Abstract—In this letter, we aim to address synthetic aperture radar (SAR) despeckling problem with the necessity of neither clean (speckle-free) SAR images nor independent speckled image pairs from the same scene, a practical solution for SAR despeckling (PSD) is proposed. Firstly, to generate speckled-to-speckled (S2S) image pairs from the same scene in the situation of only single speckled SAR images are available, an adversarial learning framework is designed. Then, the S2S SAR image pairs are employed to train a modified despeckling Nested-UNet model using the Noise2Noise (N2N) strategy. Moreover, an iterative version of the PSD method (PSDi) is also proposed. The performance of the proposed methods is demonstrated by both synthetic speckled and real SAR data. SAR block-matching 3-D algorithm (SAR-BM3D) and SAR dilated residual network (SAR-DRN) are used in the visual and quantitative comparison. Experimental results show that the proposed methods can reach a good tradeoff between speckle suppression and edge preservation.

Index Terms—Synthetic aperture radar (SAR), image despeckling, adversarial learning, Noise2Noise (N2N), Nested-UNet.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) images are usually corrupted with speckle noise, which leads to the degradation of image quality and affects the performance in various applications of remote sensing, e.g., classification [1] and change detection [2]. Several methods have been proposed to mitigate the speckle in SAR images, including filtering methods [3], [4], wavelet shrinkage [5], [6], SAR block-matching 3-D algorithm (SAR-BM3D) [7].

Recently, convolutional neural network (CNN) based supervised learning has been employed for SAR image despeckling, which can reduce speckle noise by learning relationships between speckled and clean ground truth images with CNN models, e.g., SAR convolutional neural network (SAR-CNN) [8], SAR image despeckling network (SAR-IDN) [9], and SAR dilated residual network (SAR-DRN) [10]. However, there are few clean SAR images in practical applications. Hence, speckle-free optical images with the single channel are usually employed as the clean ground truth for SAR image despeckling, and synthetic speckled images can be obtained by adding speckle noise to original images. Then, such speckled-to-clean image pairs can be used as training data in CNN despeckling models. However, because of the differences in imaging mechanism and image features of SAR and optical images, i.e., gray-level distribution and spatial correlation [11], it is not the optimal solution to achieve SAR image despeckling by directly employing the aforementioned CNN models, which are trained on optical images using supervised learning.

More recently, the Noise2Noise (N2N) [12] strategy has shown its ability for image denoising without using any clean ground truth images. By employing the N2N strategy, CNN model can still achieve high denoising performance (e.g., zero-mean distribution noise) with mean square error (MSE) loss, as long as two independent noisy images from the same underlying clean image are available. This provides the possibility for SAR image despeckling without using clean ground truth images. However, to acquire two independent speckled SAR images from same scene is quite difficult. Hence, it is still impossible to solve the SAR image despeckling problem by merely employing the N2N strategy, as only single speckled SAR images are available.

In this letter, we introduce a more practical despeckling solution for SAR image in the situation of only single speckled images are available, namely, a practical solution for SAR despeckling (PSD). Our main contributions are summarized as follows:

1) To generate speckled-to-speckled (S2S) SAR image pairs, an adversarial learning framework that consists of two generators and a discriminator is presented, which are trained by an alternating optimization strategy.
2) By using the obtained S2S SAR image pairs, an advanced Nested-UNet model [13] is trained to achieve despeckling with the N2N strategy. In addition, an iterative version of the PSD method (PSDi) is proposed, which can further improve the despeckling performance.
3) Visual and quantitative experiments conducted on synthetic speckled and real SAR data show that the proposed methods notably suppress speckle noise with better preserving edge, which outperform SAR-BM3D and SAR-DRN.

II. METHODOLOGY

In this section, firstly, the SAR speckle noise model is introduced. Then, we present implement details of the proposed PSD method as shown in Fig. 1 which consists of two parts: 1) an adversarial learning framework for S2S SAR image
Adversarial learning for S2S SAR image pairs generation

| Input: Speckled Images Y | Random Speckle N |
|-------------------------|------------------|
| \( g_\theta (Y) \) | \( Y \) |
| \( L_{TV} (\theta_1) \) | \( g_\theta (Y) \times N \) |
| \( L_{cyc} (\theta_1, \theta_2) \) | \( \hat{Y} = g_\theta (Y) + N \) |
| And the network architecture of \( g_\theta \) is the same as the de-speckling model \( g_{\theta_d} \), which will be described in Section II-C. |

\[ Y = N \odot X, \] (1)

where \( \odot \) denotes the Hadamard product (i.e., entry-wise product) of two matrices. It is well-known that, for a SAR image, \( N \) often follows a Gamma distribution with unit mean and variance \( \frac{1}{L} \), where \( L \) is the number of looks in SAR imaging process, where a smaller \( L \) indicates stronger speckle. The aim of despeckling is to estimate \( X \) from the observed image \( Y \). Based on the established speckle noise model in (1), our proposed approach in Section II-B can be used to generate S2S SAR image pairs from same scenario.

