

γ -Net: Superresolving SAR Tomographic Inversion via Deep Learning

Kun Qian, Yuanyuan Wang^{id}, *Member, IEEE*, Yilei Shi, *Member, IEEE*, and Xiao Xiang Zhu^{id}, *Fellow, IEEE*

Abstract—Synthetic aperture radar tomography (TomoSAR) has been extensively employed in 3-D reconstruction in dense urban areas using high-resolution SAR acquisitions. Compressive sensing (CS)-based algorithms are generally considered as the state-of-the-art in super-resolving TomoSAR, in particular in the single look case. This superior performance comes at the cost of extra computational burdens, because of the sparse reconstruction, which cannot be solved analytically, and we need to employ computationally expensive iterative solvers. In this article, we propose a novel deep learning-based super-resolving TomoSAR inversion approach, γ -Net, to tackle this challenge. γ -Net adopts advanced complex-valued learned iterative shrinkage thresholding algorithm (CV-LISTA) to mimic the iterative optimization step in sparse reconstruction. Simulations show the height estimate from a well-trained γ -Net approaches the Cramér-Rao lower bound (CRLB) while improving the computational efficiency by one to two orders of magnitude comparing to the first-order CS-based methods. It also shows no degradation in the super-resolution power comparing to the state-of-the-art second-order TomoSAR solvers, which are much more computationally expensive than the first-order methods. Specifically, γ -Net reaches more than 90% detection rate in moderate super-resolving cases at 25 measurements at 6 dB SNR. Moreover, simulation at limited baselines demonstrates that the proposed algorithm outperforms the second-order CS-based method by a fair margin. Test on real TanDEM-X data with just six interferograms also shows high-quality 3-D reconstruction with high-density detected double scatterers.

Index Terms—Complex-valued learned iterative shrinkage thresholding algorithm (LISTA), compressive sensing (CS), synthetic aperture radar (SAR) tomography (TomoSAR), super-resolution.

Manuscript received December 7, 2021; revised March 3, 2022; accepted March 28, 2022. Date of publication April 1, 2022; date of current version April 19, 2022. This work was supported in part by the China Scholarship Council under Grant 201908080038; in part by the European Research Council (ERC), Acronym: So2Sat, under Grant ERC-2016-StG-714087; in part by the Helmholtz Association through the Framework of Helmholtz AI-Local Unit “Munich Unit at Aeronautics, Space and Transport (MASTr)” under Grant ZT-I-PF-5-01; in part by the Helmholtz Excellent Professorship “Data Science in Earth Observation-Big Data Fusion for Urban Research” under Grant W2-W3-100; in part by the German Federal Ministry of Education and Research (BMBF) in the framework of the International Future AI Laboratory through the “Artificial Intelligence for Earth Observation (AI4EO): Reasoning, Uncertainties, Ethics and Beyond” under Grant 01DD20001; and in part by the German Federal Ministry of Economics and Technology in the framework of the “National Center of Excellence ML4Earth” under Grant 50EE2201C. (Corresponding author: Xiao Xiang Zhu.)

Kun Qian is with the Data Science in Earth Observation, Technical University of Munich, 80333 Munich, Germany.

Yuanyuan Wang and Xiao Xiang Zhu are with the Data Science in Earth Observation, Technical University of Munich, 80333 Munich, Germany, and also with the Department of EO Data Science, Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, 82234 Weßling, Germany (e-mail: xiaoxiang.zhu@dlr.de).

Yilei Shi is with the Chair of Remote Sensing Technology, Technical University of Munich, 80333 Munich, Germany.

Digital Object Identifier 10.1109/TGRS.2022.3164193

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) tomography (TomoSAR) [1] has been widely employed for large-scale 3-D urban mapping. It utilizes a stack of SAR acquisitions to reconstruct the reflectivity profile γ along the elevation direction for every azimuth-range pixel. In urban areas, there are usually only a few significant scatterers overlaid in a resolution cell along the elevation direction. Based on this fact, compressive sensing (CS)-based sparse reconstruction algorithms [2]–[4] were introduced to TomoSAR inversion so that we can best unleash the potential of high-resolution SAR data like TerraSAR-X in urban areas. Budillon *et al.* [5] and Zhu and Bamler [6] presented the first simulation of CS TomoSAR, [6] presented the first real data example and [7] proved the super-resolution power of CS for TomoSAR inversion. In recent years, different CS-based methods for solving TomoSAR inversion have been extensively studied, such as SLIMMER [8], truncated singular value decomposition (TSVD)-based CS [9], and alternating direction method of multipliers (ADMM)-based L_1 algorithm [10]. These CS-based algorithms show superiority in super-resolution capability as well as elevation estimate accuracy over the conventional L_2 regularization methods. However, CS-based algorithms usually suffer from high-computational expense and are more challenging to be extended to large-scale processing. An efficient approach was proposed in [11] to address this issue, which is an integration of persistent scatterer interferometry (PSI) and “SLIMMER.” This approach speeds up the processing by pre-classifying the pixels and reducing the percentage of pixels that require SLIMMER for sparse reconstruction. Nevertheless, it did not boost the TomoSAR inversion fundamentally. The same authors of Wang *et al.* [11] also proposed a data-driven method [12], which is based on the CAESAR algorithm [13]. It applies kernel principle component analysis (KPCA) to separate the contribution of individual scatterers before inversion, thus reducing the computational cost logarithmically. Although these algorithms bring a perspective of data-driven approaches in TomoSAR, they still do not strictly solve the SAR tomographic inversion. Their super-resolution capability is also not investigated. Therefore, there has not been a fully data-driven TomoSAR algorithm to date. Hence, we would like to explore the potential of modern data science algorithm such as deep learning, for TomoSAR in this article.

A. Related Work

Recently, deep learning has rapidly developed and been extensively applied in various fields of remote sensing [14],

including SAR data processing [15], thanks to its strong learning power. In particular, a deep neural network can act as an effective nonlinear function and is capable of representing many complicated mathematical models including the CS problems [16]. Several recent studies [17]–[19] have documented the application of deep neural networks in solving sparse reconstruction-related problems in signal processing and remote sensing, i.e., 3-D millimeter-wave sparse imaging and 3-D microwave reconstruction. Motivated by this fact, community started to investigate TomoSAR inversion algorithms based on deep learning and their application since a few years ago. Budillon *et al.* [20] proposed a method to utilize neural networks to detect single scatterer and estimate the corresponding elevation. In [20], TomoSAR inversion was treated as a typical classification problem to detect single scatterer with the classes indicating all the possible discretized positions within the elevation extension of the illuminated scene. Because of its problem formulation, this method cannot be employed in true SAR 3-D imaging, i.e., layover separation. An efficient line spectral estimation algorithm based on deep neural networks was proposed in [21] and applied to tackle the TomoSAR inversion. Experiment results in [21] showed that the method can separate overlaid scatterers and achieves moderate reconstruction performance, whereas the super-resolution power of the proposed method was not systematically analyzed. More recently, a novel super-resolving TomoSAR imaging framework based on CS and deep neural networks was proposed in [22]. It employed CS-based algorithms for preliminary reconstruction and split the elevation range to several subregions with spatial filters. Then a group of deep neural networks-based regression models were trained and applied to each subregion to achieve final super-resolution reconstruction. This method was shown to have unprecedented super-resolution capability. However, the drawback of the proposed algorithm is also obvious. First, the computational complexity of the proposed method is of same order of magnitude to other CS-based algorithms, although the authors increased the sampling distance between two neighbor discrete grid point. The second drawback is that the strong super-resolution power is attributed to adequate training samples, whereas it is arduous to simulate data that imitate the real scattering scenario of high fidelity. Hence, strong overfitting to the training data is expected.

B. Contribution of This Article

The aim of this article is to introduce a computationally efficient and generic TomoSAR algorithm based on deep learning and provide a systematic analysis of its super-resolution power. To this end, we propose a deep learning-based approach to address super-resolving TomoSAR inversion. We unroll iterative shrinkage thresholding algorithm (ISTA) as a complex-valued feedforward neural network with side-connection, named as γ -Net. γ -Net could be trained using data simulated by spatial baselines of given stacked SAR acquisitions. Once well-trained, γ -Net can be directly used for further inference. The main contributions of this article are listed as follows.

- 1) We are the first to introduce a deep unfolded neural network called γ -Net to solve super-resolving TomoSAR inversion. We improved the conventional soft-thresholding function by the piecewise linear function to mitigate the loss of information caused by the learning architecture. Simulations indicate that the piecewise linear function contributes to improving the convergence rate and reducing reconstruction error.
- 2) We are the first to perform a systematic evaluation of a deep learning-based TomoSAR inversion algorithm. We investigated the generalization ability w.r.t. amplitude ratio, phase difference of the interfering scatterers, super-resolution power, and elevation estimation accuracy of the proposed algorithm, i.e., γ -Net. Experiments demonstrate that γ -Net not only approaches almost the same performance in nominal condition comparing to the state of the art but also outperforms the state-of-the-art at limited number of measurements.
- 3) We carried out rigorous analysis of algorithm complexity and proved that γ -Net improves the computational efficiency by one to two orders of magnitude comparing to first-order CS-based methods and shows no degradation in super-resolution power, nor in estimation accuracy comparing to second-order CS-based methods, which are much more computationally expensive than first-order methods. Further time consumption comparison to the second-order CS-based method establishes the superiority of γ -Net in computational efficiency and evidences that γ -Net is able to realize large-scale super-resolving TomoSAR processing, whereas second-order CS-based solvers can only be applied to small laboratory samples.

The remainder of the article is outlined as follows: In Section II, the TomoSAR imaging model and inversion are briefly introduced. Section III provides an overview on the formulation of γ -Net. Results of systematic evaluation, using simulated and real data, are presented and discussed in Sections IV and V. Finally, the conclusion of this article is drawn in Section VI.

II. SAR IMAGING MODEL AND PROBLEM FORMULATION

First, we would like to introduce the TomoSAR imaging model (see Fig. 1). For a single SAR acquisition at aperture position b_n , the complex-valued measurement at an azimuth-range pixel for the n th acquisition is the integral of the reflected signal along the elevation direction and can be expressed as follows:

$$g_n = \int_{\Delta s} \gamma(s) \exp(-j2\pi \zeta_n s) ds, \quad n = 1, \dots, N \quad (1)$$

where $\gamma(s)$ depicts the reflectivity profiles along the elevation direction. $\zeta_n = -2b_n/(\lambda r)$ denotes the elevation frequency. λ and r refer to the wavelength and the range, respectively. By discretizing the reflectivity profile along the elevation direction s , the approximated system model reads

$$g_n \approx \delta s \cdot \sum_{l=1}^L \gamma(s_l) \exp(-j2\pi \zeta_n s_l), \quad n = 1, \dots, N \quad (2)$$

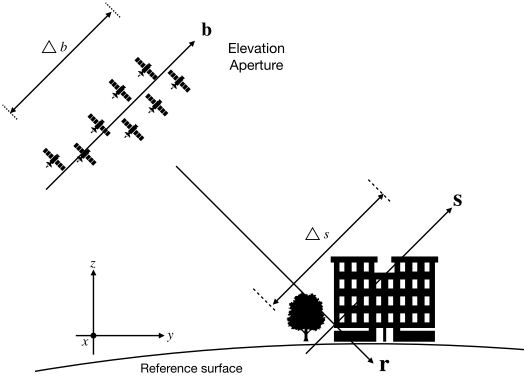


Fig. 1. SAR imaging geometry. The elevation synthetic aperture is built up by SAR data acquired from slightly different viewing angles. Flight direction is orthogonal into the plane.

where L is the number of the discrete elevation indices and $\delta s = \Delta s / (L - 1)$ is the sampling distance with Δs depicting the whole extent of the reflectivity profile along the elevation direction. In the presence of noise ϵ , the discrete TomoSAR imaging model can be expressed as follows:

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \epsilon \quad (3)$$

where $\mathbf{g} \in \mathbb{C}^{N \times 1}$ is the complex-valued SAR measurement vector, $\mathbf{R} \in \mathbb{C}^{N \times L}$ is the steering matrix with $R_{nl} = \exp(-j2\pi \zeta_n s_l)$, and $\boldsymbol{\gamma} \in \mathbb{C}^{L \times 1}$ denotes the discrete reflectivity profile vector. TomoSAR inversion is aimed at retrieving the reflectivity profile for each range-azimuth cell, then estimating the corresponding scattering parameters such as the number of scatterers and their elevation and reflectivity.

