

A GAN-based Method for SAR Image Despeckling

Feng Gu

Key Laboratory of Digital Earth Science,
of Institute of Remote Sensing and
Digital Earth
Chinese Academy of Science
Beijing, China
gufeng@radi.ac.cn

Hong Zhang*

Key Laboratory of Digital Earth Science,
of Institute of Remote Sensing and
Digital Earth
Chinese Academy of Science
Beijing, China
zhanghong@radi.ac.cn

Chao Wang*

Key Laboratory of Digital Earth Science,
of Institute of Remote Sensing and
Digital Earth
Chinese Academy of Science
Beijing, China
wangchao@radi.ac.cn

Abstract—Synthetic Aperture Radar (SAR) images are contaminated by multiplicative noise known as speckle. Most filtering methods are limited to noise statistics and requires complex parameter tuning to achieve the desired visual effects. To solve the above problem, a generative adversarial network (GAN) based method is proposed for SAR image despeckling. Firstly, homogeneous regions are selected manually and speckle samples are produced. Then, GAN is trained to learn the distribution of the speckle samples and generate the “realistic-looking” ones. Third, a convolutional neural network (CNN) is designed that specializes in removing the speckle. Experiments on simulated SAR images and real SAR images show the good performance of the proposed method with respect to both the visual effect and quantitative analysis.

Keywords—*SAR image despeckling, generative adversarial network, convolutional neural network*

I. INTRODUCTION

Due to the coherent interference in the returned signal from various elementary scatterers [1], SAR images are vulnerable to speckled noise, which complicates the image processing and application problems. The classic denoising algorithms are based on the spatial domain and Fourier analysis [2-3]. They have good low pass and noise resistance properties, but they also smooth the edges and textures.

In the field of optimization, image denoising tasks can be expressed as cost minimization problems and can be solved using different models, one of which is sparse representation. The sparse-based image denoising model uses the K-singular value decomposition (K-SVD) algorithm to obtain a sparse and redundant representation that effectively describes the content of a noise-free image [4]. Another branch of optimization-based denoising method is total variation (TV) regularization, which is based on the principle that fake details have big total variations, namely, the integral of the absolute gradient of the image is high.

In addition to the aforementioned methods, by searching the similarity of the image itself, the strategy of the nonlocal mean (NLM) achieved a breakthrough in image denoising and has gradually become mainstream. The NLM tries to find patterns of pixels all over the image and calculates the average over these self-similarities rather than using the pixels close to the current pixel [5]. Different algorithms have been proposed to extend the NLM to SAR image despeckling, among which the probabilistic patch-based (PPB) algorithm [6] and the SAR version of the block-matching 3D (SAR-BM3D) algorithm [7] are the most representative ones.

Although NLM filters are more effective at suppressing noise and maintaining details, they need to correctly select the search window size and other auxiliary parameters according to the noise statistics, which largely depends on the complex SAR imaging mechanism including the

backscattering correlation, antenna pattern and signal transmission medium. The same problem exists in the recently popular deep learning technologies for SAR image despeckling. The deep learning-based method owns great performance due to the decent network architecture that learns the underlying noise distribution. However, because there are no noise-free SAR images in reality, deep learning-based methods use optical photographs as simulated clean SAR images and artificially add speckled noise in advance to form paired training data, which causes one problem: how to specify the noise statistics. On the one hand, the noise level of an SAR image varies even in the same imaging mode. On the other hand, the Gamma noise model only holds in theory and the situation is more complicated in the real world.

Therefore, how to learn the complex SAR noise pattern for a specific imaging mode without human intervention is the key to fully displaying the superiority of deep learning-based methods in SAR image despeckling. Fortunately, generative adversarial networks (GANs) [8] and their improved variations [9-10] make this possible. GANs are state-of-the-art generative algorithms that can model the distribution of the given training data. In this paper, a three-step SAR denoising framework is proposed. First, homogeneous regions are selected manually and noise samples are produced. Second, a GAN is trained to model the distribution of the noise from step one. Finally, the noise that is generated from the trained GAN can be artificially added to the noise-free optical remote sensing images to produce a simulated noisy image, which can then be used to train a deep CNN for despeckling.

