




An Unsupervised CNN-Based Multichannel Interferometric Phase Denoising Method Applied to TomoSAR Imaging

Jie Li , Zhongqiu Xu , Zhiyuan Li, Zhe Zhang , *Member, IEEE*, Bingchen Zhang, and Yirong Wu, *Member, IEEE*

Abstract—Tomographic synthetic aperture radar (TomoSAR) is an advanced SAR interferometric technique to retrieve 3-D spatial information. However, the standard deviation in the reconstructed elevation could be high due to the noise in the interferometric phases, which makes the denoising filter crucial before tomographic reconstruction. In this article, we propose an unsupervised multichannel SAR interferometric phase denoising method based on the convolution neural network. It utilizes the weighted least-squares (WLS) regularization combining with the covariance of multichannel interferometric phases to minimize the standard deviation of phase noise, which leads to the accurate and complete TomoSAR reconstruction. This network is trained by real SAR images and the results of both simulated and real observations verify the effectiveness of our proposed method.

Index Terms—Convolution neural network (CNN), interferometric phase, SAR, tomographic synthetic aperture radar (TomoSAR).

I. INTRODUCTION

TOMOSAR is an essential technique for retrieving spatial information from multibaseline interferometric SAR images acquired with different view angles, which have been intensively developed in the past two decades and shows promising results [1], [2], [3], [4]. The elevation information is contained in the interferometric phases, thus the quality of interferograms is critical to tomographic reconstruction along the elevation direction. However, SAR images are often corrupted by speckle [5] and system thermal noise, which could induce the noise in

interferometric phases [6], [7] and further leads to higher standard deviation in the elevation estimated with tomographic focusing methods [8].

Great efforts have been put into interferometric phase denoising and plenty of filters have appeared [9]. The interferometric phase is traditionally estimated by an operation called multilooking [10] consisting in averaging pixels in the range and/or azimuth directions. Considering that the multilook filter reduces the size of the image, this operation cannot be performed on a large number of pixels. Since then, famous filters, such as Lee [11], [12], maximum-likelihood method [13], have been deployed in the spatial domain. Meanwhile, filters in the frequency domain [14], [15] and in the wavelet domain [16], [17] has also been widely developed, possessing the superior quality of enhancing the phase signatures to be more separable from the noise. However, these methods have not overcome the neighbor-connection limitation. The concept of nonlocal filtering [18], [19] has been migrated to deal with the phase noise in interferograms [20]. The weights are defined from intensity and interferometric phase, and are iteratively refined based both on the similarity between noisy patches and on the similarity of patches from the previous estimate.

The multichannel interferometric phases could carry much more information [21], [22], [23]. For example, TomoSAR exploits a set of interferometric phases to recover the elevation profile inside the pixel. These multichannel interferometric phases are of high interest for a wide range of applications for Earth observation. Therefore, the filtering methods are migrated to deal with multiple interferograms from tomographic data, such as the multilook filtering [24] and the nonlocal filtering [25], [26], [27] using the multidimensional probability density distribution (PDF) or the covariance matrix to achieve the better denoising performance.

Recent advances in deep learning have revolutionized the approaches to SAR processing tasks [28], [29]. The methods in [30], [31] combine convolutional neural networks (CNNs) together with the concept of residual learning for the joint estimation of the interferometric phase and coherence. In [32], the CNN with a multiobjective cost function is used for the task of SAR despeckling. In [33], the deep CNN learns the self-similarity function of real interferometric phase from the input interferogram to achieve the purpose of interferometric phase denoising. However, ground-truth datasets are required to train these network and the output of a well-trained network could

Manuscript received 18 August 2022; revised 15 December 2022 and 5 February 2023; accepted 29 March 2023. Date of publication 3 April 2023; date of current version 20 April 2023. This work was supported by the National Natural Science Foundation of China under Grant 61991421. (*Corresponding author: Zhe Zhang.*)

Jie Li, Zhongqiu Xu, and Zhiyuan Li are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, also with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: lijie195@mailsucas.ac.cn; xuzhongqiu18@mailsucas.ac.cn; lizhiyuan19@mailsucas.ac.cn).

Zhe Zhang is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the Suzhou Aerospace Information Research Institute, Suzhou 215123, China (e-mail: zhangzhe01@aircas.ac.cn).

Bingchen Zhang and Yirong Wu are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: zhangbc@aircas.ac.cn; wyr@mails.ie.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3263964

be close to the ground truth. Nevertheless, there is no noise-free SAR dataset as ground truth in practice. A large number of optical and simulated datasets are used for the training process [34]. However, these reference datasets are quite different from real ones in the imaging mechanism, gray-level distribution, operational wavelength, etc., which results in a greatly reduced reliability of the network. In [35], the proposed network uses the noisy dataset as a reference in the training process. This method is built on the basic principle that the expectation of the noisy reference is theoretically equal to the clean signal. In [36], [37], the unsupervised filters based on CNN generative model have been proposed for single-channel interferometric phase filtering.

Clearly, these networks have been developed for the single interferogram denoising task and the deep learning-based approaches for multichannel interferometric phases denoising are understudied. Considering the application of multichannel interferometric phases in TomoSAR, we propose an unsupervised multichannel SAR interferometric phase denoising method (MS-IPDM) based on CNN in this article. The network is trained to learn the multichannel denoised interferometric phases based on the WLS regularization. Covariance matrix of multichannel interferometric phases is applied to minimize the standard deviation of multichannel interferometric phase noise. Based on the unsupervised training, real acquisitions are fed into the network for the training process. And, the results of simulated and real SAR images ensure the effectiveness of the proposed method.

II. INTERFEROMETRIC PHASE MODEL

For the SAR acquisition, the focused complex-valued measurement g_m of an azimuth–range pixel (x_0, r_0) for the m th acquisition at aperture position b_m is the tomographic projection of the reflected signal along the elevation direction. Considering the TomoSAR imaging model, we can obtain the reflectivity function $\gamma(s)$ along elevation s with an extent of Δs , regardless of the deformation term for simplicity

$$g_m = \int_{\Delta s} \gamma(s) \exp(-j2\pi\xi_m s) ds \quad (1)$$

where, the spatial frequency $\xi_m = -2b_m/\lambda r$ is proportional to the respective aperture position $b_m (m = 1, 2, \dots, M)$. λ is the wavelength of radar signals, and r denotes the range between radar and the observed object, respectively. After the proper phase calibration, the interferometric phase can be expressed as a function of the elevation information and TomoSAR uses the multichannel interferometric phases to obtain the tomographic reconstruction along elevation direction, which is shown in Fig. 1.

However, the phase noise degrades the quality of interferograms severely and leads to errors in the heights estimated with tomographic focusing methods. The multichannel interferometric phase has been completely characterized in the real domain, leading to the following model [11], [38]:

$$\varphi = \phi + \varepsilon \quad (2)$$

where, the φ is the measured multichannel interferometric phase and the ϕ is the ideal interferometric phase without noise,

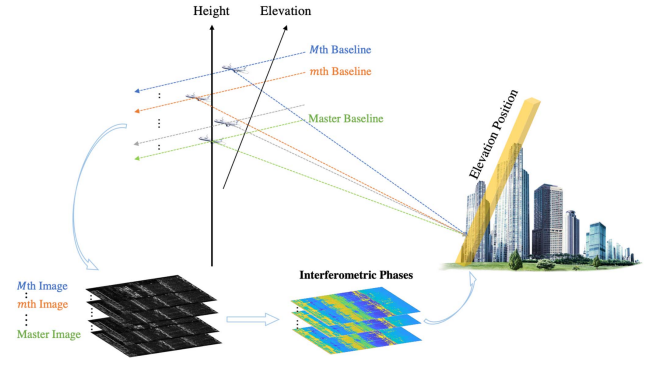


Fig. 1. Principle of TomoSAR.

$\varphi, \phi \in \mathbb{R}^{(M-1) \times 1}$. The multichannel interferometric phases are calculated based on the common reference master SAR image. ε denotes the zero-mean phase noise, whose standard deviation reflects the noise level. According to the principle of TomoSAR imaging, the standard deviation of interferometric phase noise could definitely lead to a higher standard deviation of tomographic reconstruction. And the original phase ϕ and the phase noise ε are assumed to be independent of each other.

III. PROPOSED FILTERING METHOD

A. MS-IPDM Network

The idea of applying deep learning to image processing originated from [39], where the network uses a residual learning strategy to subtract the noisy component. Note that this architecture of network handles the additive-form noise well [39], [40], [41], thus we propose the MS-IPDM to filter the interferometric phase noise based on this network. Fig. 2 shows the architecture of the network.

The residual learning strategy is designed for speeding up the training process and improve denoising performance. In our interferometric phase filtering, we apply the similar idea by creating identity shortcuts for predicting the noise of multichannel interferometric phases. Instead of directly outputting the estimated clean components, the proposed CNN-based method is trained to predict noise. Therefore, the model implicitly filters the latent clean interferometric phases with hidden operations within the network.