For TomoSAR reconstruction in urban areas, it is shown in [7] that there are rarely more than a few (0–4) scatterers overlaid along the elevation direction in each resolution unit, namely, the reflected signal along the elevation direction is sufficiently sparse. The ideal sparse solution of $\boldsymbol{\gamma}$ is obtained by solving (3) with the L_0 -norm regularization, which is, however, a NP-hard problem. For our application, it is shown in [2]–[4], [7] that the L_0 -norm minimization can be approximate by the L_1 -norm minimization, which can be expressed as follows:

$$\hat{\boldsymbol{\gamma}} = \arg \min_{\boldsymbol{\gamma}} \{ \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \} \quad (4)$$

where λ is a regularization parameter balancing the sparsity and data-fitting terms. It should be adjusted according to the noise level as well as the desired sparsity level. The choice of a proper λ is described in great detail in [23]. The L_2 – L_1 mixed norm minimization (4) is also known as basis pursuit denoising (BPDN) [23] and can be formulated as least absolute shrinkage and selection operator (LASSO) in some condition. Conventional solvers for (4) are either first- or second-order methods. First-order methods are typically based on linear approximations, such as ISTA [24] and ADMM [25]. An example for the second-order methods is the primal-dual interior-point method (PDIPM) [26]. Second-order methods often suffer from high-computational cost, thus impeding their application in large-scale processing.

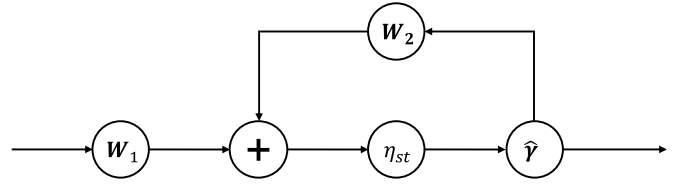


Fig. 2. RNN structure of ISTA by viewing an iteration of ISTA as a layer of the RNN.

III. TOMOSAR INVERSION VIA γ -NET

A. Background of Complex-Valued ISTA

ISTA [24] is a popular method to solve the L_2 – L_1 mix norm minimization. Each iteration of ISTA is defined by

$$\hat{\boldsymbol{\gamma}}_i = \eta_{st}(\hat{\boldsymbol{\gamma}}_{i-1} + \beta \mathbf{R}^H \mathbf{b}_{i-1}, \theta_i) \quad (5)$$

with

$$\mathbf{b}_i = \mathbf{g} - \mathbf{R}\hat{\boldsymbol{\gamma}}_i$$

where $\hat{\boldsymbol{\gamma}}_0 = \mathbf{0}$, β is the stepsize, η_{st} is the soft-thresholding function applied to each element of $\hat{\boldsymbol{\gamma}}_i$, and θ is the threshold in the soft-thresholding function. The complex-valued version of the soft-thresholding function η_{st} is defined by

$$\eta_{st}(\hat{\boldsymbol{\gamma}}_i, \theta_i) = e^{j \cdot \angle(\hat{\boldsymbol{\gamma}}_i)} \max(|\hat{\boldsymbol{\gamma}}_i| - \theta_i, 0) \quad (6)$$

where j is the imaginary number. In each iteration of ISTA, the estimate is first optimized via gradient decent and then the soft-thresholding function is applied to prune the elements with small magnitude, thus promoting the sparsity of the final estimate.

B. Complex-Valued LISTA Formulation for TomoSAR

As presented in the review article [27], (5) can be rewritten as the following form:

$$\hat{\boldsymbol{\gamma}}_i = \eta_{st} \{ \mathbf{W}_1^i \mathbf{g} + \mathbf{W}_2^i \hat{\boldsymbol{\gamma}}_{i-1}, \theta_i \} \quad (7)$$

where $\mathbf{W}_1^i = \beta \mathbf{R}^H$ and $\mathbf{W}_2^i = \mathbf{I} - \beta \mathbf{R}^H \mathbf{R}$.

If we regard the soft-thresholding function in (7) as an activation function, we find that (7) is the basic form of the i th layer of a recurrent neural network (RNN). Therefore, ISTA can be viewed as a RNN illustrated in Fig. 2.

Inspired by the connection between ISTA and RNN, a learning-based model named Learned ISTA (LISTA) was proposed in [28]. Fig. 3 demonstrates the learning architecture of a K -layer LISTA, which unrolls the RNN and truncates it into K iterations, thus leading to a K -layer side-connected feedforward neural network. The major difference between ISTA and LISTA is that the weight matrices \mathbf{W}_1^i , \mathbf{W}_2^i as well as the threshold θ_i in each layer of LISTA are not predetermined. Those parameters are learned in the LISTA neural network from training data.

The loss function of LISTA over the training data $\{(\mathbf{g}_i, \boldsymbol{\gamma}_i)\}_{i=1}^T$ is the mean square error (MSE) loss described as follows:

$$\underset{\Psi}{\text{minimize}} \mathcal{L}(\Psi) = \frac{1}{T} \sum_{i=1}^T \|\hat{\boldsymbol{\gamma}}(\Psi, \mathbf{g}) - \boldsymbol{\gamma}\|_2^2 \quad (8)$$

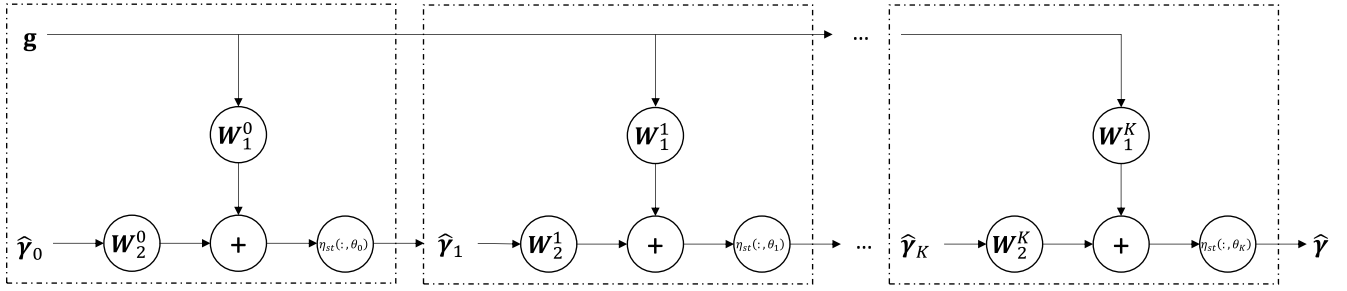


Fig. 3. Unfolded LISTA architecture. A K -layer LISTA unrolls the RNN and truncates it into K iterations, thus leading to a side-connected feedforward neural network.

where T denotes the number of samples in the training data and $\Psi = \{\mathbf{W}_1, \mathbf{W}_2, \theta\}$ is the set of free parameters to be learned. Many recent works [28]–[32] have demonstrated that LISTA is able to achieve the same estimation accuracy within two to three order-of-magnitude fewer iterations than the original ISTA. Moreover, empirical results show that LISTA has better generalization ability.

However, to apply LISTA to solve TomoSAR inversion, we need to extend LISTA to complex-valued domain. Complex-valued learned iterative shrinkage thresholding algorithm (CV-LISTA) shares the same learning architecture as LISTA, except that each neuron in CV-LISTA has two channels, which refer to the real and imaginary part of a complex number, respectively. We applied the following adaptations to (7):

$$\tilde{\mathbf{y}}_i = \eta_{st}\{\tilde{\mathbf{W}}_1^i \tilde{\mathbf{g}} + \tilde{\mathbf{W}}_2^i \tilde{\mathbf{y}}_{i-1}, \theta_i\} \quad (9)$$

where

$$\begin{aligned} \tilde{\mathbf{W}}_j^i &= \begin{bmatrix} \Re(\mathbf{W}_j^i) & -\Im(\mathbf{W}_j^i) \\ \Im(\mathbf{W}_j^i) & \Re(\mathbf{W}_j^i) \end{bmatrix} \\ \tilde{\mathbf{g}} &= \begin{bmatrix} \Re(\mathbf{g}) \\ \Im(\mathbf{g}) \end{bmatrix} \\ \tilde{\mathbf{y}} &= \begin{bmatrix} \Re(\hat{\mathbf{y}}) \\ \Im(\hat{\mathbf{y}}) \end{bmatrix} \end{aligned} \quad (10)$$

with $j = 1, 2$ and $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary operators, respectively.

C. γ -Net Formulation for TomoSAR

Through our research and experiments, we discovered a few drawbacks of CV-LISTA applying to TomoSAR and proposed several novel improvements in γ -Net. The improvements are mainly threefold. First, the \mathbf{W}_1 and \mathbf{W}_2 matrices in the aforementioned CV-LISTA are highly correlated. Hence, as proposed in [33], a weight coupling structure is introduced to reduce redundant trainable parameters in γ -Net. Second, we tried to use the acceleration technique, support selection, which was originally developed for LASSO to boost the convergence in γ -Net. Last, we replaced the conventional soft-thresholding function by the piecewise linear function because the conventional soft-thresholding function causes information loss, which leads to large reconstruction error and decreases the convergence rate of the model. We will discuss the improvements in detail in the following paragraphs.

1) *Weight Coupling*: Instead of training γ -Net as pure “black-box” networks, we simplify the CV-LISTA and propose γ -Net by exploiting the dependency among the trainable weights. Details can be found in [33] that the weights to be learned $\{(\mathbf{W}_1^i, \mathbf{W}_2^i)\}_{i=0}^K$ in each layer asymptotically satisfy the following partial weight coupling structure:

$$\mathbf{W}_2^i = \mathbf{I} - \mathbf{W}_1^i \mathbf{R}. \quad (11)$$

By employing the partial weight coupling structure, we can simplify the i th layer of γ -Net to

$$\tilde{\mathbf{y}}_i = \eta_{st}\{\tilde{\mathbf{y}}_{i-1} + \tilde{\mathbf{W}}^i (\tilde{\mathbf{g}} - \tilde{\mathbf{R}} \tilde{\mathbf{y}}_{i-1}), \theta_i\} \quad (12)$$

with

$$\tilde{\mathbf{R}} = \begin{bmatrix} \Re(\mathbf{R}) & -\Im(\mathbf{R}) \\ \Im(\mathbf{R}) & \Re(\mathbf{R}) \end{bmatrix}$$

where $(\Re(\mathbf{W}^i), \Im(\mathbf{W}^i), \theta^i)$ are the parameters to be learned in the i th layer, and the trainable weight \mathbf{W}^i is initialized using the system measurement matrix \mathbf{R} with $\mathbf{W}^i = \beta \mathbf{R}^H$. The coupled structure contributes to eliminating the number of free parameters to be trained, thus accelerating the training procedure significantly. Theoretically speaking, (11) can only be satisfied for deep layers. However, extensive simulations in [33] demonstrate that the application of the partial weight coupling structure to every layer will not degrade the theoretical and experimental performance.