The remainder of this paper is organized as follows. In Section 2, the details of the methodology are described. Section 3 first introduces the experimental settings and then presents the testing results along with discussions. Section 4 concludes the paper.

II. METHOD

A. Noise generation using the GAN

Let \mathbf{Y} be the noisy SAR intensity image and \mathbf{X} be the noise-free image. The multiplicative speckle model is defined as follows:

$$Y(s) = X(s)N(s) \quad (1)$$

where N represents the speckle and s indicates the spatial location. Generally, N follows a Gamma distribution with unit mean and fixed variance $1/L$ where L is the number of looks. However, the actual L is changed in a range and not fixed even in the same SAR image which causes difficulty in noise statistics estimation and repeated parameter fine-tuning to achieve desired filtering result.

Thus, in this paper, GAN is used to model the distribution of the unknown noise and automatically generate the synthetic noise samples. Using the Logarithmic transformation, the noise can be extracted as follows:

$$\text{Log}(U(s)) = \text{Log}(\hat{Y}(s)) - \text{mean}(\text{Log}(\hat{Y}(s))) \quad (2)$$

where \hat{Y} refers to the homogeneous regions, $\text{mean}(\cdot)$ calculates the mean value and $\text{Log}(\cdot)$ represents the Logarithmic transformation. There are a large number of homogeneous areas in the high-resolution SAR images, such as the oceans, farmland, bare land and so on, which meets the requirement of the GAN for the number of training samples.

The architecture of the GAN introduces a generator G that is trained to produce “realistic-looking” samples by introducing an adversary, namely, the discriminator network D that detects if a given sample is real or fake. In our despeckling framework, this is represented as a min-max optimization as follows:

$$\min_G \max_D \mathcal{L}_{GAN}(D, G) = \mathbb{E}_{x \sim p(x)}[\log D(x)] + \mathbb{E}_{z \sim p(z)}[1 - \log D(G(z))] \quad (3)$$

$$\tilde{x} = G(z), z \sim N(0,1) \quad (4)$$

where $p(\cdot)$ represents the distributions of the data, and x and \tilde{x} are the real noise samples and noise samples that generated from the generator, respectively. \mathbb{E} denotes the expectation, and \tilde{x} is generated by the generator given a latent vector z , which obeys standard normal distribution. The structure of the GAN that is used in this paper and more details can be found in [8]. Because the basic components of the GAN are consistent with the CNN, the loss function \mathcal{L}_{GAN} can be trained using the prevailing back-propagation algorithm.

The sketched map of the GAN is depicted in Figure 1. At the beginning of the training process, G has not learned anything about the noise distribution. Thus, the noise samples generated by G can be easily identified by D , namely, $D(G(z))=0$. As the number of anti-iterative training of G and D is increased, the capabilities of G and D will be simultaneously enhanced and the fake noise samples will become increasingly realer. Finally, D cannot distinguish between true and false noise samples. This means that the mapping from the random normal input z to the synthetic noise sample is constructed.

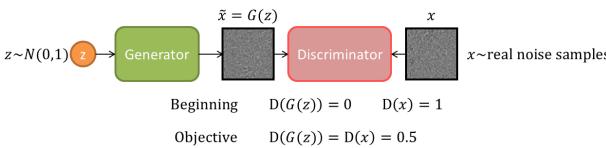


Fig. 1. Sketched map of the GAN.

B. Noise removal using the CNN

There are many CNN structures that have been introduced in SAR image denoising. In our previous research [11], we also designed a network called residual encoder-decoder network (RED-NET) for multisource SAR image filtering. CNN owns a great ability to learn the high-level features of the images and has been proved to be effective at suppressing the additive Gaussian noise. However, the traditional pair training strategy limits its application in SAR despeckling, which needs to fix the

variance of the added noise while restricting the denoising ability of the CNN model.