The inputs of the network are 128×128 phase patches, generated by multichannel interferometric phase patches of SAR raw data and the output are the corresponding multichannel filtered interferometric phase patches. There are $(M - 1)$ number of channels in the input and output layers, i.e., $M = 2$ when the proposed method deals with the single-channel interferometric phase. The dimension of output and input must be equal, which is achieved by the 3×3 convolutions with a stride of 1. The network removes the pooling layers and combines batch normalization (BN) for fast training and better denoising [42]. It sets the depth of the network according to the patch size used in the algorithms [39]. Each layer uses the filters of size $3 \times 3 \times 64$, extracting 64 feature maps, except the first and last layers. The network is oriented to the denoising of multichannel interferometric phases. The configurations of the network are listed in Table I.

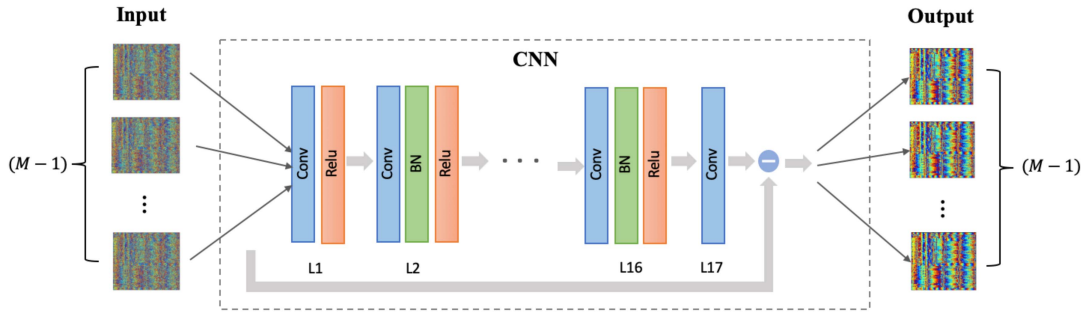


Fig. 2. Architecture of the network.

TABLE I
NETWORK CONFIGURATIONS

	Layer	Filter size	Filters	Output size
L1	Conv	$3 \times 3 \times (M-1)$	64	$128 \times 128 \times 64$
	Relu			$128 \times 128 \times 64$
L2-L16	Conv	$3 \times 3 \times 64$	64	$128 \times 128 \times 64$
	BN			$128 \times 128 \times 64$
	Relu			$128 \times 128 \times 64$
L17	Conv	$3 \times 3 \times 64$	$(M-1)$	$128 \times 128 \times (M-1)$

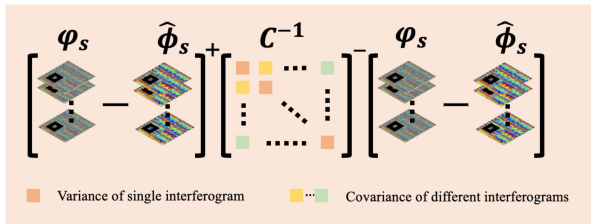


Fig. 3. WLS-based regularization applied in the loss criterion.

B. Loss Criterion Applied in MS-IPDM

To suppress the interferometric phase noise, we need to minimize the standard deviation of multichannel interferometric phase noise. Taking account of the model of measured interferometric phase φ in (2), we define the loss function based on the WLS regularization without considering the ground truth, i.e., the clean version of corresponding multichannel interferometric phases $\hat{\phi}$. Specifically

$$\mathcal{L}(\varphi, \hat{\phi}) = \|\varphi - \hat{\phi}\|_{\Psi}^2 = (\varphi - \hat{\phi})^T \mathbf{W}_{\Psi} (\varphi - \hat{\phi}) \quad (3)$$

where, $\hat{\phi}$ denotes the output multichannel interferometric phases of the network. \mathbf{W}_{Ψ} is the weighting matrix, which is defined by the inverse of the covariance \mathbf{C} of the multichannel interferometric phases, i.e., $\mathbf{W}_{\Psi} = \mathbf{C}^{-1}$. The covariance can be expressed as follows:

$$\mathbf{C} = \text{diag}[\sigma_1^2, \sigma_2^2, \dots, \sigma_{M-1}^2] \quad (4)$$

where, $\sigma_i^2, i = 1, 2, \dots, M-1$ is defined by correlation coefficient in [43], $\sigma_i^2 = (1 - |\rho_i|^2)/2L|\rho_i|^2$, where L is the number of looks and ρ_i denotes the corresponding complex coefficient. The loss function is illustrated in Fig. 3. It is not hard to see that σ_i^2 is inversely proportional to the coherence, thus the weighting matrix of this term is proportional to the correlation coefficient. Based on the loss criterion, the output would fit the input to a

greater degree in the regions with high coherence, and less well in the regions with poor coherence. By setting the WLS-based loss function, the standard deviation of multichannel interferometric phase noise could be minimized, which means that the zero-mean noise could be effectively suppressed.

During the training process, the parameters of the network could be optimized in order to minimize the loss function. Thus, the output of the network would have the low standard deviation of multichannel interferometric phase noise. Meanwhile, there is no need for ground truth in the training process according to the loss function (3), as we can learn from the input data itself, without requiring its clean version as the training dataset. Therefore, our network is completely unsupervised and can be trained directly on real SAR data, which are of supreme importance to the development of deep learning in the SAR image field. The framework of the proposed CNN method is provided in Fig. 4.

IV. EXPERIMENTS

A. Experimental Settings and Data Description

In our experiments, we use the Aerospace Information Research Institute, Chinese Academy of Sciences array data acquired on the city of YunCheng in Shanxi province, as the dataset. All the input interferometric phases for training, which are of the image size of 3100×1220 , are formed with the first track as a common reference. To ensure the generalization ability of the network and to ease the computational burden, the input is divided into patches of 128×128 pixels, with a stride of 64. This dataset has 34 225 image patches, including 24 260 patches for training and the remaining for test. The proposed model is implemented in the PyTorch package and run on an NVIDIA 2060Ti GPU with a 6 GB RAM.

Meanwhile, the number of channels in the input layer and output layer is set to be 7 based on the training dataset for simplicity. And the network is trained for 100 epochs using the Adam optimizer, with a learning rate set as 0.001. A batch size of 24 is used in all experiments. In this section, we describe the test experiments conducted on the simulated and real SAR datasets to reveal the denoising performance of the proposed MS-IPDM.

B. Simulated Data

In this section, we simulate a simple building scene composed of planar facets representing the ground, wall, and roof.

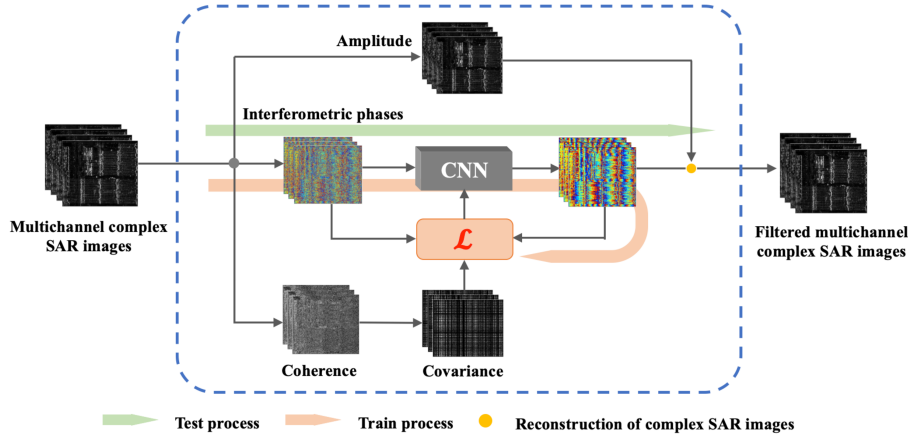


Fig. 4. Framework of the proposed MS-IPDM based on CNN. The process of amplitude is not discussed in detail here.

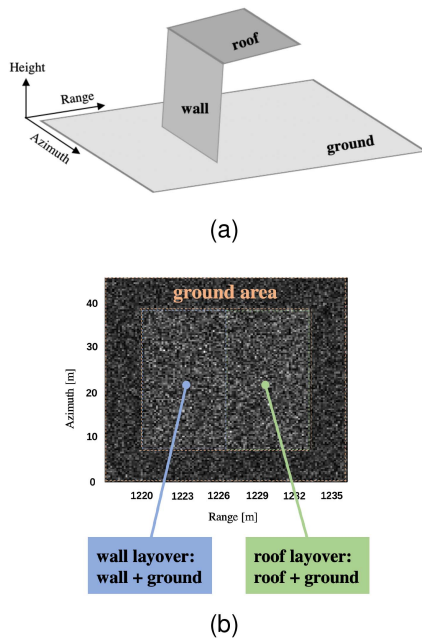


Fig. 5. Simulated scene. (a) 3-D scene generated by tiling rectangular facets of different footprints and orientations. (b) Corresponding normalized intensity of one image with the description of the different areas.

The framework of simulation is shown in Fig. 6. Taking into account the number of channels in the input layer of the network, we assume 15-track complex SAR images to verify the generalization of the proposed method, which has a different elevation resolution from that of the training data. The number of interferometric phases of the simulated data is twice that of the actual data, which are divided into two groups and fed into the network sequentially. The difference in the number of input channels leads to the different elevation resolution to validate the effectiveness of the proposed network. The building roof is 50 m in the height direction. Fig. 5 shows the simulated scene and the corresponding intensity of an image in the stack. There are three areas in the intensity image.