2) *Support Selection*: In addition to the application of the weight coupling structure, we introduce a special thresholding scheme in γ -Net, called *support selection*, which is inspired by “kickin” in linearized Bregman iteration [34]. Meaning, we will select a certain percentage of entries with largest magnitude at each layer of γ -Net before the shrinkage step. Hereafter, the selected part will be trusted as “true support” and directly fed into the following layer, bypassing the shrinkage step. The remaining entries will go through the shrinkage step as usual. Assuming that ρ^i percentage of entries are trusted in the i th layer, the support selection can be formally defined as

$$\eta_{ss\theta_i}^{\rho^i}(\tilde{\mathbf{y}}_i) = \begin{cases} \tilde{\mathbf{y}}_i, & i \in S^{\rho^i}(\tilde{\mathbf{y}}) \\ \eta_{st}(\tilde{\mathbf{y}}_i, \theta_i), & i \notin S^{\rho^i}(\tilde{\mathbf{y}}) \end{cases} \quad (13)$$

where $S^{\rho^i}(\mathbf{y})$ contains the entries with the ρ^i largest magnitudes. It is worth mentioning that the percentage ρ^i is a

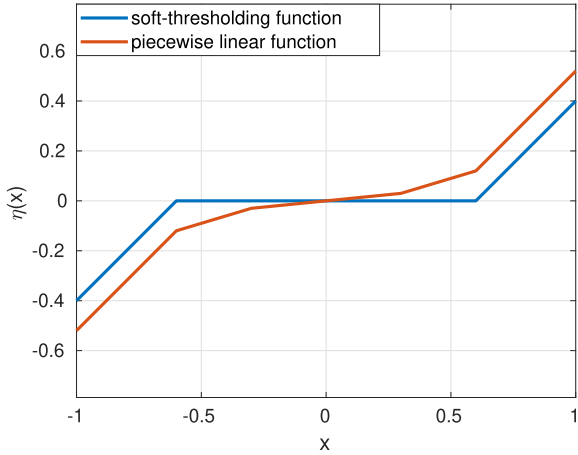


Fig. 4. Comparison between the piecewise linear function and soft-thresholding function. Instead of pruning the elements with small magnitude, the piecewise linear function just further minimizes them, thus possibly avoiding the information loss.

hyperparameter that requires manual tuning. When we apply the support selection to γ -Net, then (12) is modified as

$$\tilde{\mathbf{y}}_i = \eta_{ss\theta_i}^{p_i} \{ \tilde{\mathbf{y}}_{i-1} + \tilde{\mathbf{W}}^i (\tilde{\mathbf{g}} - \tilde{\mathbf{R}} \tilde{\mathbf{y}}_{i-1}), \theta_i \}. \quad (14)$$

Simulation experiments in [33] support that introducing the support selection on the one hand improves the convergence rate both theoretically and empirically, on the other hand, it contributes to reducing the recovery error, thus improving the estimation accuracy.

3) *Piecewise Linear Thresholding Function*: The conventional soft-thresholding function simply prunes elements with small magnitude to zero, which is very likely to result in information loss. To maintain useful information as much as possible and execute the shrinkage step in the meanwhile, we replace the soft-thresholding function by the piecewise linear function $\eta_{pwl}(\hat{\mathbf{y}}, \theta)$ in γ -Net, which is defined as

$$\eta_{pwl}(\hat{\mathbf{y}}) = \begin{cases} \theta_3 \hat{\mathbf{y}}, & |\hat{\mathbf{y}}| \leq \theta_1 \\ e^{j \cdot \hat{\mathbf{y}}} [\theta_4 (|\hat{\mathbf{y}}| - \theta_1) + \theta_3 \theta_1], & \theta_1 < |\hat{\mathbf{y}}| \leq \theta_2 \\ e^{j \cdot \hat{\mathbf{y}}} [\theta_5 (|\hat{\mathbf{y}}| - \theta_2) + \theta_4 (\theta_2 - \theta_1) + \theta_3 \theta_1], & |\hat{\mathbf{y}}| > \theta_2. \end{cases} \quad (15)$$

To clarify, the symbol j in (15) refers to the imaginary number.

Fig. 4 compares both functions. As can be seen, instead of pruning elements with small magnitude, the piecewise linear function only down scale them. Hence, it mitigates the information loss. However, it leads to the consequence that the final output of γ -Net being not strictly sparse. Most elements of the final output are not driven to zero strictly but to some extremely small values. Therefore, an additional post-processing step for cleaning the elements with extremely small magnitude is necessary when we employ the piecewise linear function.

Fig. 5 compares γ -Net performance in term of the normalized MSE (NMSE) under the two shrinkage functions.

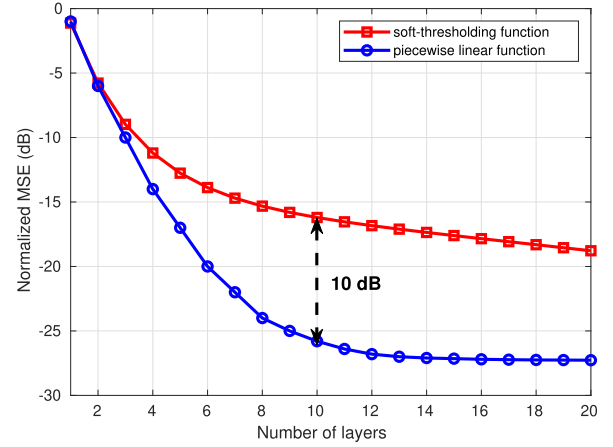


Fig. 5. Performance of γ -Net using different shrinkage function. The piecewise linear function conduces to faster convergence and improves the estimation accuracy.

To clarify, the performance of γ -Net with different shrinkage functions was verified on a set of noise-free data so that the results reflect the ideal performance. The NMSE is defined as

$$\text{NMSE} = \frac{1}{T} \sum \frac{\|\hat{\mathbf{y}} - \mathbf{y}\|_2^2}{\|\mathbf{y}\|_2^2}. \quad (16)$$

From this figure, it can be seen that γ -Net with the piecewise linear function achieves lower NMSE at the same number of layers. In another word, the piecewise linear function improves the estimation accuracy of γ -Net or the convergence rate. Specifically, γ -Net with the piecewise linear function requires only about 12 layers to achieve convergence. However, it is obvious that much more layers are required when the conventional soft-thresholding function is employed.

D. Algorithm Summary

To achieve super-resolution ability and high elevation estimation accuracy, it is usually required to sample the elevation range much denser than the elevation resolution unit, thus rendering the steering matrix \mathbf{R} severely overcomplete, reducing its restricted isometric property (RIP) and increasing its coherence. The violation of RIP and incoherence introduces outliers to estimates of the reflectivity profile $\hat{\mathbf{y}}$ [35]. Additionally, outliers will be caused by the noise interference as well. Hence, we need further perform model order selection [36] to suppress the undesired outliers and estimate the number and location of scatterers precisely, which are typical steps in TomoSAR. The proposed super-resolving TomoSAR inversion algorithm is a combination of γ -Net and model order selection, and re-estimation. The basic workflow of the proposed algorithm is show in Algorithm 1. The model order selection is conducted based on Bayesian information criterion (BIC) [37].

IV. PERFORMANCE EVALUATION

A. Data Simulation and Training

1) *Simulation Setup*: We simulate the data using a setting similar to [8]. Specifically, SAR measurements with 25 spatial baselines are simulated. The spatial baselines are regularly

Algorithm 1 Summary of the Proposed Algorithm**Simulate training data****Sampling** the elevation extent**Generate** steering matrix \mathbf{R} with

$$R_{nl} = \exp(-j2\pi \xi_n s_l), \text{ where } \xi_n = -2b_n/(\lambda r)$$

Simulate reflectivity profile $\boldsymbol{\gamma}$ **Simulate** SAR measurements with $\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ **Finish** the generation of training data $\{(\mathbf{g}_i, \boldsymbol{\gamma}_i)\}_{i=1}^T$ **Training of $\boldsymbol{\gamma}$ -Net****Over given training samples** $\{(\mathbf{g}_i, \boldsymbol{\gamma}_i)\}_{i=1}^T$

$$\underset{\boldsymbol{\Psi}}{\text{minimize}} \mathcal{L}(\boldsymbol{\Psi}) = \frac{1}{T} \sum_{i=1}^T \|\hat{\boldsymbol{\gamma}}(\boldsymbol{\Psi}, \mathbf{g}) - \boldsymbol{\gamma}\|_2^2$$

where $\boldsymbol{\Psi} = [\Re(\mathbf{W}), \Im(\mathbf{W}), \boldsymbol{\theta}]$ **for** each pixel in the image: **do****Preliminary estimate** via $\boldsymbol{\gamma}$ -Net:

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}\text{-Net}(\mathbf{g})$$

Model order selection to remove outliers:

$$\hat{P} = \underset{P}{\text{argmin}} \{\sigma_\varepsilon^{-2} \|\mathbf{g} - \mathbf{R}\hat{\boldsymbol{\gamma}}\|_2^2 + 1.5 P \ln N\}$$

Determine the number of scatterers**Final estimation** of their elevation**end for**

distributed in the range from -135 to 135 m, thus leading to a Rayleigh resolution of around 42 m. We simulated ca. four million training samples, half of which are single scatterer and the others are overlaid double scatterers. The simulation details of single and double scatterers are listed below. We randomized many parameters in the simulation, to make the simulation more realistic.

- 1) *Single Scatterers*: for a single scatterer, the scattering coefficient is a complex number $\gamma = A \cdot \exp(j\phi)$, with the amplitude being deterministic and the scattering phase following an uniform distribution, i.e., $\phi \sim U(0, 2\pi)$. To randomize the amplitude in the simulation, we simulate it with a uniform distribution as well, i.e., $A \sim U(1, 4)$, although a real SAR amplitude image shows more Rayleigh or Gamma distribution. The elevation of the simulated scatterer is regularly distributed in the range from 0 to 200 m with 1 m sampling. Once the location of the scatterer is determined, the echo signal is simulated at 11 different levels of SNR, which is regularly distributed between $[0, 10]$ dB.
- 2) *Double Scatterers*: For double scatterers, the generation of the two single scatterers is identical to the previous step, i.e., for each individual scatterer, the phase follows an uniform distribution $\phi \sim U(0, 2\pi)$ and the amplitude follows an uniform distribution $A \sim U(1, 4)$, respectively. As a result, different amplitude ratio and phase difference of the simulated double scatterers can be covered. We also vary the elevation distance between the two single scatterers. The elevation distance varies from 0.1 until 1.2 Rayleigh resolution, with a regular sampling of 0.1 Rayleigh resolution. The elevation of the first scatterer follows an uniform distribution in the range of 0 to 200 m. To avoid the off-grid bias, we assume that all scatterers locate on-grid with 1 m sampling.

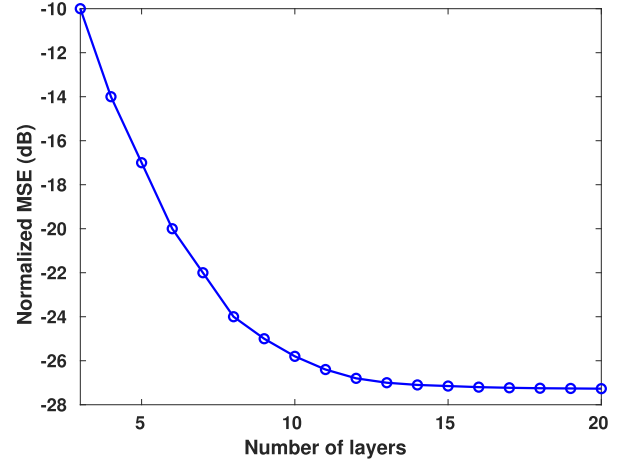


Fig. 6. $\boldsymbol{\gamma}$ -Net performance w.r.t. the number of layers. After 12 layers, the performance improvement of $\boldsymbol{\gamma}$ -Net is marginal with the increase number of layers. Instead, the increase of layers leads to heavier computational burden.

2) *Training*: The training was carried out using Pytorch [38] and the Adam optimizer [39]. The learning rate was initialized at 0.0005 and adjusted adaptively during the training. In the training procedure, we gradually increase the number of the layers from 3 to 20 to determine an optimal network structure. We validate the performance of $\boldsymbol{\gamma}$ -Net with different number of layers on a validation dataset. The validation dataset contains 0.2 million noise-free samples simulated using the same settings mentioned in the simulation setup, so that we can compare the theoretical performance of $\boldsymbol{\gamma}$ -Net with different number of layers. Fig. 6 illustrates the performance of $\boldsymbol{\gamma}$ -Net w.r.t. the number of its layers. Closer inspection of Fig. 6 shows that the NMSE first decreases rapidly and then starts to converge to a minimum at around twelve layers. Simultaneously, the increase of the number of layers leads to heavier computation cost. Therefore, $\boldsymbol{\gamma}$ -Net employed in this article contains just twelve layers. On the one hand, $\boldsymbol{\gamma}$ -Net with twelve layers is able to guarantee the estimation accuracy; on the other hand, it maintains the computational efficiency.