With the noise samples that are generated from the GAN, this problem can be easily solved. If we define the generated noise sample set from Section II.A as $X = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n\}$, and the noise-free optical image patch set as $Y = \{y_1, y_2, \dots, y_n\}$, then the training pair $\{X, Y\}$ can directly train the CNN model. Note that the order of the combinations of X and Y can be changed arbitrarily, which further augments the training data. Without the loss of generality, this paper uses the residual encoder-decoder network (RED-NET) as the CNN model, the architecture of which is presented in Figure 2.

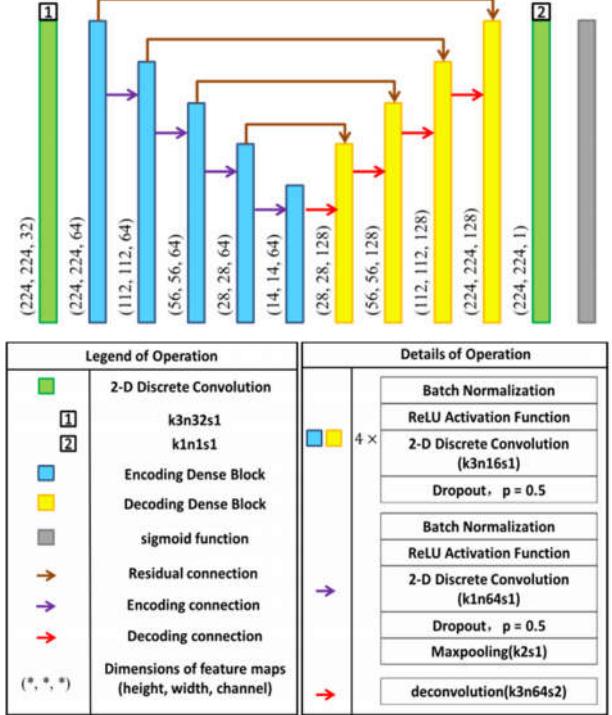


Fig. 2. The architecture of the RED-NET. k denotes the kernel size, n denotes the number of output channels and s denotes the stride.

RED-NET is fully convolutional and consists of a downsampling path (blue part in Figure 2) and an upsampling path (yellow part in Figure 2). The downsampling path follows a typical CNN paradigm, consisting of repeated convolutions, batch normalization, the rectified linear unit (ReLU) and the dropout layer [12]. In each downsampling step, the maxpooling operation with a stride of 2 will reduce the spatial size of the input while the number of output feature channels is unchanged. The purpose of downsampling is to contract the spatial information and extract the high-level content features, which are beneficial to smoothing the image, but the details of the ground objects are lost. Therefore, when restoring the resolution of the output using the deconvolution, the downsampling pathway will also combine the feature and spatial information using a sequence of residual connections, which concatenates the corresponding feature maps in the downsampling path to supplement the lack of details. The per-pixel Euclidean loss is used to train the RED-NET network as follows:

$$\text{loss}(\theta) = \frac{1}{2WH} \| \text{net}_\theta(x) - y \|_2^2 \quad (5)$$

where x and y denote the noisy input image and the noise-free label image of size $W \times H$ respectively. θ

denotes the weights that the RED-NET (net_θ) needs to update to generate the despeckled output $net_\theta(x)$.

III. EXPERIMENTAL RESULTS

Experiments are carried out on two simulated SAR images and two real SAR images. The results are compared with the state-of-the-art methods, including PPB [6], SAR-BM3D [7], and POTDF [13]. The executable program of the method is downloaded from the author's official home page and the default parameters are used in the corresponding code.