To demonstrate the suppression performance of interferometric phase noise, we simulate the SAR images with 0.6 rad

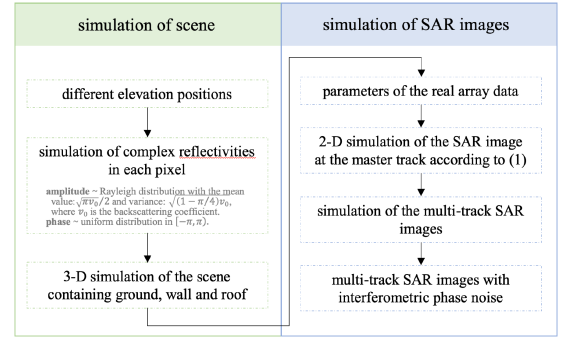


Fig. 6. Framework of simulation.

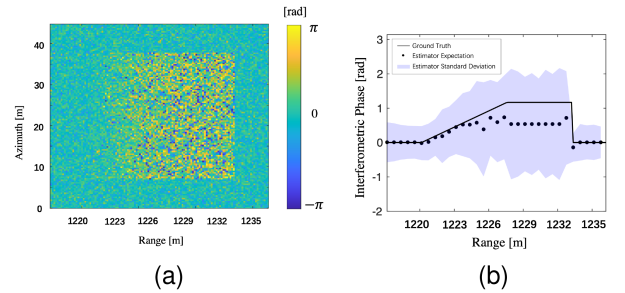


Fig. 7. Simulated data without filtering. (a) Interferometric phase. (b) Statistical answer on the unfiltered interferometric phases.

standard deviation in the multichannel interferometric phases. Due to phase noise, the single-look (unfiltered) interferometric phase has a high standard deviation of the multichannel interferometric phase noise, which can be seen in Fig. 7. The ground truth, the mean, and an interval of standard deviation are represented on the graphics for the three estimated components. Fig. 8 presents the example of the multilook (5×5) interferogram, the nonlocal filtered interferogram [20], the CNN-IPDM (interferometric phase denoising method based on CNN-InSAR [37]), the Gen-IPDM (interferometric phase denoising method based on GenInSAR [36]), the adapted SS-IPDM (single-channel interferometric phase denoising method), and the proposed MS-IPDM filtered interferograms.

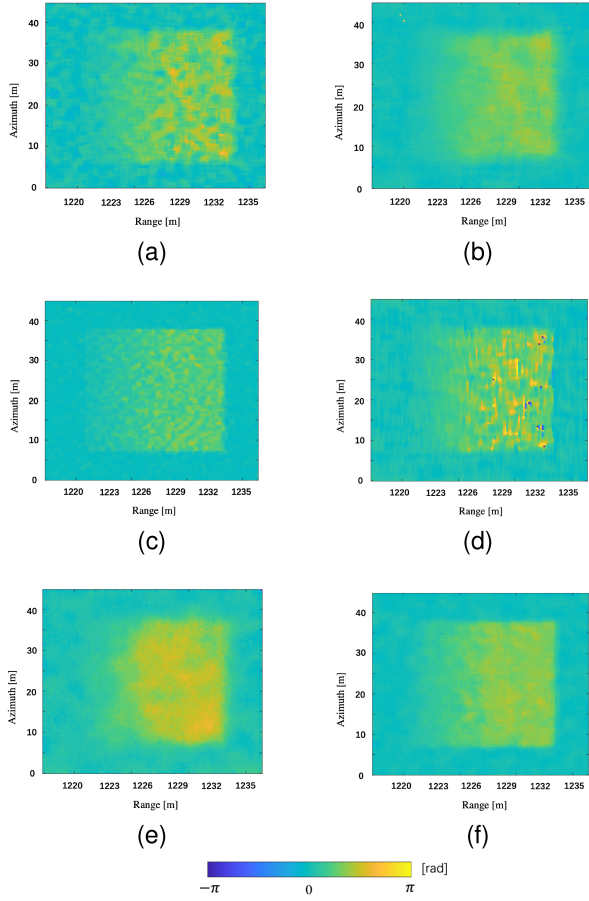


Fig. 8. Interferograms of the simulated data after different filters. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM. SS-IPDM is introduced by changing the number of channels in the input layer to be single.

The multilook filter averages 5×5 pixels in the range and azimuth directions. The nonlocal filter performing a weighted averaging of similar pixels [20] based on the multichannel statistical model of interferometric phases. The CNN-IPDM is a supervised interferometric phase denoising method adapted from CNN-InSAR [37]. The training dataset is simulated following the original paper. The Gen-IPDM is an unsupervised interferometric phase denoising method adapted from GenInSAR [36]. The SS-IPDM is a method achieved by changing the number of channels to be single in the input layer, and the MS-IPDM is the proposed multichannel interferometric phase denoising method.

Fig. 9 shows the statistical answer of different filters on the samples from 5 to 35 m along the azimuth direction. The statistics have been measured on one of the denoised interferometric phases. We find that SS-IPDM can reduce the standard deviation of interferometric phase noise partly but it represents a bias sometimes, which is similar to the performances of single-channel CNN-IPDM and Gen-IPDM. Meanwhile, nonlocal filter is unbiased with a lower standard deviation, which considers the joint PDF of multichannel interferometric phase noise. And the proposed MS-IPDM can do better, which uses the covariance of multichannel interferometric phase to suppress the noise.

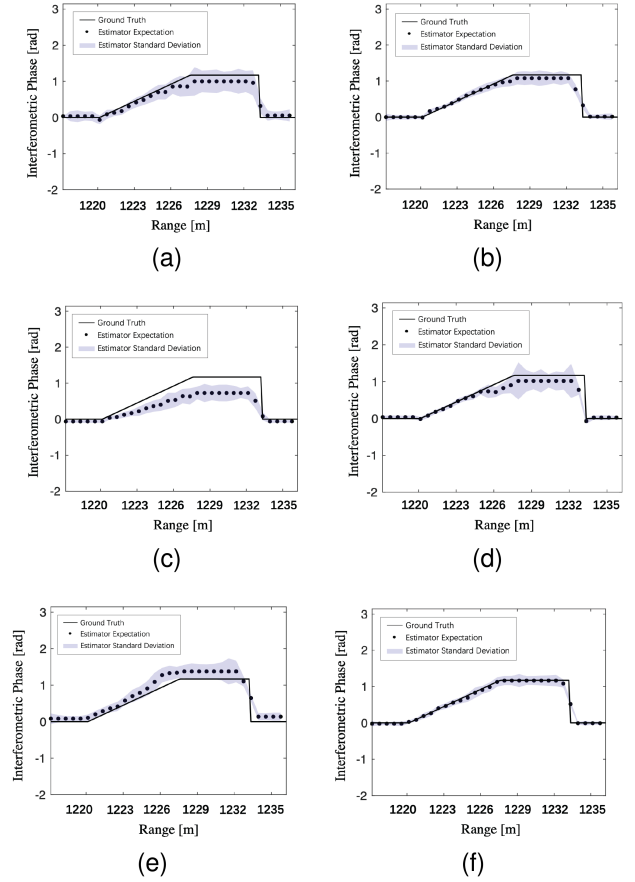


Fig. 9. Statistical answers on the different filtered interferometric phases. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM.

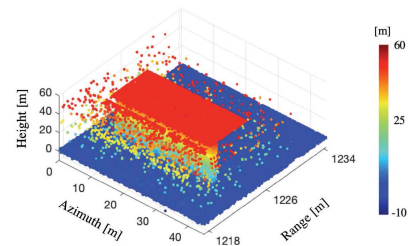


Fig. 10. 3-D reconstruction of unfiltered SAR images.

Table II shows the filtering performances of the interferometric phases for the adoptive methods. $RMSE_{wa}^p$ denotes the root-mean-square error of the restored interferometric phases on the whole area, and PCE_{wa}^p [36] describes the phase cosine error of the restored interferometric phases on the whole area. It can be noticed that nonlocal filter and MS-IPDM have small phase errors, which is consistent with the statistical answers on the different filtered interferometric phases in Fig. 8.