\[ \hat{Y} = g_{\theta_1} (Y) \odot N, \] (2)

B. Adversarial learning for S2S SAR image pairs generation

In this section, to generate S2S SAR image pairs, we introduce an adversarial learning framework, which consists of two generators \( g_{\theta_1}, g_{\theta_2} \), and a discriminator \( f_\omega \), as shown in Fig. 1 (left). \( \theta_1, \theta_2 \) and \( \omega \) denote the parameters (i.e., weights and biases) of \( g_{\theta_1}, g_{\theta_2} \) and \( f_\omega \), respectively.

The generator \( g_{\theta_1} \) is used to generate the “fake” speckled SAR images \( \hat{Y} \), expressed as follows

\[ \hat{Y} = g_{\theta_1} (Y) \odot N, \] (2)

and the network architecture of \( g_{\theta_1} \) is the same as the de-speckling model \( g_{\theta_d} \), which will be described in Section II-C.

The overall flowchart of the proposed PSD method. (left) Adversarial learning for S2S SAR image pairs generation: two generators \( g_{\theta_1}, g_{\theta_2} \), and a discriminator \( f_\omega \) are trained by using the adversarial loss \( L_{\text{WGAN}} \), the backward cycle consistency loss \( L_{\text{cyc}} \), and the total variation (TV) loss \( L_{\text{TV}} \). (right) N2N despeckling without clean ground truth images: the despeckling model \( g_{\theta_d} \) is trained by using the MSE loss \( L_{\text{MSE}} \).

In addition, to smooth the output of \( g_{\theta_1} \), the generator \( g_{\theta_1} \) is also trained by the total variation (TV) loss \( L_{\text{TV}} \), which is defined as

\[ L_{\text{TV}} (\theta_1) = \sum_{w=1}^{W-1} \sum_{k=1}^{H-1} \sqrt{|g_{\theta_1} (Y)_{w+1,k} - g_{\theta_1} (Y)_{w,k}|^2 + |g_{\theta_1} (Y)_{w,k+1} - g_{\theta_1} (Y)_{w,k}|^2}, \] (5)

where \((g_{\theta_1} (Y))_{w,h}\) is the pixel value in \( g_{\theta_1} (Y) \). \( L_{\text{TV}} \) can reduce the difference of adjacent pixel values in the despeckled images \( g_{\theta_1} (Y) \).

With the defined loss functions, an alternating optimization strategy is applied to optimize \( g_{\theta_1}, g_{\theta_2} \) and \( f_\omega \), which can be described as an adversarial min-max problem, expressed as follows

\[ \min_{\theta_1, \theta_2} \max_\omega \left[ \alpha L_{\text{WGAN}} (\theta_1, \omega) + \beta L_{\text{cyc}} (\theta_1, \theta_2) + \gamma L_{\text{TV}} (\theta_1) \right], \] (6)

where \( \alpha, \beta, \) and \( \gamma \) are the predefined weights for \( L_{\text{WGAN}}, L_{\text{cyc}} \) and \( L_{\text{TV}} \), respectively. To prevent \( g_{\theta_1} (Y) \) from being over-smooth, the predefined weight \( \gamma \) for \( L_{\text{TV}} \) should be far less than \( \alpha \) and \( \beta \). After reaching a steady state through adversarial learning, i.e., until the discriminator \( f_\omega \) cannot
C. N2N despeckling without clean ground truth images

After obtaining the S2S SAR image pairs \( \hat{Y}_1, \hat{Y}_2 \), as shown in Fig. 1 (right), we can employ the N2N strategy \([12]\) to train the despeckling model \( g_{\theta_d} \) without using any clean ground truth images. Here, MSE loss is used to optimize \( g_{\theta_d} \), which is formulated as

\[
L_{\text{MSE}}(\theta_d) = \| \hat{Y}_2 - g_{\theta_d}(\hat{Y}_1) \|_2^2,
\]

where \( \theta_d \) denotes the parameters (i.e., weights and biases) of \( g_{\theta_d} \).