A. Experimental dataset and parameter Settings

Gaofen-3 (GF-3), the first Chinese civil C-band SAR, circles the Earth in a sun-synchronous orbit with an altitude of 755 km and a 29-day revisit period [14]. Among its 12 imaging modes, the Full polarized Strip I (QPSI) mode collects SAR images over a large coverage area with a high resolution of 8 m [15]. We choose four images from the QPSI mode for the experiment (HH polarization), among which two images are used to generate the real noise samples while the others are chose for testing the performance of the different despeckling methods.

Considering the fact that there is no noise-free SAR image in reality, to force the manifold distribution of the simulated SAR images closer to that of real SAR images, the NWPU-RESISC45 dataset [16] is used as our clean SAR image dataset. Combined with the noise from the GAN, it makes up the training dataset to train the RED-NET. The NWPU-RESISC45 dataset is a public dataset for remote sensing image scene classification that includes over 30,000 images covering 45 land cover types (about 700 images in each class), and it is used for training the RED-NET.

For the GAN, the input size is set to 64×64 , and the learning rates of the generator and the discriminator are both $1e-4$. Other specific training details completely follow literature [8]. For training the RED-NET, the “Xavier” algorithm [17] is used to initialize the weights and the input size is set to 224×224 . A total of 31,500 image patches are collected for training the RED-NET with a batch size of 24. The learning rate of the RED-NET is exponentially decayed every 15 training epochs from $1e-4$ to $1e-6$. The training process of RED-NET takes 50 epochs with 1,300 iterations. Adam [18] is used as the gradient descent optimization method to train both the GAN and RED-NET.

B. Experiments on Simulated SAR Images

Two images from the NWPU-RESISC45 dataset, *palace* and *parking lot*, are selected to produce the simulated SAR images by adding noise from the GAN. Two commonly used criteria were selected to evaluate the denoising performance: the Structural Similarity Index (SSIM) and the peak signal-to-noise ratio (PSNR) [19]. These two metrics are defined as follows:

$$PSNR = 10\log_{10} \left[\frac{\max(\tilde{X}^2)}{\frac{1}{|S|} \sum_{s \in S} [\tilde{X}(s) - X(s)]^2} \right] \quad (6)$$

$$SSIM = \frac{(2\mu_X\mu_{\tilde{X}} + c_1)(2\sigma_{X\tilde{X}} + c_2)}{(\mu_X^2 + \mu_{\tilde{X}}^2 + c_1)(\sigma_X^2 + \sigma_{\tilde{X}}^2 + c_2)} \quad (7)$$

where s is the coordinator that indicates the spatial location. μ_X and $\mu_{\tilde{X}}$ are the average intensities of the clean image X and the despeckled image \tilde{X} , while σ is their variance. c_1

and c_2 are default constants. The PSNR is an approximation index for human perception and a higher PSNR generally indicates that the denoised image is higher quality. The SSIM is a method for measuring the structural similarity between two images, the range of which is $[0, 1]$. The closer the SSIM is to 1, the better the reconstruction performance will be.

Figures 3 and 4 show the visual despeckled results for *palace* and *parking lot*, respectively. The evaluation indexes of the despeckled images are given in Table I. It can be clearly seen that the proposed method achieves better denoising performance in smoothing the noise and maintaining the details. The PPB oversmooths the image and the structure of the palace, and the indicator line of the parking lot has been erased. Although the details are kept in the result of the POTDF, there are still fake pepper-and-salt particles in the image. The result of the SAR-BM3D is relatively good, but the details are not well preserved. We can observe that the indicator arrow is incomplete and the indicator line is discontinuous in the SAR-BM3D result. As shown in Table I, the proposed method outperforms the PPB, SAR-BM3D, and POTDF on both the SSIM and PSNR. The above quantitative and qualitative assessments show the superiority of the proposed method on simulated SAR images.