Figs. 10 and 11 show the 3-D reconstruction of unfiltered and different filtered data, obtained with the iterative shrinkage and thresholding algorithm [44]. Clearly, the 3-D reconstruction of the single-look interferogram is extremely noisy. Filtering allows to strongly mitigate this effect. It is noted that the results obtained from the adoptive methods are able to preserve the

TABLE II
 FILTERING PERFORMANCES OF THE SIMULATED INTERFEROMETRIC PHASES FOR THE ADOPTIVE METHODS

Metric (rad)	Single look	Multilook	Nonlocal	CNN-IPDM	Gen-IPDM	SS-IPDM	MS-IPDM
RMSE $_{wa}^p$	1.1800	0.6422	0.4871	0.7051	0.6159	0.5814	0.3267
PCD $_{wa}^p$	0.2949	0.1084	0.0722	0.1376	0.1038	0.0812	0.0574

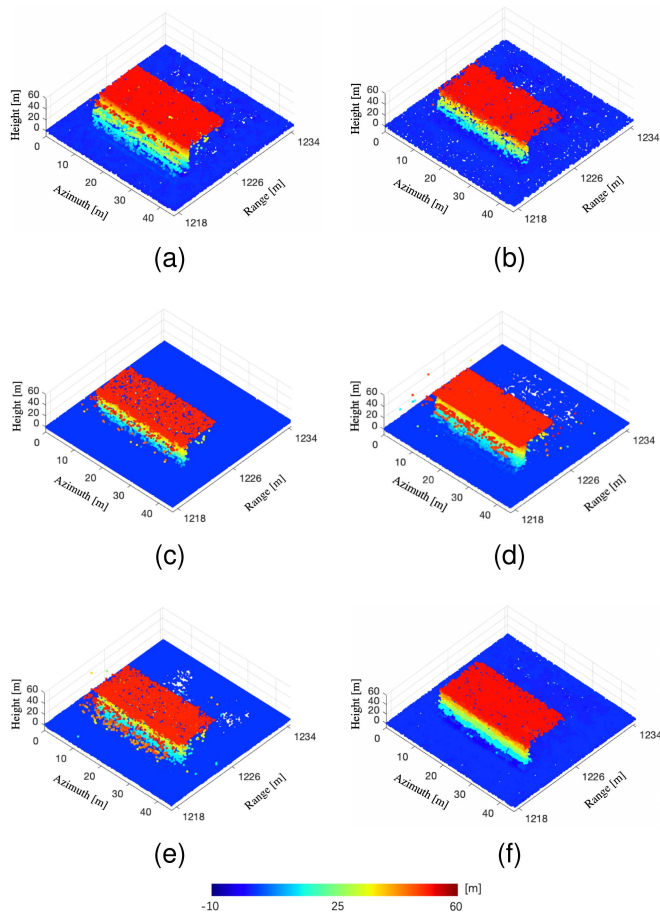
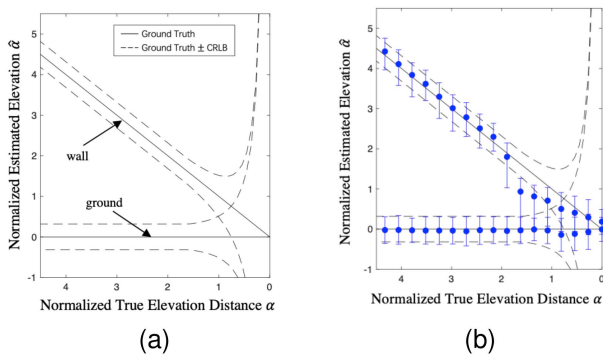
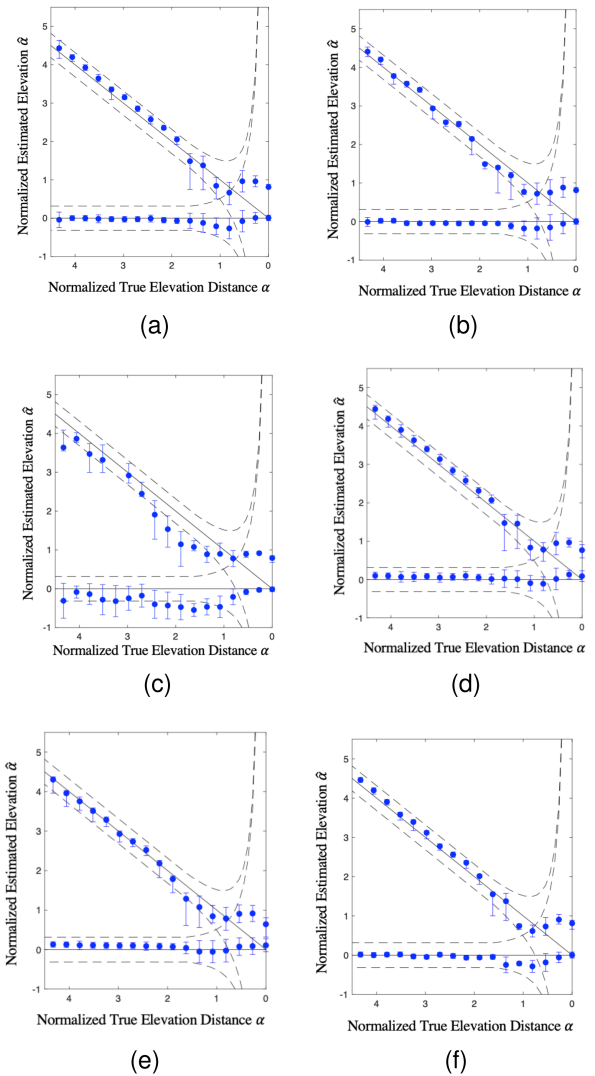


Fig. 11. 3-D reconstruction of different filtered SAR images. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM.


 Fig. 12. Normalized true elevation α of simulated wall and ground (solid lines), as well as the CRLB (dashed lines) of normalized elevation estimates, both w.r.t. normalized true elevation distance α . (a) Ground truth. (b) Result of unfiltered data. Each dot depicts mean value of all estimates, with error bar indicating its standard deviation.

 Fig. 13. Normalized true elevation α of reconstructed wall and ground (solid lines), as well as the CRLB (dashed lines) of normalized elevation estimates, both w.r.t. normalized true elevation distance α . (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM.

structure of the scene while removing noise on the reconstructed reflectivity profile. However, the better restoration of interferometric phases leads to the better tomographic reconstruction. As expected, the nonlocal filter suppresses the outliers well and MS-IPDM achieves a more accurate and complete result.

The standard deviation of interferometric phases could definitely result in the standard deviation of reconstructed tomographic elevation. To show the filtering performance, we analyze the Cramér-Rao lower bounds (CRLB) for the estimated

TABLE III
PERFORMANCE OF 3-D RECONSTRUCTION OF SIMULATED DATA FOR THE ADOPTIVE METHODS

Metric (pixel)	Single-look	Multilook	Nonlocal	CNN-IPDM	Gen-IPDM	SS-IPDM	MS-IPDM
MAE_{ga}^h	0.4481	0.0074	0.0203	0.6879	0.0390	0.0281	0.0040
MAE_{rl}^h	1.1388	1.1627	0.8415	2.4659	1.3095	2.3702	0.7548
STD_{ga}^h	0.6973	0.0861	0.2288	0.4634	0.0624	0.1687	0.0198
STD_{rl}^h	2.1891	2.0042	1.0990	3.6819	2.6433	3.6610	0.9849
$RMSE_{ga}^h$	0.6974	0.0862	0.2288	0.8294	0.0625	0.1700	0.0198
$RMSE_{rl}^h$	2.2386	2.0039	1.0989	3.8757	2.6617	3.6741	1.0095

The bold value is used for the clear presentation of the better performance of our proposed method.

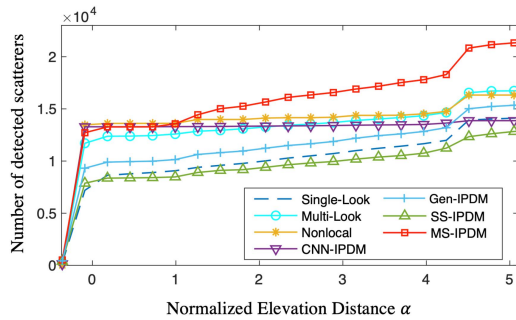


Fig. 14. Number of detected scatterers for the 3-D reconstruction based on different filters.

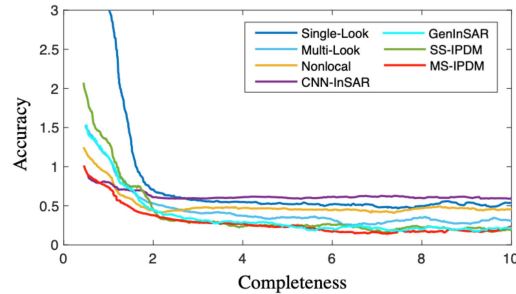
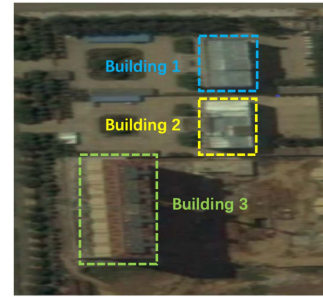


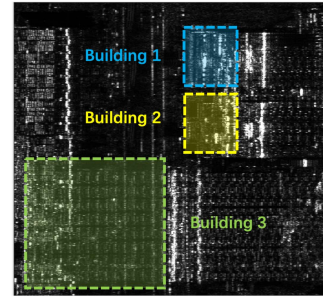
Fig. 15. Accuracy versus completeness for the 3-D reconstruction based on different filters.

elevation of two scatterers in the wall layover area. The true elevation of simulated wall and ground is shown as two solid line segments w.r.t. their normalized elevation distance α in Fig. 12(a), respectively. It is noted that α is the elevation distance between two scatterers normalized w.r.t. the Rayleigh resolution unit. And, the dashed lines mark true elevation $\pm 1 \times \text{CRLB}$ with the simulated interferograms.