Due to the characteristic of speckle noise, as expressed in \([1]\), the expected value of speckled image is the same as that of underlying clean (despeckled) image, hence, the despeckled images can be generated by using \( g_{\theta_d} \) without any clean ground truth images. This is particularly true when the despeckling model \( g_{\theta_d} \) is trained on a large dataset. A modified Nested-UNet \([13]\) is adopted as the despeckling model \( g_{\theta_d} \) as shown in Fig. 2. Nested-UNet is chosen because of its advantage in improving the gradients flow throughout the network by its convolution layers on skip pathways and dense skip connections on skip pathways. The original Nested-UNet is polished to be more effective for SAR image despeckling by removing batch normalization layers of each convolutional block, which has been proven in other image processing tasks, e.g., image super-resolution \([18]\).

Except the above PSD method, we also propose an iterative version of PSD method, named PSDi. For PSD method, the training data (the input and ground truth images) of \( g_{\theta_d} \), i.e., S2S SAR image pairs \( \hat{Y}_1, \hat{Y}_2 \), are generated by using \( g_{\theta_i}(Y) \). However, since no ground truth (clean or speckled) images are available in the training process of \( g_{\theta_i} \), the despeckling ability of \( g_{\theta_i} \) is not decent. Compared with \( g_{\theta_i} \), \( g_{\theta_d} \) can provide more powerful despeckling ability since ground truth (speckled) images are used in its training process. Therefore, to make the distribution of generative S2S SAR image pairs closer to that of real SAR images, we apply \( g_{\theta_d} \) instead of \( g_{\theta_i} \) in the proposed PSDi. Then, the new S2S image pairs \( \hat{Y}_1, \hat{Y}_2 \) can be obtained as follows

\[
(\hat{Y}_1, \hat{Y}_2) = (g_{\theta_d}(Y) \circ N_1, g_{\theta_d}(Y) \circ N_2),
\]

where \( N_1 \) and \( N_2 \) are two independent speckle matrices with the same number of looks \( L \).

III. EXPERIMENTS AND RESULTS

A. Experimental setup

In this work, 2.2 × 10^4 1-look SAR image patches (cropped to 96×96 pixels) obtained from the Sentinel-1 are used to train the proposed networks. In the training process of the adversarial learning (Section II-B), to keep the noise level in the original 1-look SAR images \( Y \) consistent with that in the generated speckled images \( \hat{Y} \), the number of looks \( L \) for each generated speckled image is set to be 1. In the training process of the N2N despeckling (Section II-C), to make the network available to different speckle levels, the number of looks \( L \) for each training image pair \( (\hat{Y}_1, \hat{Y}_2) \) is randomly set to be 1, 2, 4, 8 and 16, respectively. The predefined weights \( \alpha, \beta, \) and \( \gamma \) are set to be 1, 1 and 0.1, respectively. The networks \( g_{\theta_1}, g_{\theta_2}, g_{\theta_3}, \) and \( g_{\theta_4} \) are trained using the Adam optimizer by setting \( \beta_1 = 0.9, \beta_2 = 0.999, \) and \( \epsilon = 10^{-8} \), where the learning rates are all initialized as 10^{-4} and reduced by half with each 8 epochs. The network \( f_{\omega} \) is trained using the RMSProp optimizer with an initial learning rate of 5 × 10^{-5}, and its clipping value is set as 0.02, and the critic iteration is

---

**TABLE I**

| The number of looks | 1     | 4     | 16    |
|---------------------|-------|-------|-------|
| **Index**           | **PSNR** | **SSIM** | **PSNR** | **SSIM** | **PSNR** | **SSIM** |
| Speckled Input      | 11.08  | 0.0519 | 15.17  | 0.1261  | 20.73   | 0.3052   |
| SAR-BM3D \([7]\)    | 16.99  | 0.1799 | 19.88  | 0.2525  | 22.24   | 0.3604   |
| SAR-DRN \([10]\)    | 25.25  | 0.5831 | 28.39  | 0.7143  | 29.77   | 0.8061   |
| PSD (Proposed)      | 25.47  | 0.7023 | 28.45  | 0.7571  | 30.04   | 0.8200   |
| PSDi (Proposed)     | 25.67  | 0.7196 | 28.67  | 0.7763  | 29.92   | 0.8284   |

---

Fig. 3. Illustration of despeckling results on synthetic speckled image corrupted by 1-look speckle. Full results (top row), Magnified results (bottom row). (a) Speckle-free image, (b) Speckled image, (c) SAR-BM3D \([7]\), (d) SAR-DRN \([10]\), (e) PSD (proposed), (f) PSDi (proposed).
5. The training process is set to have 32 epochs with mini-batch size 16 in the adversarial learning (Section II-B) and the N2N despeckling (Section II-C).