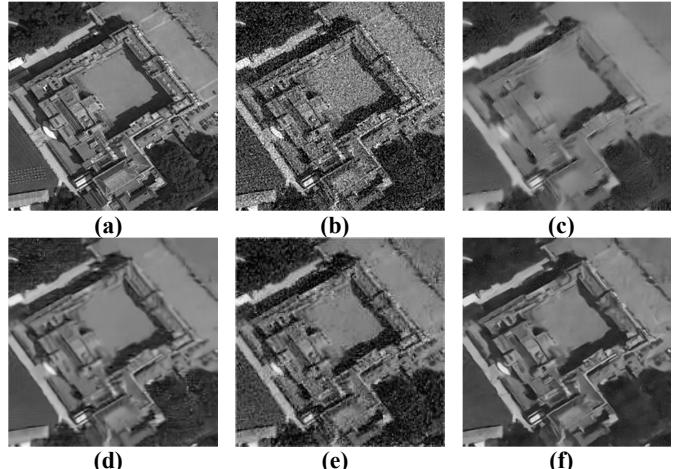


Fig. 3. Despeckling results for the *Palace* image. (a) Original image. (b) Image corrupted by speckled noise. The despeckled images that were obtained by (c) the PPB, (d) the SAR-BM3D, (e) the POTDF, and (f) the proposed method.

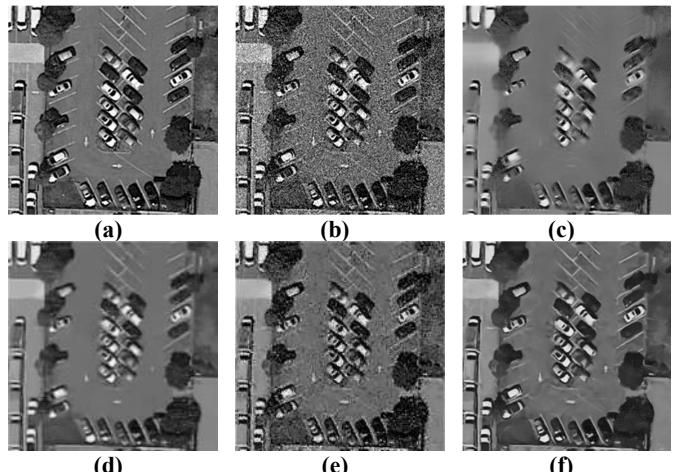


Fig. 4. Despeckling results for the *parking lot* image. (a) Original image. (b) Image corrupted by speckled noise. The despeckled images that were obtained by (c) the PPB, (d) the SAR-BM3D, (e) the POTDF, and (f) the proposed method.

TABLE I. SSIM AND PSNR OF THE DESPECKLED SIMULATED SAR IMAGES USING DIFFERENT METHODS

| Methods | | PPB | SAR BM3D | POTDF | Proposed method |
|--------------------|------|---------------|---------------|---------------|-----------------|
| <i>palace</i> | SSIM | 0.5397 | 0.6597 | 0.6234 | 0.6937 |
| | PSNR | 20.38 | 22.92 | 22.30 | 24.01 |
| <i>parking lot</i> | SSIM | 0.6025 | 0.7080 | 0.6357 | 0.7376 |
| | PSNR | 20.32 | 22.34 | 21.11 | 23.87 |

C. Experiments on Real SAR Images

In this section, we conduct experiments on two real SAR images. To view the despeckling results, two subimages are cropped from the original images to check the details. Since noise-free SAR images are not accessible, the equivalent number of looks (ENL), which indicates the ability of a filter over homogenous areas, is used to analyze the noise suppression ability of the different methods for real SAR images. The ENL is defined as follows:

$$ENL = \left(\frac{\mu_{region}}{\sigma_{region}} \right)^2 \quad (8)$$

where μ_{region} and σ_{region} are the mean and standard deviation of the selected region, respectively. Figure 5 shows the two subimages of size 1024×1024 and the three selected homogeneous regions.