As shown in Figs. 12(b) and 13, the elevation estimates $\hat{\alpha}$ of wall and ground are plotted w.r.t. their normalized true elevation difference α . Each dot depicts mean value of all estimates, with error bar indicating its standard deviation. Regardless of the superresolution properties of the tomographic focusing method, we find that the tomographic reconstruction of single-look interferometric phases has a higher standard deviation. After



(a)



(b)

Fig. 16. Experimental area with three buildings to be reconstructed. (a) Optical image (Copyright Google). (b) Intensity map.

the adoptive filters, the deviation has been suppressed partly, especially with the proposed MS-IPDM.

To provide a quantitative analysis, a comparison of the obtained results with ground truth is also performed based on different metrics in Table III. MAE_{ga}^h denotes the mean absolute error of reconstructed height properties in the ground area, STD_{ga}^h describes the standard deviation of reconstructed height properties in the ground area, and $RMSE_{ga}^h$ describes the root-mean-square error of the reconstructed height properties in ground area. MAE_{rl}^h represents the mean absolute error of the reconstructed height properties in roof layover area, STD_{rl}^h denotes the standard deviation of the reconstructed height properties in roof layover area, and $RMSE_{rl}^h$ describes the root-mean-square error of the reconstructed height properties in roof layover area. It can be noticed that MS-IPDM reduces deviation well while keeping the lower errors, which means that there are

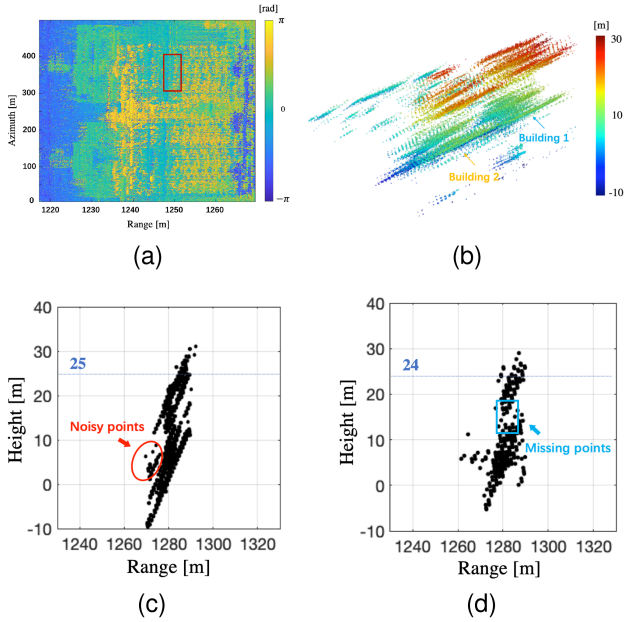


Fig. 17. Real data containing building 1 and building 2 without filtering. (a) Interferometric phase. (b) 3-D points cloud. (c) Range–height plane of building 1 reconstruction with the height of 25 m. (d) Range–height plane of building 2 reconstruction with the height of 24 m.

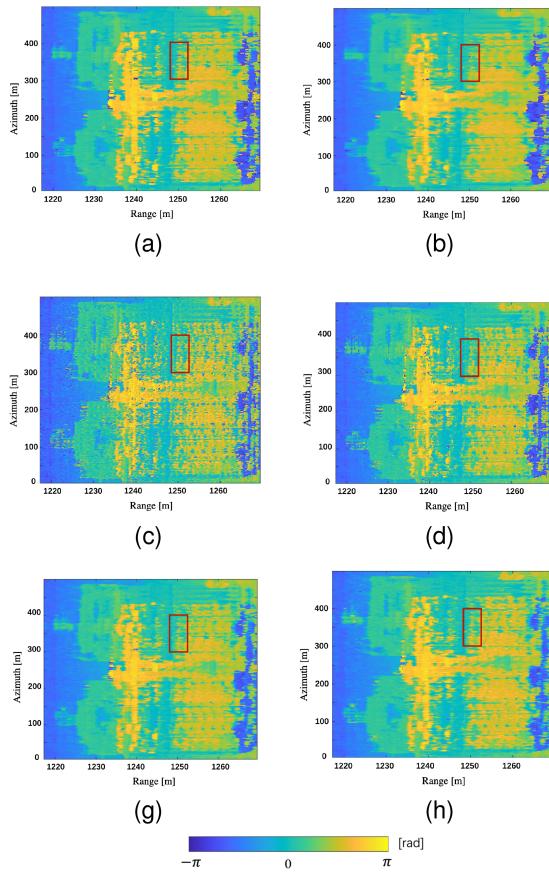


Fig. 18. Interferograms of the real data containing building 1 and building 2 after different filters. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM. The red rectangles are the examples of the effect in denoising and detail-preserving, which represent the regular shapes of the building.

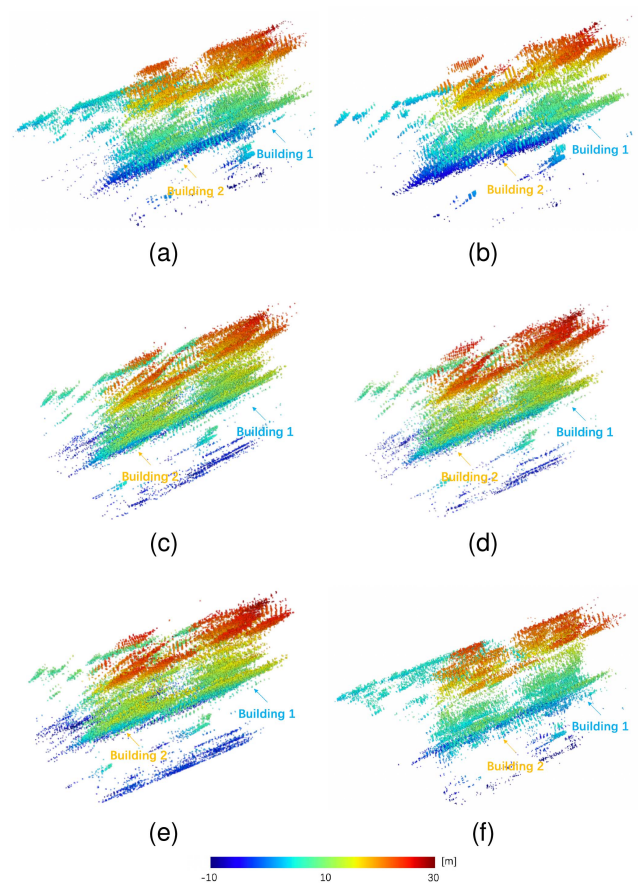


Fig. 19. 3-D points cloud of the area containing building 1 and building 2 based on different filtered interferograms. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM.

fewer noisy points in the 3-D reconstruction of the proposed method. In this case, the tomographic reconstruction could be precise and accurate.

Meanwhile, we also showed the number of detected scatterers to analyzed the missing points in Fig. 14. Note that the reconstruction after nonlocal filter has impressive detection capability, i.e., many of the scatterers in the ground area are detected. However, the performance of detection in the wall layover decreases accordingly. In contrast, reconstruction after the proposed MS-IPDM has the better detection, especially in the wall layover area, indicating that the missing points are reduced effectively.

Taking into account the outliers and missing points in the 3-D reconstruction, we plotted the curves of accuracy as a function of completeness [45] in Fig. 15. The lower values represent the improved performance. It can be seen that the proposed MS-IPDM leads to a better tradeoff between accuracy and completeness, which is consistent with the suppression of noisy and missing points in the tomographic reconstruction.

C. Real Data

Furthermore, we present the 3-D reconstruction results on the three buildings of experimental array data, which are not existent in the training dataset. The radar system operates at

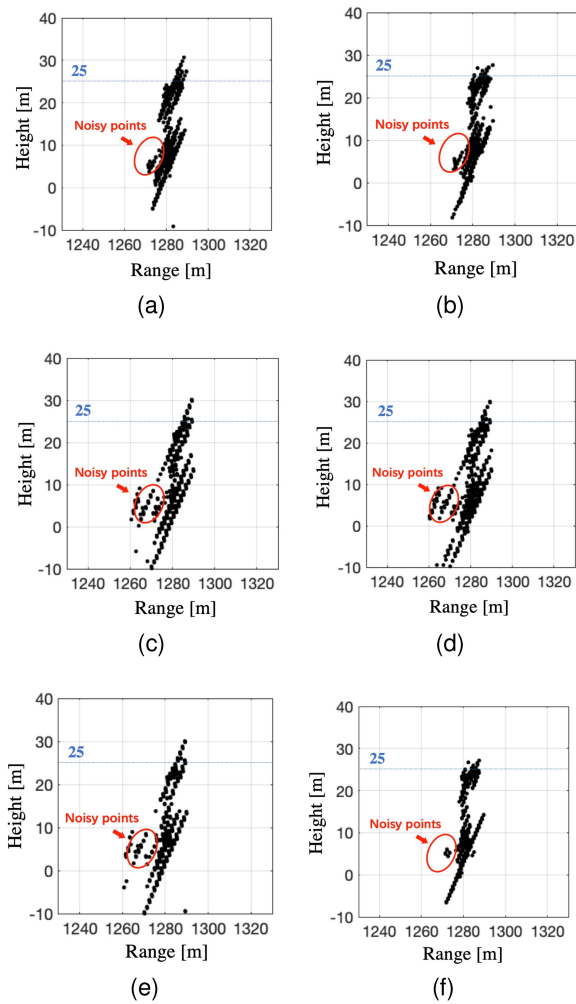


Fig. 20. Range–height plane of the 3-D reconstruction results of the building 1. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM. Here we assume that the ground truth of the building 1 could be fitted with a vertical line at the range–height plane.