B. Single Scatterer Analysis

In addition to the simulated training data, we simulated four sets of testing data for the single scatterer analysis with SNR = $\{0, 3, 6, 10\}$ dB. Each set is composed of 0.2 million samples. We use the proposed algorithm to detect the single scatterer and estimate the corresponding elevation. Fig. 7 demonstrates the estimated reflectivity profile using the trained $\boldsymbol{\gamma}$ -Net and singular value decomposition (SVD)-Wiener [36] (a conventional nonsuperresolving algorithm). As we can see, although both of $\boldsymbol{\gamma}$ -Net and SVD-Wiener are able to detect the position of the single scatterer, $\boldsymbol{\gamma}$ -Net reconstructs spectral lines instead of sinc-like point response function, thus mitigating the sidelobe problem. Moreover, from Fig. 7(a)–(d), we can see that the outliers caused by noise interference exist in the reflectivity profile estimate of $\boldsymbol{\gamma}$ -Net. Therefore, further model order selection step is required.

Table I provides the results after model order selection. The Cramér–Rao lower bound (CRLB), the estimates mean

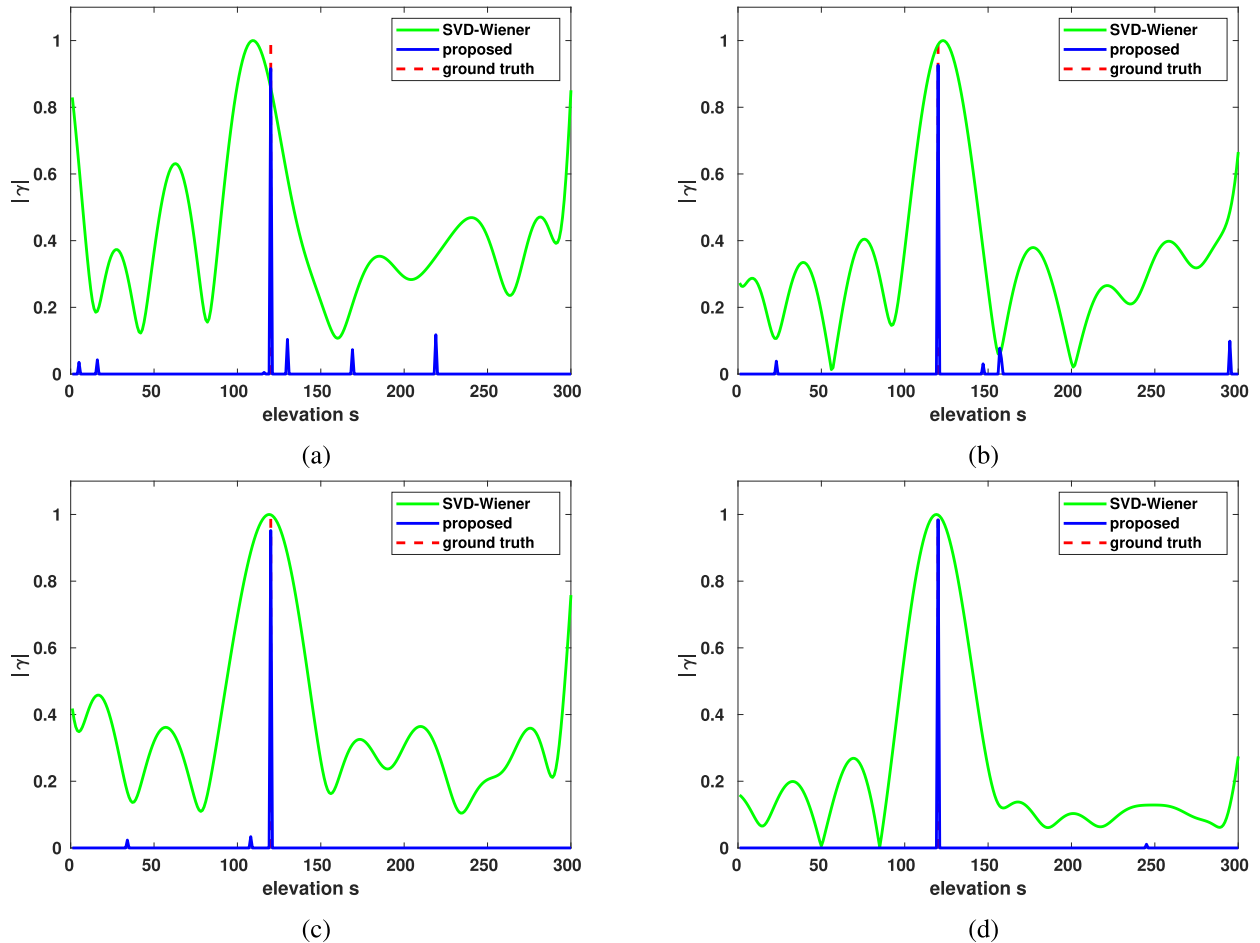


Fig. 7. Estimated reflectivity profile of simulated data with single scatterer at different SNR. (a) SNR = 0 dB. (b) SNR = 3 dB. (c) SNR = 6 dB. (d) SNR = 10 dB.

TABLE I

STATISTICS OF THE ESTIMATE OF SINGLE SCATTERER USING THE PROPOSED ALGORITHM. μ AND σ DENOTE THE SAMPLE MEAN AND THE CORRESPONDING STANDARD DEVIATION, RESPECTIVELY. THE PROPOSED ALGORITHM IS ABLE TO DETECT THE SINGLE SCATTERER IN NEARLY ALL CASES WITH THE STANDARD DEVIATION APPROACHING THE CRLB AND BIAS APPROACHING ZERO

SNR (dB)	effective detection rate	CRLB (normalized)	σ (normalized)	μ (normalized)
0	94.19%	7×10^{-2}	9×10^{-2}	9×10^{-3}
3	96.34%	5×10^{-2}	6×10^{-2}	5×10^{-3}
6	98.81%	3×10^{-2}	3×10^{-2}	2×10^{-3}
10	99.79%	2×10^{-2}	2×10^{-2}	6×10^{-4}

(μ) and standard deviation (σ) in Table I are normalized to the Rayleigh resolution. Since the goal of TomoSAR is to have a good elevation estimate, and also a high detection rate, we define the term *effective detection rate*. An effective detection of single scatterer should satisfy the following two conditions: 1) only one scatterer is detected and 2) the estimated elevation should not exceed ± 3 times CRLB w.r.t. the ground truth. It is apparent from this table that the proposed algorithm is able to detect almost all single scatterer at different SNRs. Further statistics on mean value μ and standard deviation σ of the estimation error indicate high estimation accuracy of the proposed algorithm with the bias approaching zero and the standard deviation approaching the CRLB.

C. Double Scatterers Analysis

For the double scatterer analysis, we simulate two-scatterer mixtures in the experiments and a systematic evaluation was carried out regarding the distance between simulated double scatterers, different scatterers amplitude ratio as well as phase difference between the double scatterers.

1) Performance With Respect to Scatterers Distance:

In this experiment, we performed a well-known TomoSAR benchmark test [7], [36]. We simulated double scatterers with increasing elevation distance between the two layovered scatterers, to mimic a facade-ground interaction. Since we focus on the super-resolution regime, the elevation distance d_s between the two overlaid scatterers is set to be no larger than 1.2 times of the Rayleigh resolution. Two different scenarios

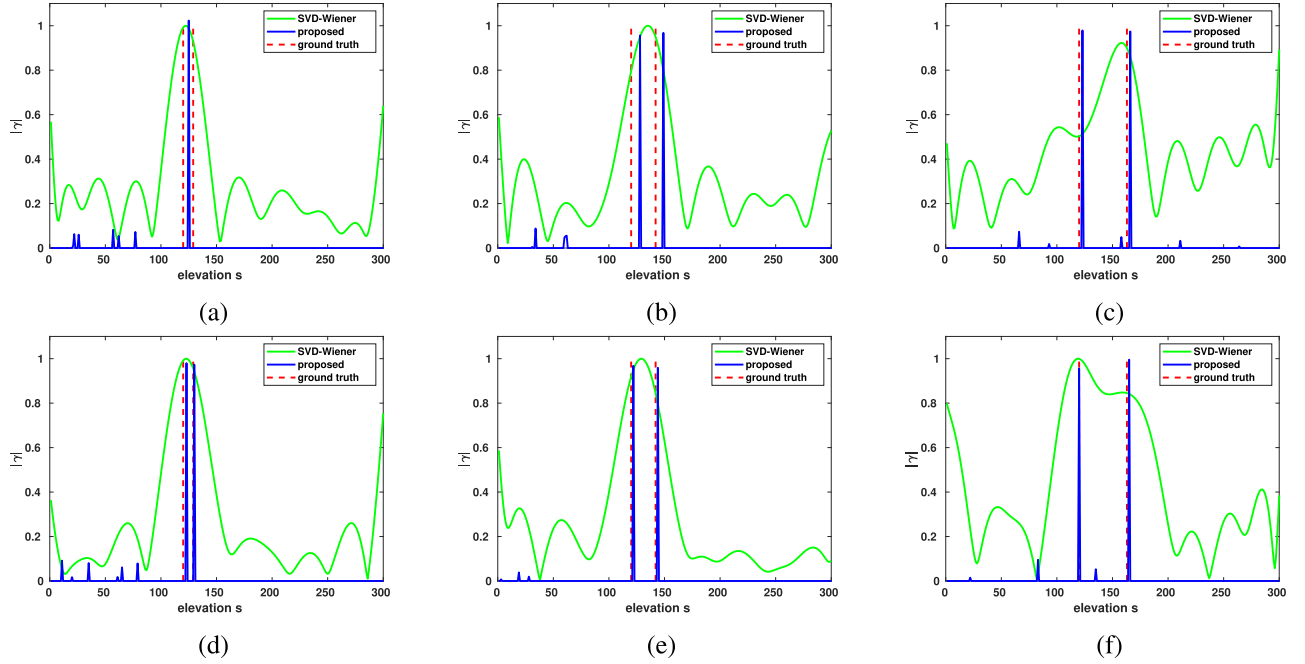


Fig. 8. Estimated reflectivity profile of simulated data with overlaid double scatterers. (a) SNR = 0 dB, $\alpha = 0.2$. (b) SNR = 0 dB, $\alpha = 0.5$. (c) SNR = 0 dB, $\alpha = 1$. (d) SNR = 6 dB, $\alpha = 0.2$. (e) SNR = 6 dB, $\alpha = 0.5$. (f) SNR = 6 dB, $\alpha = 1$.

were taken into consideration with $\text{SNR} \in \{0, 6\}$ dB, which represent typical SNR levels in a high-resolution spaceborne SAR image. Fig. 8 demonstrates some examples of the estimated reflectivity profile at the normalized elevation distance $\alpha = [0.2, 0.5, 1.0]$. The normalized distance α is defined as the ratio of the elevation distance between the double scatterers and the Rayleigh resolution ρ_s , formally expressed as

$$\alpha = \frac{d_s}{\rho_s}. \quad (17)$$

From Fig. 8, one can see that both the trained γ -Net and SVD-Wiener are able to distinguish the overlaid double scatterers in the nonsuperresolving case, i.e., last column, when $\alpha = 1.0$. But comparing to SVD-Wiener, γ -Net provides much higher elevation estimation accuracy. Moreover, when we move the double scatterers closer into the Rayleigh resolution, SVD-Wiener fails to separate them. In the contrast, γ -Net is still capable of detecting the double scatterers in most cases, which exhibits its super-resolution power.

Hereafter, we compare the proposed algorithm with the state-of-the-art SLIMMER algorithm [8] focusing on the detection rate and the estimation accuracy. Similar to the single scatterer case, we use the *effective detection rate* to fairly evaluate the detection rate. An effective detection of double scatterer is defined as.

- 1) The hypothesis test correctly decides two scatterers for a double scatterers' signal.
- 2) The estimated elevation of *both* detected double scatterers are within ± 3 times CRLB w.r.t. their true elevation.
- 3) Both elevation estimates are also within $\pm 0.5 d_s$ w.r.t. their true elevation.