The despeckled results of the two real SAR subimages (#1 and #2) using the different methods are shown in Figures 6 and 7. The estimated ENL values from three selected homogeneous regions are listed in Table II. Visually, consistent with the simulation experiment, the PPB and SAR-BM3D are obviously oversmooth. In addition, the POTDF still suffers from the pixel distortion that appears in the form of black speckles. From Table II, we can find that the highest ENL is obtained by PPB algorithm due to its excessive suppression of speckled noise. The POTDF filter performs poor in the selected regions. The proposed method as well as the SAR-BM3D has relatively strong noise removal ability. Meanwhile, they also provide sharper results with respect to edges and textures.

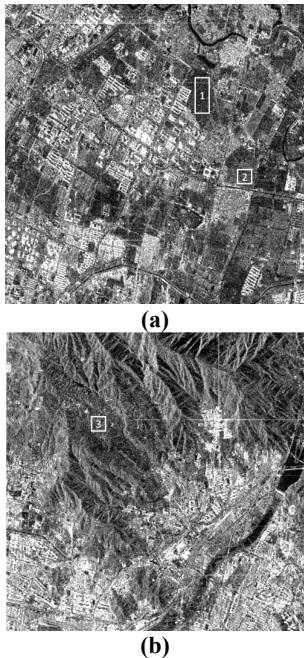


Fig. 5. Two real SAR subimages. The rectangles represent the test homogeneous blocks.

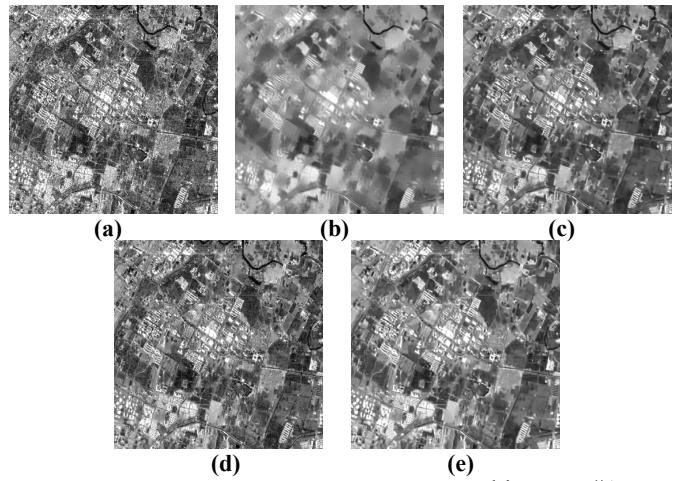


Fig. 6. Despeckling results on the real SAR subimages #1. (a) Original image. The despeckled images that were obtained by (b) the PPB, (c) the SAR-BM3D, (d) the POTDF, and (e) the proposed method.

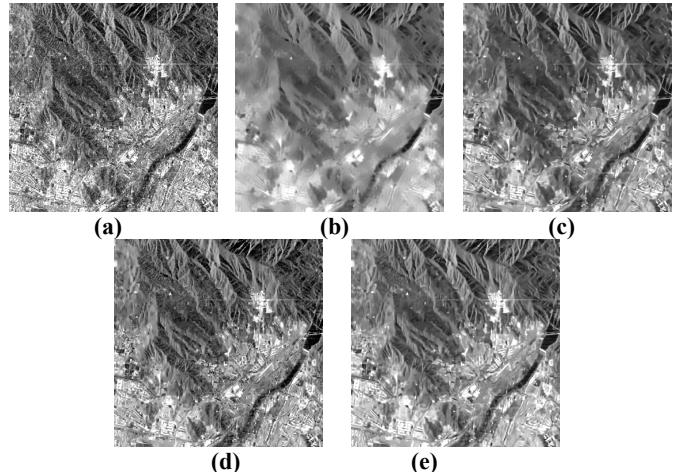


Fig. 7. Despeckling results on the subimage of #2. (a) Original image. The despeckled images that were obtained by (b) the PPB, (c) the SAR-BM3D, (d) the POTDF, and (e) the proposed method.