15 GHz and has eight channels in the cross-track direction. The distance between the adjacent channels is 0.8 m. The height of the radar platform is 972 m and the local incidence angle is 30° . The corresponding intensity maps of the areas are shown in Fig. 16. As we know, the height of building 1 is 25 m, the height of building 2 is 24 m, and the height of building 3 is 50 m.

Fig. 17 shows the unfiltered interferometric phases of building 1 and building 2, corresponding 3-D reconstruction and range–height planes of the two buildings. Here we assume that the ground truth of the buildings could be fitted with vertical lines at the range–height plane, whose height is known. It can be seen that there is much phase noise in the interferometric phase. Correspondingly, noisy and missing points appear in the 3-D reconstruction, which are obvious in the range–height plane of the reconstruction.

Fig. 18 shows the different filtered interferometric phases of building 1 and building 2. We find that the proposed MS-IPDM leads to less phase noise in the interferogram and could not be oversmoothing like the multilook filter. For example, the noise is depressed effectively and the details of the building

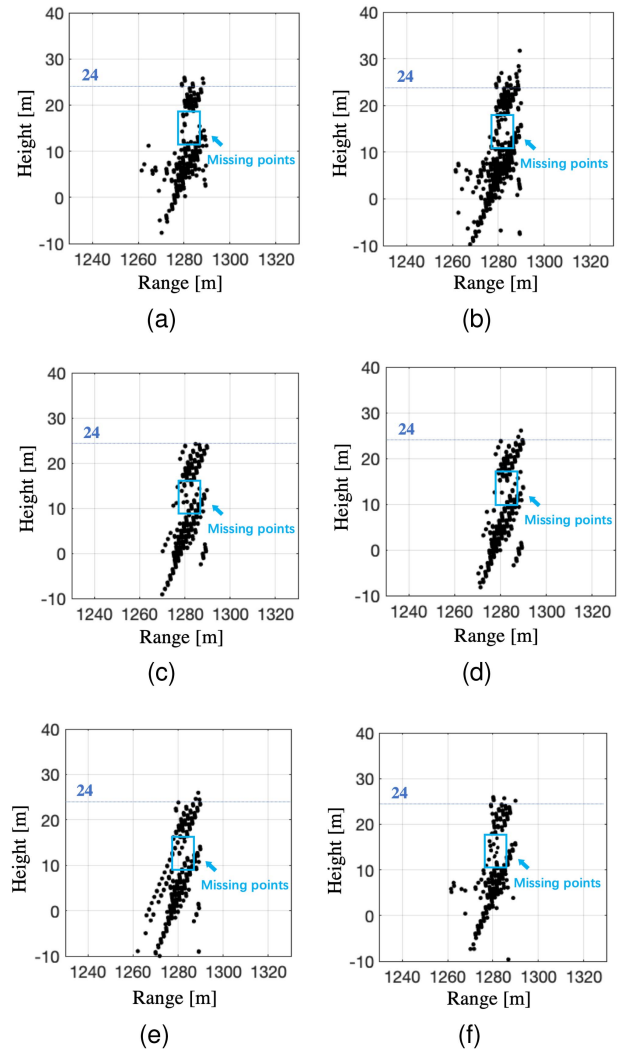


Fig. 21. Range–height plane of the 3-D reconstruction results of the building 2. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM. Here we assume that the ground truth of the building 2 could be fitted with a vertical line at the range–height plane.

are preserved well. As seen in the red rectangles, they denote the regular shapes of the building. The 3-D reconstructed results are presented in Fig. 19. It can be observed that the buildings are well reconstructed and fewer noisy points make the visual interpretation easy. Meanwhile, fewer points are missing, leading to relatively complete walls. The range–height planes of the 3-D reconstruction results of building 1 and building 2 are reported in Figs. 20 and 21.

On the one hand, most of the noisy points can be removed by the filtering methods, as shown in the red ellipses marked. Visual inspection shows that there are fewer noisy points around the buildings based on the interferograms after the proposed filter, which is especially evident in the reconstructed results of the building 1. On the other hand, the empty walls of the buildings marked by the blue rectangles can be filled by the applied filtering methods, which is evident in the reconstructed results of the building 2. In summary, the nonlocal filter leads

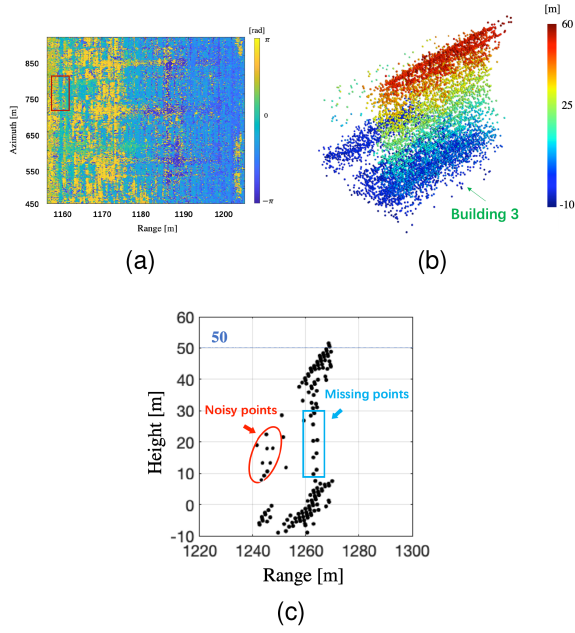


Fig. 22. Real data containing the building 3 without filtering. (a) Interferometric phase. (b) 3-D reconstruction. (c) Range–height plane of building 3 with the height of 50 m.

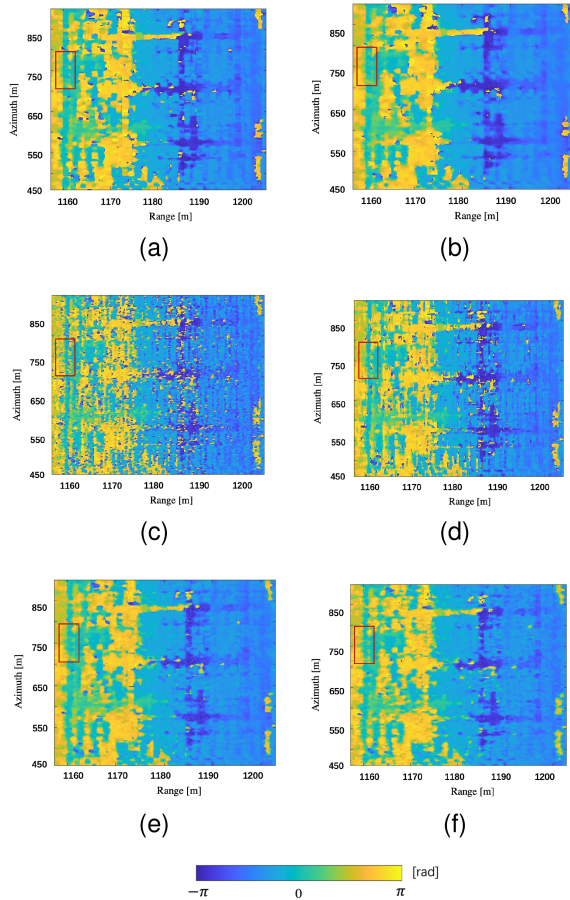


Fig. 23. Interferograms of the real data containing the building 3 after different filters. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM. The red rectangles are the examples of the effect in denoising and detail-preserving, which represent the areas around the building roof.

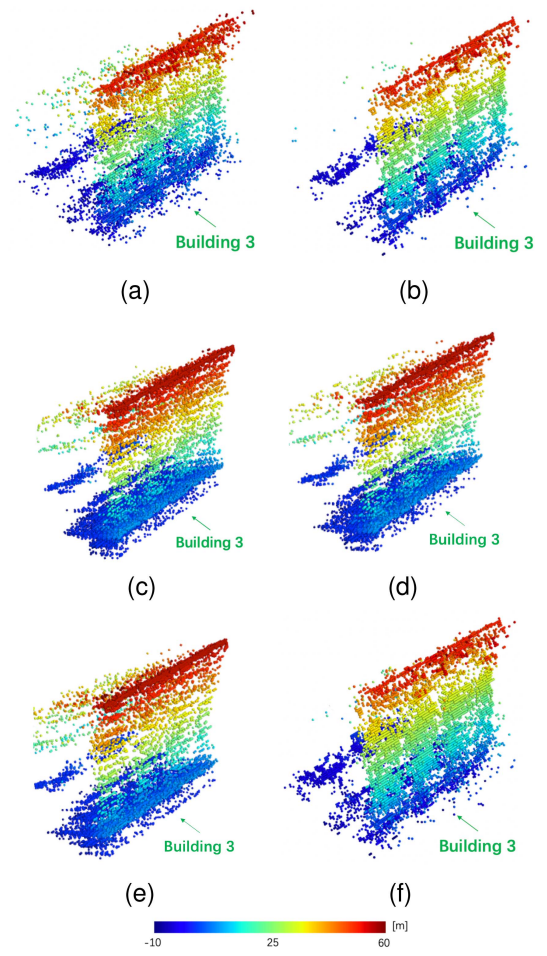


Fig. 24. 3-D points cloud of the building 3 in Fig. 23 based on different filtered interferograms. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM.

to visually comparable results whereas MS-IPDM results in a lower amount of noisy points and more complete buildings.