The third criterion is seldom seen in the literature. However, it is necessary, because in extremely super-resolving cases,

three times CRLB will become much larger than the elevation distance. Hence, it cannot be used as an accountable measure for reasonable estimates. $\pm 0.5 d_s$ is a much stricter constraint in such cases, which will reflect the true performance of the algorithm.

Fig. 9 compares the effective detection rate P_d of SLIMMER and the proposed algorithm for the case $N = 25$. For each pair of (SNR, α) , 0.2 million Monte Carlo trials for the worst case in TomoSAR inversion, i.e., the double scatterers have the same amplitude and phase, were simulated. The effective detection rate P_d is presented as a function of the normalized distance. The red and blue polylines illustrate the results of the proposed algorithm and SLIMMER, respectively. As we can see from Fig. 9, the proposed algorithm has comparable super-resolution power as SLIMMER.

Fig. 10 demonstrates the elevation estimates of simulated façade and ground w.r.t. the true normalized elevation distance. In each subplot of Fig. 10, the two red line segments represent the true elevation of the simulated façade and ground, respectively, while the dashed lines show the true elevation $\pm 1 \times \text{CRLB}$ (normalized). Exhaustive details of the derivation of the CRLB can be found in [8]. The elevation estimates of simulated façade and ground are plotted with each dot depicting the sample mean of all estimates at the given normalized distance and the error bar indicating the corresponding standard deviation. Points below an effective detection rate of 10% were not plotted in the figure. As it is shown in Fig. 10, the proposed algorithm shows higher elevation estimation accuracy than SLIMMER. To be specific, at 0 dB SNR, although both the proposed algorithm and SLIMMER have similar estimate bias, the proposed algorithm leads to much smaller variance. In high SNR case, the proposed algorithm outperforms SLIMMER in super-resolving

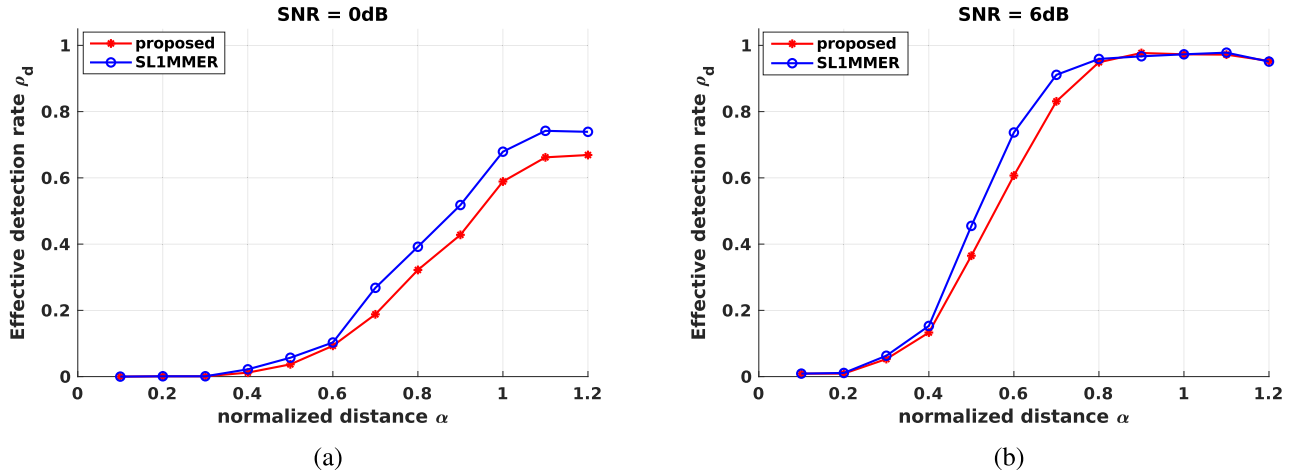


Fig. 9. Detection rate P_d as a function of the normalized elevation distance between the simulated facade and ground using the proposed algorithm (dashed star) and SLIMMER (dashed circle) with SNR = 0 and 6 dB, $N = 25$ and phase difference $\Delta\phi = 0$ (worst case) under 0.2 million Monte Carlo trials. Comparison in (a) 0 dB and (b) 6 dB.

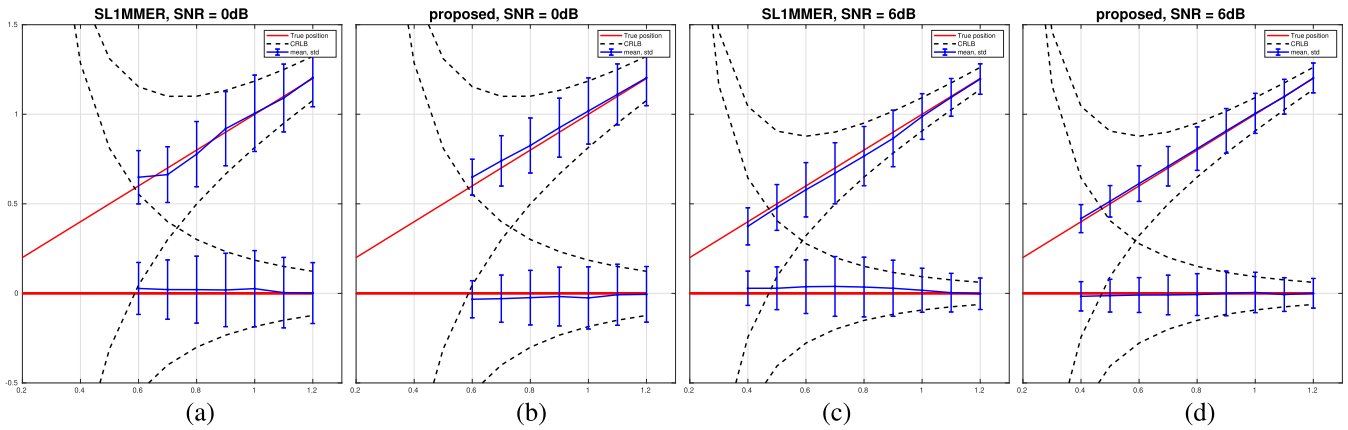


Fig. 10. Estimated elevation of simulated facade and ground, (a) SNR = 0 dB with SLIMMER, (b) SNR = 0 dB with the proposed algorithm, (c) SNR = 6 dB with SLIMMER, and (d) SNR = 6 dB with the proposed algorithm. Each dot has the sample mean of all estimates as its y value and the correspond standard deviation as error bar. The red line segments represent the true elevation of the simulated facade and ground. The dashed curves denote the true elevation $\pm 1 \times \text{CRLB}$ normalized w.r.t. the Rayleigh resolution.

cases w.r.t. the elevation estimation accuracy. As can be seen that, SLIMMER suffers from much larger elevation estimate bias as well as the standard deviation.

2) *Performance With Respect to Amplitude Ratio:* This simulation sets out to evaluate the performance of the proposed algorithm w.r.t. different amplitude ratio of the double scatterers. Fig. 11 illustrates us the effective detection rate of the proposed algorithm at different amplitude ratio. As can be seen, the effective detection rate decreases with the increasing amplitude ratio. Since γ -Net promotes sparsity by shrinking elements with small magnitude layer by layer. With the increase of the amplitude ratio between simulated double scatterers, the darker scatterer becomes less prominent, and hence easier to be ignored. Therefore, at high amplitude ratio, the proposed algorithm tends to detect single scatterer with dominant amplitude. However, from our perspective, it will not affect the application of the proposed algorithm. In real-world processing, if one scatterer is much more prominent than others in a pixel, we can usually judge that this pixel contains only a single scatterer by viewing others as noise.

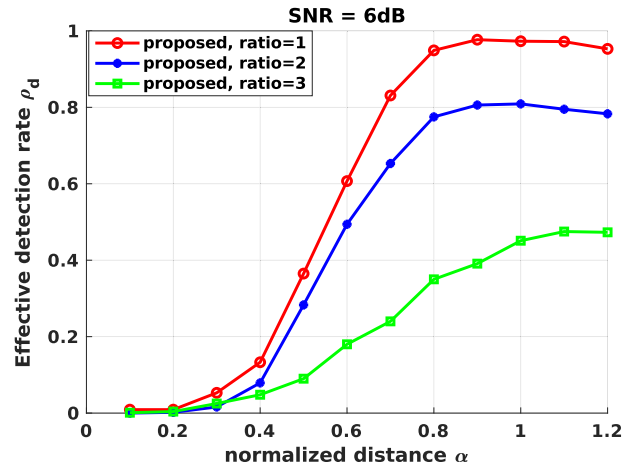


Fig. 11. Effective detection rate ρ_d as a function of amplitude ratio at 6 dB SNR.

3) *Performance With Respect to Phase Difference:* We varied the phase difference between the simulated double scatterers in this simulation to further verify the generalization

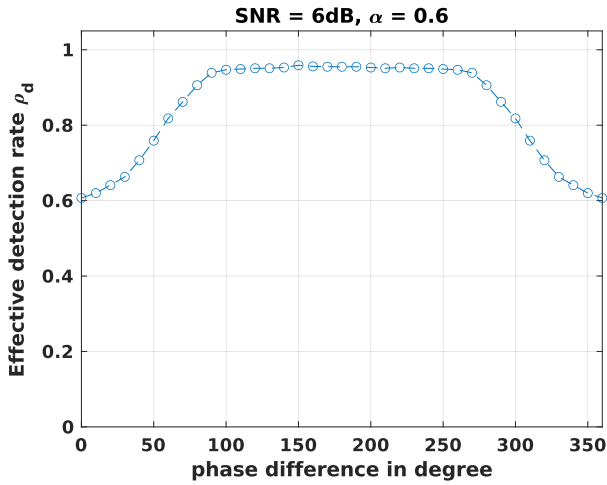


Fig. 12. Effective detection rate ρ_d as a function of phase difference $\Delta\phi$ under the case: $N = 25$, $\text{SNR} = 6$ dB and $\alpha = 0.6$.

TABLE II
STATISTICS OF EVALUATION ON FALSE DETECTION

pure noise detected as	percentage
0 scatterer	95.57%
1 scatterer	4.33%
2 scatterers	0.1%

ability of the proposed algorithm. Fig. 12 demonstrates us an example of the effective detection rate when $N = 25$ and $\text{SNR} = 6$ dB, with the normalized distance $\alpha = 0.6$. The double scatterers in the simulation are set to have identical amplitude. As we can see, although the phase difference $\Delta\phi$ affects the performance, the proposed algorithm is still capable of providing satisfactory super-resolution power even in the worst case, where the phase difference $\Delta\phi = 0$.

D. Analysis of False Detection

In this section, we will provide a quantitative assessment about false detection. We used the proposed algorithm to detect 0.2 million samples containing 0 scatterer, i.e., pure noise. As it is shown in Table II, the proposed algorithm is able to distinguish almost all samples of noise. Less than 5% samples are falsely detected as single scatterer and only about 0.1% as double scatterers. The low false alarm attributes to the powerful model order selection with known noise variance in the simulation. However, in real-world application, the noise variance needs to be estimated. Therefore, Table II shows the upper limit.

E. Performance at Limited Number of Measurements

This simulation was carried out to verify the performance of the proposed algorithm at limited number of baselines. We simulated data with only six baselines according to a real TanDEM-X images stack we have. The baseline ranges from -565.5 to 373.2 m. Fig. 13 compares the performance of the two algorithms at limited number of measurements. As one can see, in the noisy case, i.e., $\text{SNR} = 0$ dB, the

two algorithms have similar performance. However, with the increase of the SNR level, the proposed algorithm outperforms SLIMMER by a fair margin. To be specific, the proposed algorithm outperforms SLIMMER by about 20% effective detection rate in moderate super-resolving cases at 6 dB SNR. At high SNR level, i.e., 10 dB, the proposed algorithm provides 20%–30% higher effective detection rate even in extremely super-resolving cases, where the double scatterers are spaced closer than $0.5 \rho_s$.