TABLE II. ENLS OF THE DESPECKLED REAL SAR IMAGES USING DIFFERENT METHODS

| | PPB | SAR BM3D | POTDF | Proposed method |
|----------|---------------|--------------|--------------|-----------------|
| region 1 | 65.51 | 42.23 | 18.38 | 52.65 |
| region 2 | 169.84 | 62.12 | 23.41 | 86.07 |
| region 3 | 236.31 | 87.90 | 22.91 | 105.72 |

IV. CONCLUSION

This paper proposes a novel GAN-based framework for the speckle reduction in SAR images. By manually selecting the homogeneous regions, the preliminary noise samples can be obtained. Then, a GAN is introduced to learn the noise distribution and generate a pseudo noise sample that looks like real noise. Third, the RED-NET is applied to form an end-to-end mapping between the noisy and clean SAR images. Based on GAN, the noise distribution can be estimated without human intervention. Once the noise is produced by GAN, the training pairs that produced using the synthetic noise can directly supervise the training of the RED-NET. The trained RED-NET can be applied to all intensity images of the selected imaging modes and does not need to fine tune the parameters to achieve better results. Experiments are conducted on both the simulated SAR images and real SAR images. Compared with the traditional despeckling methods, the proposed framework shows better

performance in speckle suppression and detail preservation. In addition, because the algorithm is conducted on GPU, the proposed method has the characters of a short running time, which meets the demands of real-time data processing.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grants 41331176 and 41371352.

REFERENCES

- [1] J. S. Lee, T. L. Ainsworth, and Wang Y, "A review of polarimetric SAR speckle filtering," *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5303-5306, 2017.
- [2] J. S. Lee, "Refined filtering of image noise using local statistics," *Computer Graphics and Image Processing*, vol. 15, no. 4, pp. 380-389, 1981.
- [3] V. S. Frost, J. A. Stiles, K. S. Shanmugan *et al.*, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, no. 2, pp. 157-166, 1982.
- [4] M. Elad, M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736-3745, 2006.
- [5] C. Deledalle, L. Denis, F. Tupin *et al.*, "NL-SAR: A unified nonlocal framework for resolution-preserving (pol)(in)SAR denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2021-2038, 2015.
- [6] C. Deledalle, L. Denis, F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2661-2672, 2009.
- [7] S. Parrilli, M. Poderico, C. V. Angelino *et al.*, "A nonlocal SAR image denoising algorithm based on lmmse wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 606-616, 2012.
- [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] M. Arjovsky, S. Chintala, L. Bottou, "Wasserstein GAN," arXiv preprint arXiv:1701.07875, 2017.
- [10] I. Gulrajani, F. Ahmed, M. Arjovsky *et al.* "Improved training of wasserstein GANs," in *Advances in Neural Information Processing Systems*, pp. 5767-5777, 2017.
- [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] G. Huang, Z. Liu, L. Van Der Maaten *et al.* "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708, 2017.
- [13] B. Xu, Y. Cui, Z. Li *et al.*, "Patch ordering-based SAR image despeckling via transform-domain filtering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 4, pp. 1682-1695, 2015.
- [14] Z. Qingjun, "System design and key technologies of the GF-3 satellite," *Acta Geodaetica et Cartographica Sinica*, vol. 46, no. 3, pp. 269-277, 2017.
- [15] T. Wang, G. Zhang, L. Yu *et al.*, "Multi-mode GF-3 satellite image geometric accuracy verification using the rpc model," *Sensors*, vol. 17, no. 9, pp. 2005, 2017.
- [16] G. Cheng, J. Han, X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865-1883, 2017.
- [17] X. Glorot, Y. Bengio "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249-256, 2010.
- [18] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [19] Z. Wang, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.