Similarly, the unfiltered interferometric phase and 3-D reconstructed result of building 3 are represented in Fig. 22. And, the filtered interferometric phases and 3-D reconstructed results of building 3 are shown in Figs. 23 and 24. The red rectangles are the examples of the effect in denoising and detail-preserving, which represent the areas around the building roof. It is noted that the nonlocal filter handles the noisy points and missing points well, compared with other filters. Particularly, the building is better-reconstructed thanks to the proposed filter. Range–height plane results of the 3-D reconstruction in Fig. 25 show that there are fewer noisy points around the building and the missing points of the wall are filled by the MS-IPDM.

Numerical evaluation of experimental results is difficult due to the lack of reference data. To measure the filtering performance on the 3-D reconstruction, we use the MAE_b^h and STD_b^h to evaluate the obtained height of different buildings, as shown in Table IV. According to the quantitative analysis, the proposed MS-IPDM keeps a lower standard deviation than the nonlocal filter, which is consistent with the simulation analysis. Combining with the small absolute error, we can conclude that

TABLE IV
PERFORMANCE OF 3-D RECONSTRUCTION OF REAL DATA FOR THE ADOPTIVE METHODS

Metric (pixel)	Single-look	Multilook	Nonlocal	CNN-IPDM	Gen-IPDM	SS-IPDM	MS-IPDM
$MAE_{b_1}^h$	1.8196	1.4485	1.1293	1.6885	1.6640	1.5027	1.1133
$MAE_{b_2}^h$	2.1081	1.5230	1.3801	1.8197	1.6799	1.7648	1.4453
$MAE_{b_3}^h$	2.3339	2.9150	1.3675	1.6187	1.4731	1.5308	1.2105
$STD_{b_1}^h$	2.2351	1.9082	1.4727	1.7966	1.8182	1.7235	1.1689
$STD_{b_2}^h$	2.2042	2.0038	1.7815	1.8881	1.7055	1.6149	1.5850
$STD_{b_3}^h$	2.0024	1.5016	1.9897	1.9297	1.8139	1.8923	1.3184

The bold value is used for the clear presentation of the better performance of our proposed method.

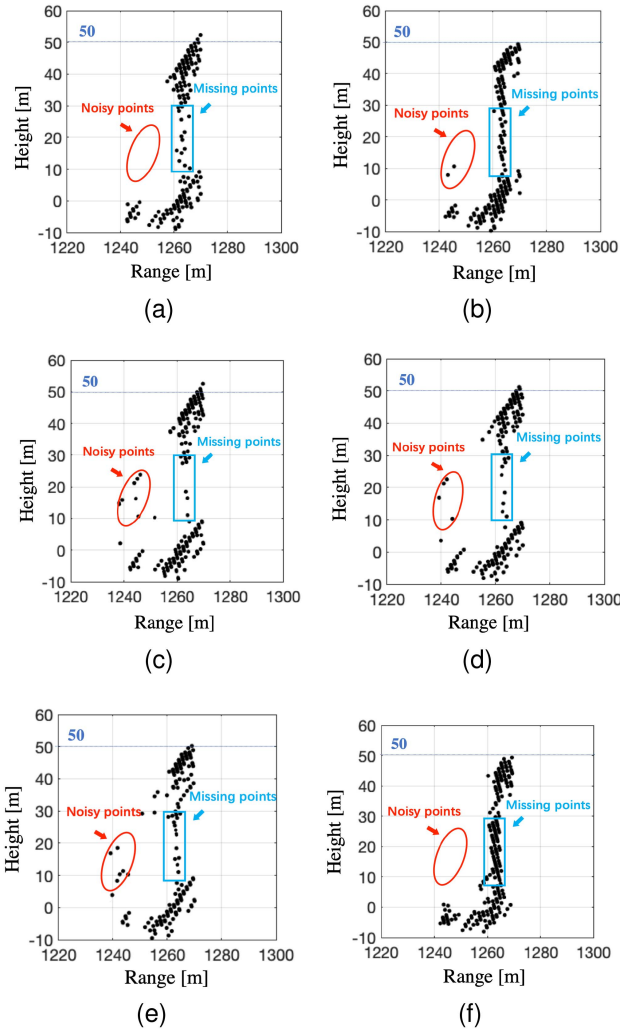


Fig. 25. Range–height plane of the 3-D reconstruction results of the building 3. (a) Multilook. (b) Nonlocal. (c) CNN-IPDM. (d) Gen-IPDM. (e) SS-IPDM. (f) MS-IPDM. Here we assume that the ground truth of the building 3 could be fitted with a vertical line at the range–height plane.

MS-IPDM improves the accuracy and precision of the tomographic reconstruction.

For the analysis of detection, we plotted the statistics of the number of detected scatterers in Figs. 26–28. In the area around

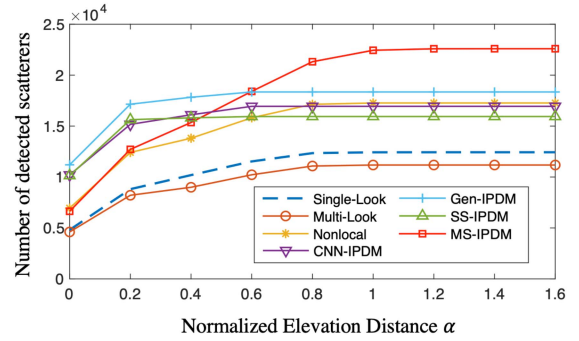


Fig. 26. Number of detected scatterers for the 3-D reconstruction of building 1 based on different filters.

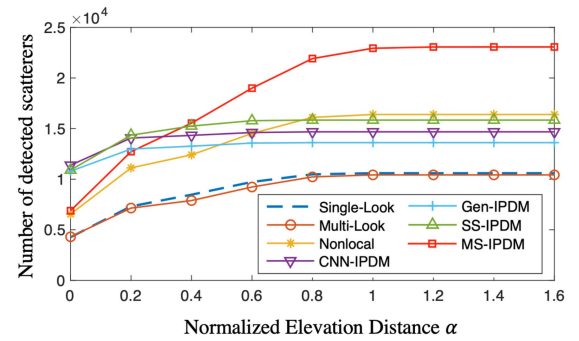


Fig. 27. Number of detected scatterers for the 3-D reconstruction of building 2 based on different filters.

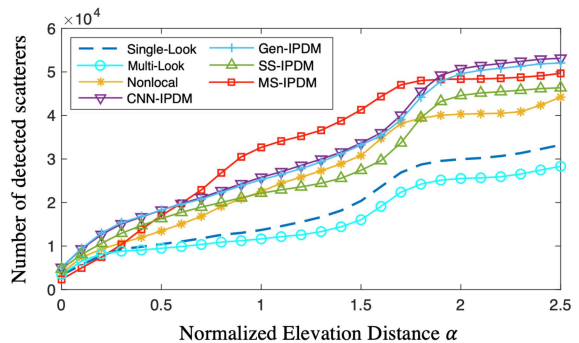


Fig. 28. Number of detected scatterers for the 3-D reconstruction of building 3 based on different filters.

ground, there are fewer noisy points in MS-IPDM thus the number of scatterers should be small. Meanwhile, the reconstruction of wall is dense, thus the number of detected scatterers grows faster than other methods, which indicates the improvement in completeness.

V. CONCLUSION

In this article, we have shown that the interferometric phase noise has a bad effect on the 3-D reconstruction of TomoSAR, resulting in the higher standard deviation of tomographic elevation. Consequently, we have proposed an unsupervised multichannel interferometric phase denoising method for TomoSAR based on CNN, which utilizes WLS regularization combining the covariance matrix of multichannel interferometric phases to minimize the standard deviation of phase noise. The network has been trained by real SAR images and tested by the simulated and real data.

We have simulated a simple scene containing different layover areas to evaluate the influence of the proposed filter generally. Considering the filtered interferograms and the quality of reconstructed 3-D points cloud, we have presented numerical results for nonfiltered and different filtered data. As expected, the proposed method could suppress the standard deviation of interferometric phases effectively and keep a low error, therefore the precision and accuracy of the reconstruction could be improved by suppressing the noisy points. Meanwhile, we found that MS-IPDM could reduce the missing points of the 3-D points cloud, thus lead to the better detection and completeness.