To verify despeckling effectiveness of the proposed methods $g_0(d)$ and PSDi ($g_{di}(d)$), experiments are conducted on both synthetic speckled data and real SAR data, as compared with the state-of-the-art methods, i.e., SAR-BM3D [7] and SAR-DRN [10]. For SAR-BM3D, it needs to know the number of looks $L$ without any additional training, whereas SAR-DRN needs clean ground truth images for training, but does not need to know $L$. Hence, in the training process of SAR-DRN, around $2.2 \times 10^4$ speckle-free optical images from ImageNet [19] are used as the ground truth images, and the input speckled images can be obtained by adding multiplicative speckle noise to original images. Other parameters in SAR-DRN are set as suggested in [10].

### B. Performance comparison on synthetic speckled data

In this experiment, 100 optical remote images (selected from the AID dataset [20]) are used for analyzing. To verify the despeckling effectiveness with the known noise level, the number of looks $L$ of these images is set as 1, 4 and 16, respectively. Since the clean ground truth images are available, the indices of peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [21] are used for evaluation. The comparison results are given in Table I. As can be seen from Table I our proposed PSD and PSDi methods outperform SAR-BM3D and SAR-DRN in all noise levels. Moreover, the PSDi method can further improve the despeckling performance compared with the PSD method. Fig. 5 shows a despeckling sample on an AID image corrupted by 1-look speckle. As can be seen, only our proposed PSD and PSDi methods can remove speckle effectively. Note that SAR-DRN needs clean ground truth images for training, whereas our proposed methods can achieve better performance without using clean ground truth images.

### C. Performance comparison on real SAR data

To validate the despeckling effectiveness of proposed methods in practical applications, we also conduct a despeckling experiment on real SAR data. Two real SAR images are employed for test. One is an agriculture area image nearby Brazil which is obtained by the ALSO-2 sensor (L-band) and corrupted by 2-look speckle. The other is an aircraft display image corrupted by 1-look speckle and courtesy of Sandia National Laboratories, Radar ISR (Ku-band). In this experiment, the images are cropped to $512 \times 512$ and $1024 \times 1024$ pixels, respectively. The despeckling results are given in Fig. 4.

To evaluate the degree of speckle reduction, the equivalent number of looks (ENL) [22] is used as the performance metric. Table II gives the comparison results in terms of ENL value, which are calculated from several homogeneous areas identified by the red boxes in Fig. 4 (a). To evaluate the edge preservation, the edge-preservation degree based on the ratio of average (EPD-ROA) [22] is used as the measurement. Table III gives the comparison results in terms of EPD-ROA value, which are calculated from several edge areas identified by the green boxes in Fig. 4 (a).

From Table II and Table III we can see though SAR-BM3D performs well in speckle reduction, the results are over-smooth in despeckling performance. SAR-DRN does not

---

**Table II**

| Data            | Original | SAR-BM3D [7] | SAR-DRN [10] | PSD (proposed) | PSDi (proposed) |
|-----------------|----------|--------------|--------------|----------------|-----------------|
| ALSO-2 Area 1   | 5.4501   | 13.7571      | 13.2693      | 14.8630        | 15.2900         |
| ALSO-2 Area 2   | 5.8544   | 23.4989      | 24.8332      | 26.7084        | 27.2935         |
| ALSO-2 Area 3   | 6.2657   | 23.8957      | 19.1659      | 25.6111        | 25.9474         |
| Sandia Area 1   | 7.7393   | 21.6378      | 18.8353      | 23.0581        | 23.5764         |
| Sandia Area 2   | 11.4771  | 112.2791     | 54.4402      | 114.6707       | 116.5001        |

**Table III**

| Data            | SAR-BM3D [7] | SAR-DRN [10] | PSD (proposed) | PSDi (proposed) |
|-----------------|--------------|--------------|----------------|-----------------|
| ALSO-2 Area 4   | 0.8802       | **0.8838**   | 0.8836         | **0.8838**      |
| ALSO-2 Area 5   | 0.8709       | 0.8728       | 0.8729         | **0.8729**      |
| Sandia Area 3   | 0.9103       | **0.9140**   | 0.9102         | 0.9104          |
| Sandia Area 4   | 0.9075       | **0.9103**   | 0.9076         | 0.9077          |

---

Fig. 5. The tradeoff between the average EPD-ROA and the average ENL.
show high speckle suppression ability on real SAR images like that on synthetic speckled images, this may because of the differences between the training data (synthetic speckled images) and the testing data (real SAR images). In summary, as shown in Fig. 5 the proposed and methods achieve overwhelming despeckling while preserving great edge features, as compared with SAR-BM3D and SAR-DRN methods. Without clean ground truth images that are used to train the despeckling network, our proposed methods still can achieve better performance while only single speckled SAR images are available. This means our proposed methods are more useful in practical applications.

