar image despeckling," in *SAR in Big Data Era: Models, Methods and Applications* pp. 1-5, 2017.

[12] C. Dong, C. C. Loy, K. He et al. "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*, pp. 184-199, 2014.