F. Performance Verification Using Real Data

For a better evaluation of the proposed algorithm, we worked with a stack of six pairs of TanDEM-X high resolution staring spotlight images, with each pair acquired in pursuit monostatic mode, with a slant-range resolution of 0.6 m and an azimuth resolution of 0.25 m. The pairs were acquired between 2014 and 2015. Table III lists the detailed baselines. The elevation aperture size of about 940 m results in about 12 m inherent elevation resolution. Six interferograms were formed and registered to a single master. Since each interferogram was acquired in pursuit monostatic mode, the motion phase is negligible. An optical image of the test site from Google Earth and the SAR mean intensity image are showed in Fig. 14. The yellow arrow indicates the range direction. The atmospheric effects and deformation are ignored since the temporal baselines are negligible. Preprocessing, such as multiple SAR images coregistration and phase calibration were carried out using the German Aerospace Center (DLR's) integrated wide area processor (IWAP) [40]. Moreover, we manually selected a coherence point on the ground as reference and set its elevation as zero.

γ -Net employed for the real data experiment was trained with data simulated using the real baseline distribution. The training data contains four million samples generated with the same strategy mentioned in the simulation setup. After training, γ -Net can be directly applied in the upcoming TomoSAR processing on real data.

We use the proposed algorithm to reconstruct the whole test site and demonstrate the super-resolution power by comparing to the results derived by SLIMMER. The complete comparison of the reconstruction results of the test site is demonstrated in Figs. 15–17. Fig. 15 depicts only the reconstruction of single scatterer detected by both algorithms. Fig. 16 demonstrates the elevation estimates of detected single scatterers combined with the top layer of detected double scatterers. Comparing Figs. 15 and 16, we can see that both algorithms are able to detect dense double scatterers, which contribute to significant information increment and complete the structure of individual buildings shown in the test site. For a better view of the separation of overlaid scatterers, we demonstrate the top and bottom layer of detected double scatterers separately in Fig. 17. As we can see, the density of double scatterers detected by the proposed algorithm and SLIMMER is almost the same, meaning that both algorithms possess similar or even the same super-resolution power. However, it is worth mentioning that the proposed algorithm is more powerful to separate close layover. At the top the building, most reflections from roof and facade are overlaid. Comparing to

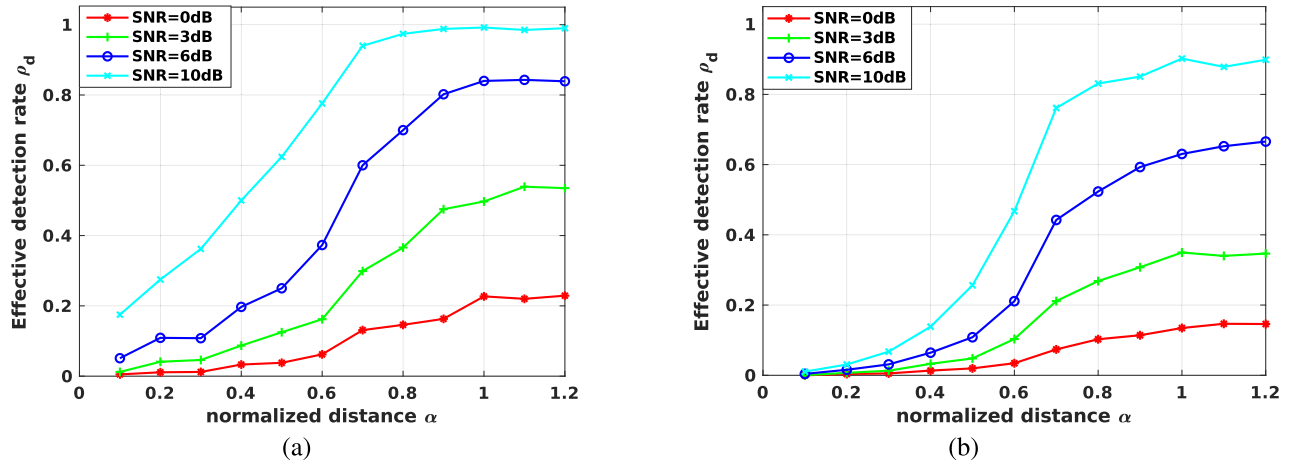


Fig. 13. Effective detection rate P_d as a function of the normalized elevation distance between double scatterers simulated with six real baselines. The simulated double scatterers are set to have identical phase and amplitude, i.e., the worst case. For each pair of (SNR, α), 0.2 million Monte Carlo trials were simulated. (a) Proposed algorithm. (b) SLIMMER.

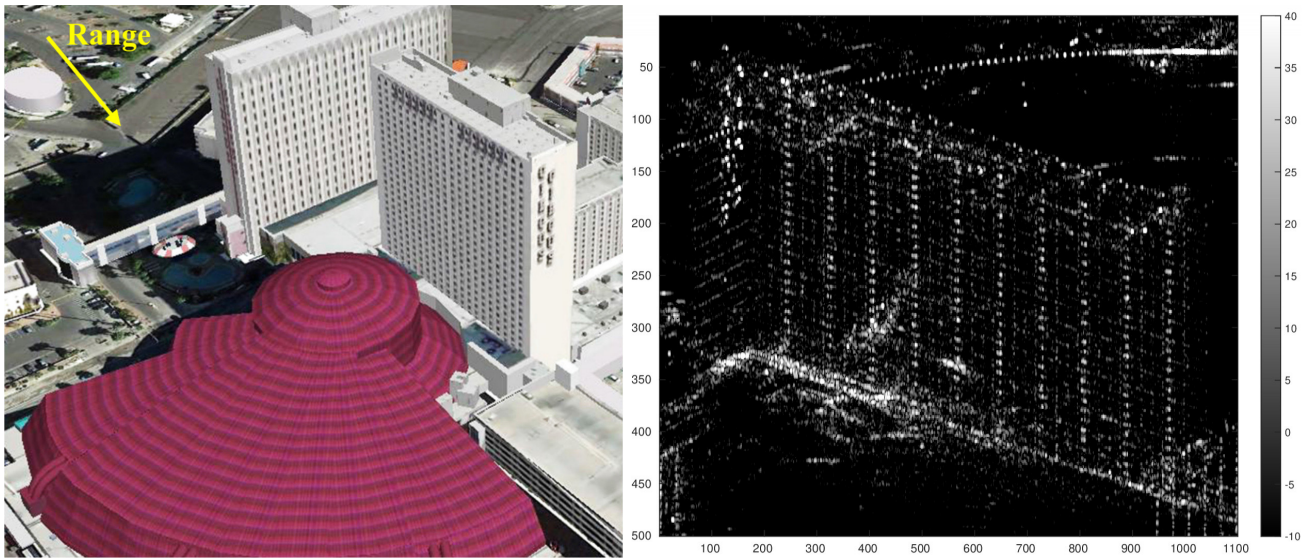


Fig. 14. Test site. (Left) Optical image from Google Earth. (Right) SAR mean intensity image.

TABLE III
DETAILED INTRODUCTION OF SPATIAL BASELINES OF THE HIGH-RESOLUTION TANDEM-X STACK

No.	1	2	3	4	5	6
Baseline [m]	-565.45	-311.43	-88.36	-7.69	82.43	373.21

SLIMMER, the proposed algorithm captures more reflection from the facade. This confirms the finding in Fig. 13 that the proposed algorithm outperforms SLIMMER at low number of measurements.

In terms of detection rate, the proposed algorithm is comparable to that of SLIMMER. Although there is no ground truth, we compare the agreement of the double scatterers detection of the both algorithms (shown in Table IV). For the whole test site, 38.97% and 37.76% of pixels are detected as double scatterers by the proposed algorithm and SLIMMER, respectively. 36.56% of the pixels were detected as double scatterers by both algorithms. Only 2.4% were only detected

TABLE IV
PERCENTAGE OF SCATTERERS DETECTION FOR THE TWO ALGORITHMS

Algorithm	Percentage of detection as		
	0 scatterer	1 scatterer	2 scatterers
proposed	30.71 %	30.32 %	38.97 %
SLIMMER	30.41 %	31.83 %	37.76 %

double scatterers by the proposed algorithm, and only 1.2% were only detected by SLIMMER.

However, as can be seen in Figs. 15–17, the reconstruction results derived from both algorithms seem slight noisy.

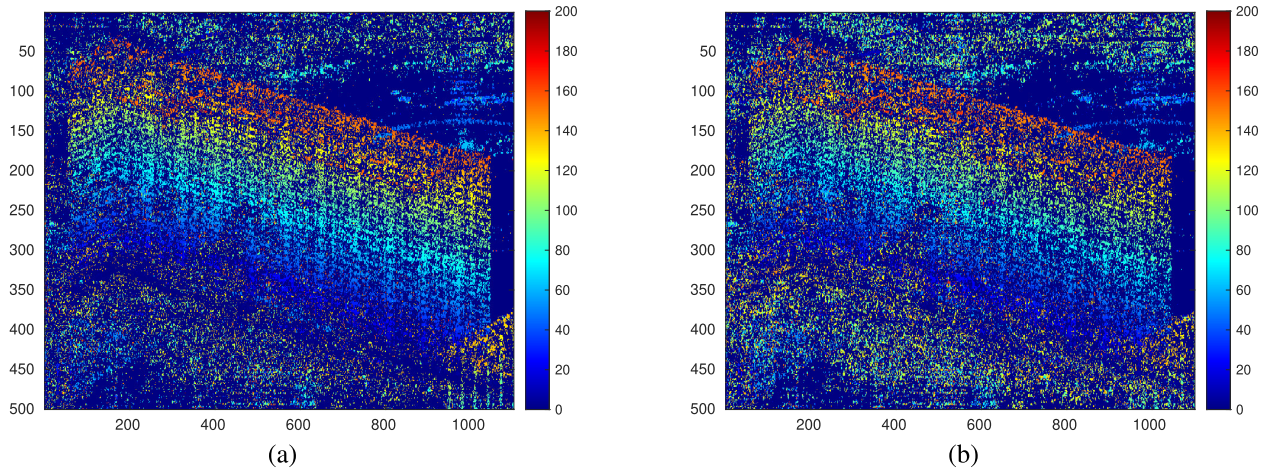


Fig. 15. Reconstructed and color-coded elevation of detected single scatterer. (a) Proposed algorithm. (b) SLIMMER algorithm.

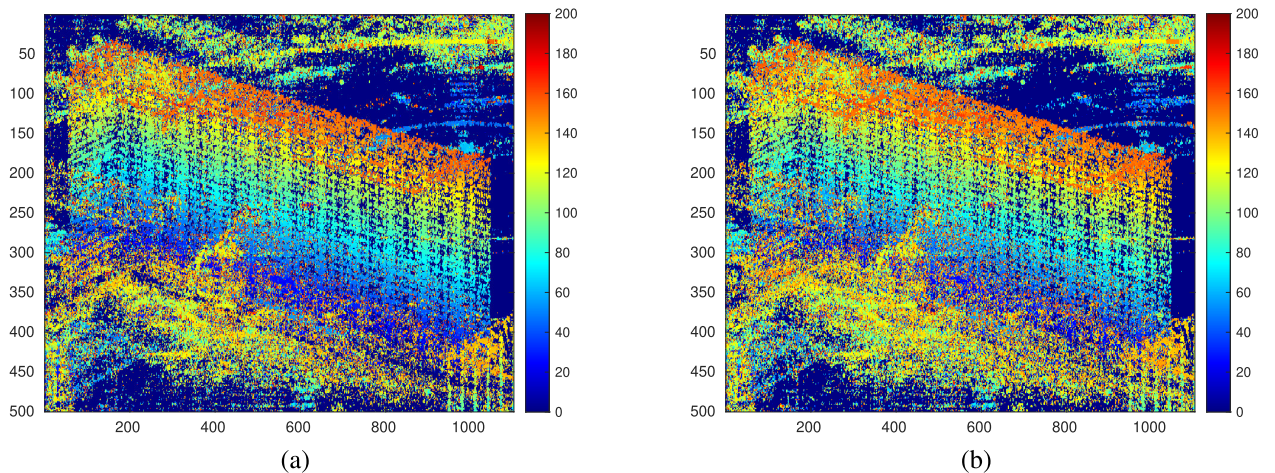


Fig. 16. Reconstructed and color-coded elevation of detected single scatterer + top layer of the detected double scatterers. (a) Proposed algorithm. (b) SLIMMER algorithm.