IV. CONCLUSION

In this letter, we presented practical deep learning-based methods for SAR image despeckling. Different from other deep learning-based despeckling methods, our proposed methods can achieve despeckling with only single speckled SAR images. In addition, a modified despeckling Nested-UNet is used as the despeckling model. Experiments conducted on both synthetic speckled data and real SAR data demonstrated the superiority of our proposed methods, as compared with state-of-the-art methods in conventional and deep learning-based, respectively. In our future study, we will apply our proposed methods for the polarimetric SAR image despeckling.

REFERENCES

[1] M. Masoud, S. Bahram, and M. Fariba, “The effect of PolSAR image despeckling on wetland classification: Introducing a new adaptive method,” Can. J. Remote Sens., vol. 43, no. 5, pp. 485–503, 2017.
[2] R. Wang, J. Chen, L. Jiao, and M. Wang, “How can despeckling and structural features benefit to change detection on bitemporal SAR images?” Remote Sens., vol. 11, no. 4, p. 421, 2019.
[3] J. Lee, “Digital image enhancement and noise filtering by use of local statistics,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-2, no. 2, pp. 165–168, 1980.
[4] D. Kuan, A. Sawchuk, T. Strand, and P. Chavel, “Adaptive noise smoothing filter for images with signal-dependent noise,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-7, no. 2, pp. 165–177, 1985.
[5] S. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” IEEE Trans. Image Process., vol. 9, no. 9, pp. 1532–1546, 2000.
[6] H. Li, W. Hong, Y. Wu, and P. Fan, “Bayesian wavelet shrinkage with heterogeneity-adaptive threshold for SAR image despeckling based on generalized gamma distribution,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 4, pp. 2388–2402, 2013.
[7] S. Parrilli, M. Poderico, C. Angelino, and L. Verdoliva, “A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 2, pp. 606–616, 2012.
[8] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, “SAR image despeckling through convolutional neural networks;” in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2017, pp. 5438–5441.
[9] P. Wang, H. Zhang, and V. Patel, “SAR image despeckling using a convolutional neural network,” IEEE Signal Process. Lett., vol. 24, no. 12, pp. 1763–1767, 2017.
[10] Q. Zhang, Q. Yuan, J. Li et al., “Learning a dilated residual network for SAR image despeckling,” Remote Sens., vol. 10, no. 2, p. 196, 2018.
[11] D. DiMartino, M. Poderico, G. Poggi et al., “Benchmarking framework for SAR despeckling,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 3, pp. 1596–1615, 2013.
[12] J. Lehtinen, J. Munkberg, J. Hasselgren et al., “Noise2Noise: Learning image restoration without clean data;” in Proc. Int. Conf. Mach. Learn., 2018, pp. 2965–2974.
[13] Z. Zhou, M. Siddiquiee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in Proc. Int. Workshop Deep Learn. Med. Image Anal. Int. Workshop Multimodal Learn. Clin. Decis. Support, 2018, pp. 3–11.
[14] S. Cha, T. Park, and T. Moon, “GAN2GAN: Generative noise learning for blind image denoising with single noisy images,” arXiv preprint arXiv:1905.10488, 2019.
[15] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in Proc. Int. Conf. Mach. Learn., 2017, pp. 214–223.
[16] K. Zhang, W. Zuo, Y. Chen et al., “Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. on Image Process., vol. 26, no. 7, pp. 3142–3155, 2017.
[17] J. Zhu, T. Park, P. Isola, and A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks;” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2223–2232.
[18] B. Lim, S. Son, H. Kim et al., “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. Workshop, 2017, pp. 1132–1140.
[19] O. Russakovsky, J. Deng, H. Su et al., “ImageNet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.
[20] G. Xia, J. Hu, F. Hu et al., “AID: A benchmark data set for performance evaluation of aerial scene classification,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 7, pp. 3965–3981, 2017.
[21] Z. Wang, A. Bovik, H. Sheikh et al., “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. on Image Process., vol. 13, no. 4, pp. 600–612, 2004.
[22] X. Ma, P. Wu, Y. Wu, and H. Shen, “A review on recent developments in fully polarimetric SAR image despeckling,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 11, no. 3, pp. 743–758, 2017.