Furthermore, we have chosen three buildings of the real data, which were not existent in the training dataset. According to the corresponding results, noisy and missing points have been suppressed effectively. Based on the known heights of buildings, we found that the standard deviation of the reconstructed heights could be lower, while keeping the small absolute error. Meanwhile, the statistics of reconstructed scatterers have proved the improvement of detection, which were in line with the simulation evaluation.

REFERENCES

- [1] A. Budillon, A. Evangelista, and G. Schirinzi, "Three-dimensional SAR focusing from multipass signals using compressive sampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 488–499, Jan. 2011.
- [2] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.
- [3] M. Schmitt and X. X. Zhu, "Demonstration of single-pass millimeterwave SAR tomography for forest volumes," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 202–206, Feb. 2016.
- [4] C. Rambour, A. Budillon, A. C. Johns, L. Denis, F. Tupin, and G. Schirinzi, "From interferometric to tomographic SAR: A review of synthetic aperture radar tomography-processing techniques for scatterer unmixing in urban areas," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 2, pp. 6–29, Jun. 2020.
- [5] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, no. 11, pp. 1145–1150, 1976.
- [6] H. A. Zebker et al., "Decorrelation in interferometric radar echoes," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 950–959, Sep. 1992.
- [7] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Problems*, vol. 14, no. 4, pp. R1–R54, 1998.
- [8] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, "Joint regularization of phase and amplitude of InSAR data: Application to 3-D reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3774–3785, Nov. 2009.
- [9] G. Xu, Y. Gao, J. Li, and M. Xing, "InSAR phase denoising: A review of current technologies and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 2, pp. 64–82, Jun. 2020.
- [10] J. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1017–1028, Sep. 1994.
- [11] J. Lee, K. P. Papathanassiou, T. L. Ainsworth, M. R. Grunes, and A. Reigber, "A new technique for noise filtering of SAR interferometric phase images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1456–1465, Sep. 1998.
- [12] C. Chao, K. Chen, and J. Lee, "Refined filtering of interferometric phase from InSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 12, pp. 5315–5323, Dec. 2013.
- [13] E. Trouvé, M. Caramma, and H. Maître, "Fringe detection in noisy complex interferograms," *Appl. Opt.*, vol. 35, no. 20, pp. 3799–3806, 1996.
- [14] R. M. Goldstein and C. L. Werner, "Radar interferogram filtering for geophysical applications," *Geophysical Res. Lett.*, vol. 25, no. 21, pp. 4035–4038, 1998.
- [15] M. Jiang et al., "The improvement for Baran phase filter derived from unbiased InSAR coherence," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 3002–3010, Jul. 2014.
- [16] C. Lopez Martinez and X. Fabregas, "Modeling and reduction of SAR interferometric phase noise in the wavelet domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 12, pp. 2553–2566, Dec. 2002.
- [17] Y. Bian and B. Mercer, "Interferometric SAR phase filtering in the wavelet domain using simultaneous detection and estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1396–1416, Apr. 2011.
- [18] C. A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.
- [19] X. Yang and D. A. Clausi, "Structure-preserving speckle reduction of SAR images using nonlocal means filters," in *Proc. IEEE 16th Int. Conf. Image Process.*, Cairo, Egypt, 2009, pp. 2985–2988.
- [20] C. A. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Nonlocal interferogram estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1441–1452, Apr. 2011.
- [21] C. A. Deledalle, L. Denis, S. Tabti, and F. Tupin, "MuLoG, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction?," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4389–4403, Sep. 2017.
- [22] T. Pan, D. Peng, W. Yang, and H. Li, "A filter for SAR image despeckling using pre-trained convolutional neural network model," *Remote Sens.*, vol. 11, no. 20, 2019, Art. no. 2379.
- [23] A. G. Mullissa, C. Persello, and J. Reiche, "Despeckling polarimetric SAR data using a multistream complex-valued fully convolutional network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Mar. 2021.
- [24] Y. Huang, L. Ferro Famil, and A. Reigber, "Under-foilage object imaging using SAR tomography and polarimetric spectral estimators," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2213–2225, Jun. 2012.
- [25] Y. Shi, X. Zhu, and R. Bamler, "Nonlocal compressive sensing-based SAR tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 3015–3024, May 2019.
- [26] H. Aghababae, G. Ferraioli, G. Schirinzi, and M. R. Sahebi, "The role of nonlocal estimation in SAR tomographic imaging of volumetric media," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 729–733, May 2018.
- [27] O. D'Hondt, C. López Martínez, S. Guillaso, and O. Hellwich, "Nonlocal filtering applied to 3-D reconstruction of tomographic SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 272–285, Jan. 2018.
- [28] X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.
- [29] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, "Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 29–51, Jun. 2021.
- [30] X. Sun, A. Zimmer, S. Mukherjee, N. K. Kottayil, P. Ghuman, and I. Cheng, "DeepInSAR-A deep learning framework for SAR interferometric phase restoration and coherence estimation," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2340.
- [31] F. Sica, G. Gobbi, P. Rizzoli, and L. Bruzzone, "φ-Net: Deep residual learning for InSAR parameters estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3917–3941, May 2021.

- [32] S. Vitale, G. Ferraioli, and V. Pascazio, "Multi-objective CNN-based algorithm for SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9336–9349, Nov. 2021.
- [33] H. Yu, T. Yang, L. Zhou, and Y. Wang, "PDNet: A lightweight deep convolutional neural network for InSAR phase denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–9, 2022.
- [34] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Mar. 2021.
- [35] J. Lehtinen et al., "Noise2Noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Mach. Learn.*, PMLR, pp. 2965–2974, 2018.
- [36] S. Mukherjee, A. Zimmer, X. Sun, P. Ghuman, and I. Cheng, "An unsupervised generative neural approach for InSAR phase filtering and coherence estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 11, pp. 1971–1975, Nov. 2021.
- [37] S. Mukherjee, A. Zimmer, N. K. Kottayil, X. Sun, P. Ghuman, and I. Cheng, "CNN-based InSAR denoising and coherence metric," in *Proc. IEEE Sensors*, 2018, pp. 1–4.
- [38] X. Luo, X. Wang, Y. Wang, and S. Zhu, "Efficient InSAR phase noise reduction via compressive sensing in the complex domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1615–1632, May 2018.
- [39] 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.
- [40] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [41] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1763–1767, Dec. 2017.
- [42] S. Li, H. Xu, S. Gao, W. Liu, C. Li, and A. Liu, "An interferometric phase noise reduction method based on modified denoising convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4947–4959, Aug. 2020.
- [43] S. Samiei Esfahany and R. Hanssen, "On the evaluation of second order phase statistics in SAR interferogram stacks," *Earth Observ. Geomatics Eng.*, vol. 1, no. 1, pp. 1–15, 2017.
- [44] M. Elad, "Why simple shrinkage is still relevant for redundant representations?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5559–5569, Dec. 2006.
- [45] C. Rambour, L. Denis, F. Tupin, and H. M. Oriot, "Introducing spatial regularization in SAR tomography reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8600–8617, Nov. 2019.



Jie Li received the bachelor's degree in electronic and information engineering from the Beijing Institute of Technology, Beijing, China, in 2019. She is currently working toward the Ph.D. degree in signal and information processing with the University of Chinese Academy of Sciences, Beijing.

Her research interests include TomoSAR imaging and deep learning.



Zhongqiu Xu received the bachelor's degree in mechatronic engineering from the Beijing Institute of Technology, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in signal and information processing with the University of Chinese Academy of Sciences, Beijing.

His research interests include radar signal processing, compressed sensing, and sparse SAR imaging.



Zhiyuan Li received the bachelor's degree in electronic and information engineering from the Beijing Institute of Technology, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in signal and information processing with the University of Chinese Academy of Sciences, Beijing.

His research interests include deep learning and image processing.



Zhe Zhang (Member, IEEE) was born in Luoyang, China, in 1988. He received the B.Sc. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2008 and 2015, respectively.

During 2015–2020, he was a Postdoctoral Research Scientist with the George Washington University, Washington, DC, USA and George Mason University, Fairfax, VA, USA. He is currently a Professor and Ph.D. Supervisor with the Aerospace Information Research Institute, Chinese Academy of Sciences and Suzhou Aerospace Information Research Institute in Suzhou, Jiangsu, China, since 2021. His research interests include sparse microwave imaging, sparse signal processing, synthetic aperture radar imaging, microwave vision, and the combination of signal processing and deep learning technologies.



Bingchen Zhang received the bachelor's degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 1996, and the M.S. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 1999.

Since 1999, he has been a Scientist with IECAS. His research interests include synthetic aperture radar (SAR) signal processing and airborne SAR system design, implementation, and data processing.



Yirong Wu (Member, IEEE) received the Ms.D. degree in microwave electromagnetic field from the Beijing Institute of Technology, Beijing, China, in 1988, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences (IECAS), Beijing, in 2001.

Since 1988, he has been with IECAS, where he is currently a Director. He has more than 20 years of experience in remote-sensing processing system design. His research interests include microwave imaging, signal and information processing, and related

applications.