The noise in Figs. 15–17 attributes to the false alarm of single/double scatterers. As we have discussed in Section IV-E, the model order selection requires the observation’s noise variance. In simulations, the noise variance is known. Hence, we can clean almost all outliers and distinguish single and double scatterers correctly. However, in real-world processing, the noise variance needs to be estimated, thus the model order selection cannot clean outliers perfectly and will introduce noise caused by false alarm of single/double scatterers.

V. DISCUSSION

A. Analysis of Computational Complexity

We assume $\mathcal{O}(1)$ to be the computational complexity of one multiplication. The computational complexity of the proposed algorithm, as well as the original ISTA, is mainly determined by $\mathcal{O}(K_s L^2)$, where K_s is the number of layers or the number of iterations. For the proposed algorithm, K_s is set as 12. Comparing to the original ISTA, which usually requires hundreds or even thousands of iteration, the computational efficiency of the proposed algorithm is two to three orders of magnitude better. Moreover, other efficient L_1 norm

minimization solvers, such as fast iterative shrinkage thresholding algorithm (FISTA) [41], ADMM [25], and randomized blockwise proximal gradient (RBPG) [42], usually need about 100 iterations to converge and achieve reasonable estimation accuracy. Comparing to those efficient solvers, the proposed algorithm is still about one order of magnitude more efficient.

In our experiments, for a single dataset containing 0.2 million Monte Carlo trials simulated using the aforementioned setup, SLIMMER requires about ten CPU hours for the TomoSAR processing. On the contrary, it takes only a few CPU minutes when a trained γ -Net is employed, despite the fact that about 9 h are required for training the model with a single NVIDIA RTX 2080 GPU. However, the fixed cost of model training diminishes when we further increase the amount of the data. Fig. 18 provides us an intuitive view of the time consumption of the two methods for TomoSAR processing. As we can see, the training procedure dominates the time consumption of the proposed algorithm and the increment of the amount of data will not burden the time consumption seriously. In the contrast, the time consumption of SLIMMER escalates with the increasing amount of data, especially when limited measurements are available.

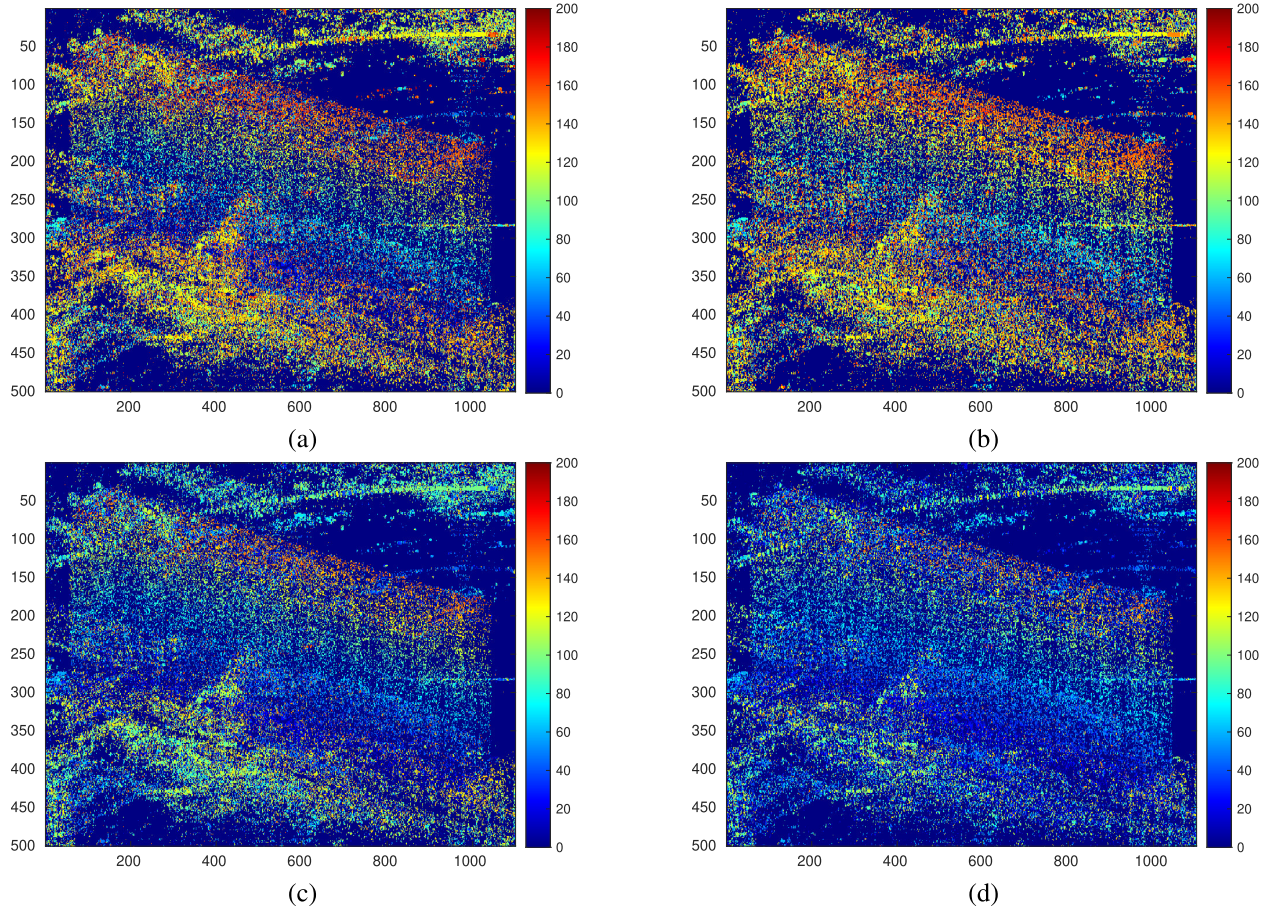


Fig. 17. Reconstructed and color-coded elevation of detected double scatterers. Top layer detected by the (a) proposed algorithm and (b) SLIMMER. Bottom layer detected by the (c) proposed algorithm and (d) SLIMMER.

In real-world TomoSAR processing, the number of pixels is usually tens or even hundreds of million, thus blocking the application of SLIMMER or other second-order CS-based methods. The proposed algorithm is able to complete the processing, including the training procedure, within matters of hours. The great superiority of the proposed algorithm in computational efficiency makes large-scale super-resolving TomoSAR processing feasible and realizable.

In addition, it is worth mentioning that the proposed algorithm maintains the elevation estimation accuracy in the meanwhile. The proposed algorithm employed the neural network with special structure, which can be trained as a more general model and is more likely to reach the global minimum and achieve better results. A detailed investigation about how deep learning improves the estimation efficiency for TomoSAR inversion will be executed in our following study.

B. Parameter Selection

1) *Step Size in γ -Net*: As it is stated in (7), a manual selected step size is required for the initialization of the trainable weights in γ -Net. To select the step size, the Lipschitz constant L_s is required, which is the largest eigenvalue of $\mathbf{R}^H \mathbf{R}$. Usually, a proper step size can be taken as $(1/L_s)$. In our experiments, we fix the step size as $(1/2L_s)$ to guarantee the convergence of γ -Net.

2) *Percentage ρ in the Support Selection*: An empirical formula is introduced in [33] to choose a proper percentage ρ^i for the i th layer of γ -Net for the support selection

$$\rho^i = \max(p \cdot i, p_{\max}) \quad (18)$$

where p is a positive constant and p_{\max} is the upper bound of the percentage of the support cardinality. Both p and p_{\max} can be selected using cross validation. From our experience, we fix the percentage ρ_i as 5% for all layers of γ -Net, which leads to satisfactory performance for our application.

C. Piecewise Linear Function

The learning architecture of γ -Net, where the output of the current layer is generated only directly from the output of the previous layer, leads to an error propagation phenomenon. Specifically, errors in the first few layers of γ -Net will be propagated and further amplified in the following layers. Moreover, the most serious problem is that once useful information is discarded in the previous layers, it is no longer possible for the upcoming layers to utilize the discarded information. The conventional soft-thresholding function simply prunes elements, whose magnitude is smaller than the threshold, to zero, which is very likely to discard information. The piecewise linear function is a smooth alternative of the soft-thresholding function. Instead of simply pruning elements with

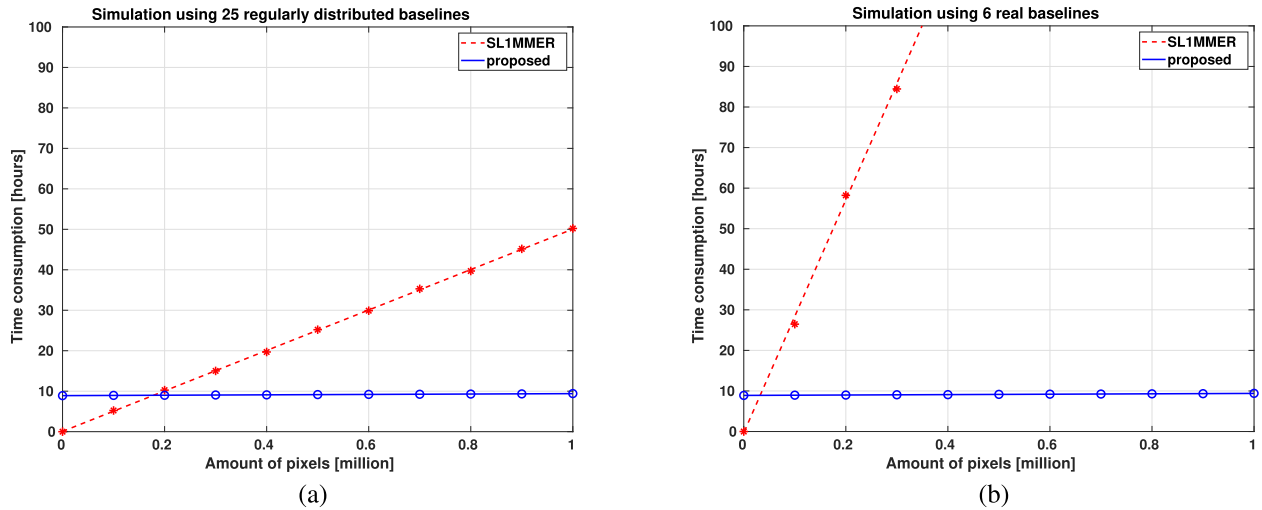


Fig. 18. Comparison of time consumption between the proposed algorithm and SL1MMER. (a) On dataset simulated using 25 regularly distributed baselines at 6 dB. (b) On dataset simulated using six real baselines at 6 dB. The training time will be affected only by the size of training data and the number of training epochs we set. Different baseline configuration does not affect the training time. On the contrary, the time consumption of SL1MMER is strongly dependent on the number of baselines. When limited images are available, the time consumption of SL1MMER escalates with the increasing number of data, whereas the inference time of the trained γ -Net is negligible. The proposed algorithm shows great computational efficiency in processing regular TomoSAR data, which usually contains tens of million pixels.

small magnitude, the piecewise linear function just further minimizes them to maintain information as much as possible and execute the shrinkage step in the meanwhile, thus moderating the information loss caused by the learning architecture and improving the performance.

D. Limitations of the Proposed Algorithm

1) *Training Time*: In the training procedure, there are $2 \times NLK + 5 \times K$ free trainable parameters. From our experience, about 2000 epochs are usually required for training the model, which takes about 9 h when a single NVIDIA RTX 2080 GPU is employed. Due to the inevitable time consumption of the training, the proposed algorithm is not recommended for processing small datasets because it consumes less time when conventional CS-based methods are used.

Similar to all other deep learning-based models, the training time could also become a burden when the task is to process many stacks with distinct baseline distributions. In our experiments, we simulated training data with the exact baselines as the test data. Ideally in real data processing, we shall train a separate model for each stack, which is very time consuming. However, our models show moderate tolerance to the baseline discrepancies between the training and the testing data. This is elaborated in more detail in Section V-D.2.

2) *Baseline Perturbation*: The biggest challenge to our deep learning model is the baseline discrepancies between the training and the testing data, because the baseline distribution is rather unique for each SAR interferometric stack. As a preliminary study, we test the proposed algorithm on testing data with slight baseline perturbation and find that the slight perturbation does not degrade the performance significantly. To be specific, we add random perturbation uniformly distributed in the range $[-10, 10]$ m, i.e., about 15% of the baseline standard deviation, to the 25 baselines. Applying

the pretrained γ -Net on the data with baselines perturbation shows that the effective detection rate decreases only 3% to 5% and the estimation accuracy as well as the bias retain nearly the same. This shows a good transferability of our trained model. However, further study is required to guarantee the performance of the proposed algorithm for large-scale processing.

3) *Application to More Complex Scenarios*: When the proposed algorithm is applied to more complex scenarios, i.e., more than two scatterers are overlaid in a single resolution unit, we are not capable of detecting and separating all of them. We tested the proposed algorithm in the three-scatterer case and found that the proposed algorithm tends to detect overlaid triple scatterers as double scatterers locating between the ground truth. Due to the fact that triple scatterers are not considered and covered in the training phase, the poor performance in coping with triple scatterers is explainable. From our perspective, the solution to this problem can be twofold. First, we can enrich the training data by introducing samples containing more scatterers. Second, we can view samples containing more than two scatterers as out-of-distribution samples, since in real-world processing only a tiny minority of pixels contain more than two scatterers. We can then use Dirichlet prior network (DPN) [43], [44] to detect these out-of-distribution data and solve it using CS solvers.

4) *False Alarm Caused by Model Order Selection*: The proposed algorithm employed γ -Net to obtain a preliminary sparse estimate of the reflectivity profile. After that, the model order selection step is followed to find the best model. However, the model order selection requires the observation's noise variance, which needs to be estimated in real data. As a result of inaccurate estimation of the noise variance, the model order selection will introduce noise caused by false alarm and the final reconstruction results will be affected accordingly.

VI. CONCLUSION

In this article, an advanced super-resolution TomoSAR inversion approach based on deep learning is proposed. We improved the complex-valued learned ISTA algorithm and proposed γ -Net by applying weight coupling structure, introducing support selection and employing the piecewise linear function instead of soft-thresholding. Experiments show that the proposed algorithm is capable of solving the L_2 - L_1 mixed norm minimization efficiently. Rigorous evaluation shows that the proposed approach is able to deliver competitive performance to the state-of-the-art in terms of the super-resolution capability and elevation estimation accuracy. This article opens a perspective on super-resolving TomoSAR inversion via deep learning and shows great potential of applying deep learning to solve other sparse reconstruction problems. In the future, we aim to extend the deep learning-based approach to higher dimensional spectral estimation problems, especially to differential TomoSAR reconstruction. Moreover, we will further exploit the power of deep learning to improve the performance, e.g., introducing long short-term memory (LSTM) unit to γ -Net to make use of historic information.

REFERENCES

- [1] 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.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] R. G. Baraniuk, "Compressive sensing," *IEEE signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Aug. 2007.
- [4] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [5] A. Budillon, A. Evangelista, and G. Schirinzì, "SAR tomography from sparse samples," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 4, Jul. 2009, pp. IV-865–IV-868.
- [6] X. X. Zhu and R. Bamler, "Very high resolution SAR tomography via compressive sensing," in *ESA Special Publication*, vol. 677, H. Lacoste, Ed. Frascati, Italy: Fringe, 2009, p. 29.
- [7] X. X. Zhu and R. Bamler, "Tomographic SAR inversion by L_1 -norm regularization—The compressive sensing approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, Oct. 2010.
- [8] X. X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247–258, Jan. 2012.
- [9] S. Zhang, Y. Zhu, G. Dong, and G. Kuang, "Truncated SVD-based compressive sensing for downward-looking three-dimensional SAR imaging with uniform/nonuniform linear array," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1853–1857, Sep. 2015.
- [10] X. Wang, F. Xu, and Y.-Q. Jin, "The iterative reweighted alternating direction method of multipliers for separating structural layovers in SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1883–1887, Nov. 2017.
- [11] Y. Wang, X. X. Zhu, and R. Bamler, "An efficient tomographic inversion approach for urban mapping using meter resolution SAR image stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1250–1254, Jul. 2014.
- [12] Y. Wang and X. Zhu, "SAR tomography via nonlinear blind scatterer separation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5751–5763, Jul. 2021.
- [13] G. Fornaro, S. Verde, D. Reale, and A. Pauciuolo, "CAESAR: An approach based on covariance matrix decomposition to improve multibaseline-multitemporal interferometric SAR processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2050–2065, Apr. 2015.
- [14] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2018.
- [15] X. 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.
- [16] Y. Wu, M. Rosca, and T. Lillicrap, "Deep compressed sensing," in *Proc. 36th Int. Conf. Mach. Learn.* (Proceedings of Machine Learning Research), vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds. Long Beach, CA, USA, Jun. 2019, pp. 6850–6860.
- [17] M. Wang *et al.*, "RMIST-net: Joint range migration and sparse reconstruction network for 3-D mmW imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.
- [18] M. Wang *et al.*, "CSR-net: A novel complex-valued network for fast and precise 3-D microwave sparse reconstruction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4476–4492, 2020.
- [19] R. Fu, Y. Liu, T. Huang, and Y. Eldar, "Structured LISTA for multi-dimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 69, pp. 3459–3472, 2021.
- [20] A. Budillon, A. C. Johnsy, G. Schirinzì, and S. Vitale, "SAR tomography based on deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 3625–3628.
- [21] J. Gao, Y. Ye, S. Li, Y. Qin, X. Gao, and X. Li, "Fast super-resolution 3D SAR imaging using an unfolded deep network," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP)*, Dec. 2019, pp. 1–5.
- [22] C. Wu, Z. Zhang, L. Chen, and W. Yu, "Super-resolution for MIMO array SAR 3-D imaging based on compressive sensing and deep neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3109–3124, 2020.
- [23] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [24] I. Daubechies, M. DeFrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.
- [25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jul. 2010, doi: [10.1561/22000000016](https://doi.org/10.1561/22000000016).
- [26] S. J. Wright, *Primal-Dual Interior-Point Methods*. Philadelphia, PA, USA: SIAM, 1997.
- [27] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.
- [28] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.* Madison, WI, USA: Omnipress, 2010, pp. 399–406.
- [29] Z. Wang, Q. Ling, and T. Huang, "Learning deep ℓ_0 encoders," in *Proc. 30th AAAI Conf. Artif. Intell.*, Jan. 2016, pp. 2194–2200.
- [30] Z. Wang, D. Liu, S. Chang, Q. Ling, Y. Yang, and T. S. Huang, "D3: Deep dual-domain based fast restoration of JPEG-compressed images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2764–2772.
- [31] Z. Wang, S. Chang, J. Zhou, M. Wang, and T. S. Huang, "Learning a task-specific deep architecture for clustering," in *Proc. SIAM Int. Conf. Data Mining*, 2016, pp. 369–377.
- [32] W. Zhangyang, Y. Yingzhen, C. Shiyu, L. Qing, and T. Huang, "Learning a deep ℓ_∞ encoder for hashing," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Jan. 2016, pp. 2174–2180.
- [33] X. Chen, J. Liu, Z. Wang, and W. Yin, "Theoretical linear convergence of unfolded ista and its practical weights and thresholds," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2018, pp. 9079–9089.
- [34] S. Osher, Y. Mao, B. Dong, and W. Yin, "Fast linearized Bregman iteration for compressed sensing and sparse denoising," *Commun. Math. Sci.*, vol. 8, no. 1, pp. 93–111, 2010.
- [35] X. X. Zhu, "Very high resolution tomographic SAR inversion for urban infrastructure monitoring—A sparse and nonlinear tour," Ph.D. dissertation, Fakultät für Bauingenieur- und Vermessungswesen, Technische Universität München, Munich, Germany, May 2011.
- [36] X. X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.
- [37] K. P. Burnham and D. R. Anderson, "Multimodel inference: Understanding AIC and BIC in model selection," *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, Nov. 2004.
- [38] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035.

- [39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [40] F. Rodriguez Gonzalez, N. Adam, A. Parizzi, and R. Brcic, "The integrated wide area processor (IWAP): A processor for wide area persistent scatterer interferometry," in *Proc. ESA Living Planet Symp.*, Edinburgh, U.K., 2013. [Online]. Available: <http://www.livingplanet2013.org/abstracts/850572.htm>
- [41] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
- [42] Y. Shi, X. X. Zhu, W. Yin, and R. Bamler, "A fast and accurate basis pursuit denoising algorithm with application to super-resolving tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6148–6158, Oct. 2018.
- [43] A. Malinin and M. Gales, "Predictive uncertainty estimation via prior networks," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2018, pp. 7047–7058.
- [44] J. Nandy. (2020). *Improving Dirichlet Prior Network for Out-of-Distribution Example Detection*. [Online]. Available: <https://openreview.net/forum?id=Bye4iaEFwr>



Kun Qian received the B.Sc. degree in remote sensing and information engineering from Wuhan University, Wuhan, China, and the B.Sc. degree in aerospace engineering and geodesy from the University of Stuttgart, Stuttgart, Germany, respectively, in 2016, and the M.Sc. degree in aerospace engineering and geodesy from the University of Stuttgart, in 2018. He is pursuing the Ph.D. degree with the Data Science in Earth Observation, Technical University of Munich, Munich, Germany.

His research focus includes data-driven methods, deep unfolding algorithms and their application in multibaseline synthetic aperture radar (SAR) tomography.



Yuanyuan Wang (Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, in 2008, and the M.Sc. and Dr.-Ing. degrees from the Technical University of Munich, Munich, Germany, in 2010 and 2015, respectively.

From June to July 2014, he was a Guest Scientist with the Institute of Visual Computing, ETH Zürich, Zürich, Switzerland. He is currently with the Department of EO Data Science, Remote Sensing Technology Institute, German Aerospace Center, Weßling, Germany, where he leads the working Group Big Synthetic Aperture Radar (SAR) Data. His research interests include optimal and robust parameters estimation in multibaseline InSAR techniques, multisensor fusion algorithms of SAR and optical data, nonlinear optimization with complex numbers, machine learning in SAR, and high-performance computing for big data.

Dr. Wang was one of the best reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2016.



Yilei Shi (Member, IEEE) received the Dipl.-Ing. degree in mechanical engineering and the Dr.-Ing. degree in signal processing from the Technische Universität München (TUM), Munich, Germany, in 2010 and 2019, respectively.

He is a Senior Scientist with the Chair of Remote Sensing Technology, TUM. His research interests include fast solver and parallel computing for large-scale problems, high-performance computing and computational intelligence, advanced methods on synthetic aperture radar (SAR) and InSAR processing, machine learning, and deep learning for variety of data sources, such as SAR, optical images, and medical images, and partial differential equation (PDE)-related numerical modeling and computing.



Xiao Xiang Zhu (Fellow, IEEE) received the M.Sc., Dr.-Ing., and Habilitation degrees in the field of signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is currently a Professor for Data Science in Earth Observation (former: Signal Processing in Earth Observation) with TUM, and the Head of the Department "EO Data Science," Remote Sensing Technology Institute, German Aerospace Center (DLR), Weßling, Germany. Since 2019, she has been a Co-Coordinator of the Munich Data Science Research School, Munich, (www.mu-ds.de). Since 2019, she has been heads the Helmholtz Artificial Intelligence–Research Field "Aeronautics, Space and Transport." Since May 2020, she has been the Director of the International Future AI Laboratory "Artificial Intelligence for Earth Observation (AI4EO): Reasoning, Uncertainties, Ethics and Beyond," Munich. Since October 2020, she has been serves as a Co-Director of the Munich Data Science Institute (MDSI), TUM. She was a Guest Scientist or a Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, The University of Tokyo, Tokyo, Japan, and University of California, Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is currently a Visiting AI Professor with the ESA's Phi-Laboratory. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with their applications in tackling societal grand challenges, e.g., global urbanization, United Nations (UN) sustainable development goals (SDGs), and climate change.

Dr. Zhu is a member of Young Academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She serves in the scientific advisory board in several research organizations, among others the German Research Center for Geosciences (GFZ) and Potsdam Institute for Climate Impact Research (PIK). She is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and serves as the area editor responsible for special issues of the *IEEE Signal Processing Magazine*.