ATASI-Net: An Efficient Sparse Reconstruction Network for Tomographic SAR Imaging With Adaptive Threshold

Muhan Wang®, Zhe Zhang®, Member, IEEE, Xiaolan Qiu®, Senior Member, IEEE, Silin Gao®, and Yue Wang®, Senior Member, IEEE

Abstract— Tomographic synthetic aperture radar (SAR) technique has attracted remarkable interest for its ability of 3-D resolving along the elevation direction via a stack of SAR images collected from different cross-track angles. The emerged compressed sensing (CS)-based algorithms have been introduced into SAR tomography (TomoSAR) considering its super-resolution ability with limited samples. However, the conventional CS-based methods suffer from several drawbacks, including weak noise resistance, high computational complexity, and complex parameter fine-tuning. Aiming at efficient TomoSAR imaging, this article proposes a novel and efficient sparse unfolding network based on the analytic learned iterative shrinkage-thresholding algorithm (ALISTA) architecture with adaptive threshold, named adaptive threshold ALISTA-based sparse imaging network (ATASI-Net). The weight matrix in each layer of ATASI-Net is precalculated as the solution of an off-line optimization problem, leaving only two scalar parameters to be learned from data, which significantly simplifies the training stage. Furthermore, the introduction of an adaptive threshold for each azimuth–range pixel permits the threshold shrinkage to be not only layer-varied but also elementwise. In addition, the final learned thresholds can be visualized and combined with the SAR image semantics for mutual feedback. Finally, extensive experiments on simulated and real data are carried out to demonstrate the effectiveness and efficiency of the proposed method.

Index Terms— Adaptive threshold ALISTA-based sparse imaging network (ATASI-Net), analytic learned iterative shrinkage-thresholding algorithm (ALISTA), compressed sensing (CS), deep unfolded network, semantic, sparse reconstruction, synthetic aperture radar tomography (TomoSAR).

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) [1] uses the synthetic aperture principle to achieve high resolution in both the azimuth and range directions. However, traditional 2-D SAR images do not have resolving ability along the elevation direction, causing issues such as layover and shadowing. SAR tomography (TomoSAR) [2], as an emerging radar imaging technique, has attracted remarkable interest in recent years for its ability in achieving 3-D resolution by collecting data of the same target scene from multiple slightly different viewing, by either a physical antenna array or multiple flight passes. TomoSAR technique has been widely used in urban observation [3], forestry detection [4], and disaster monitoring [5].

Along the elevation dimension, TomoSAR imaging problem can be treated as a line spectrum estimation problem in theory. Traditionally, it can be solved via canonical spectrum estimation algorithms, such as singular value decomposition (SVD) [6], multiple signal classification (MUSIC) [7], or CAPON [8], which, however, usually experience poor performance under limited observations and low signal-to-noise ratio (SNR) circumstances. In typical scenarios of TomoSAR applications, scatterers are usually distributed sparsely along the elevation direction, and meanwhile, only a few significant scatterers fall into a range–azimuth pixel. Thus, compressed sensing (CS)-based [9] methods are widely used to solve the TomoSAR inversion problem as the state-of-the-art approach. A myriad of techniques has been proposed to tackle such sparse inversion problems, which conventionally contain two categories. The first group is the greedy approach, adopting the minimization, such as orthogonal matching pursuit (OMP) [10] and compressive sampling matching pursuit (CoSaMP) [11]. The second class of methods, known as convex optimization, adopts the regularization and forms a convex object function, such as iterative shrinkage-thresholding algorithm (ISTA) [12], approximate message passing (AMP)
algorithm [13], and alternating direction method of multipliers (ADMM) [14]. In recent years, different CS-based methods for solving TomoSAR inversion have been extensively studied. Budillon et al. [15] presented the first simulation of CS TomoSAR. Zhu and Bamler [16] conducted the first real data tomographic SAR inversion by the CS approach and developed the scale-down by $L_1$-norm minimization model selection estimation reconstruction (SIIMMER) method [17]. He et al. [18] applied the generalized OMP algorithm to the TomoSAR imaging. Han et al. [19] modified the ISTA and proposed an improved method for the 3-D reconstruction of airborne TomoSAR. Although CS-based algorithms are widely accepted and believed to be reliable, they still suffer from several drawbacks, including weak noise resistance, high computational complexity, the holding conditions of sparse assumption, and fine-tuning of reconstruction parameters.

With the development of artificial intelligence and deep learning technique, the data-driven deep network provides a new idea to overcome the limitation of complicated parameter fine-tuning issues in CS and improve its performance in many aspects. Due to the enhanced interpretability, fast implementation, and high robustness to model mismatch, a deep learning method called “deep unfolding” [20] was proposed to provide a concrete and systematic connection between iterative model-based algorithms and deep neural networks. Specifically, iteration-based sparse reconstruction algorithms of CS can be represented by an unfolded deep recurrent network. The reconstruction parameters, e.g., regularization parameter, can be learned from a given dataset via canonical training methods over a given dataset, and the sparse reconstruction becomes a simple and fast inference process. Along this line, a novel sparse unfolding network, called learned ISTA (LISTA), has been first proposed as the unfolded network of the ISTA by Gregor and LeCun [21]. In LISTA, ISTA is folded into a multilayer deep network. Following the unrolling, training data can be fed through the network, and stochastic gradient descent can be used to update and optimize its parameters, including the regularization parameter, step size, and even the measurement matrix. Borgerding et al. [22] proposed learned AMP (LAMP) by unfolding AMP to deep network. Unfolding was also applied to the ADMM algorithm to solve the magnetic resonance imaging (MRI) problem [23]. The resulting network, ADMM-CSNet, uses training data to learn filters, penalties, simple nonlinearities, and multipliers. In addition, some improved networks based on the above unfolded networks have also been proposed, including LISTA-CPS [24] and LDAMP [25]. Moving beyond natural images and medical images, the deep unfolded network has been applied successfully in remote sensing [26], [27], speech and video processing [28], [29], and so on. In these applications, unfolding and training significantly improve both the quality and speed of signal reconstruction.

Recently, the idea of deep unfolding has also triggered the attention in the applications of SAR imaging. Pu [30], [31] proposed a deep SAR imaging algorithm using the unfolded ADMM-based autoencoder structure to deal with the SAR motion compensation and SAR autofocus and eliminate the influence of motion errors, improving the quality of SAR imaging. Aiming at enhancing the desired target and improving the SCR in the reconstructed SAR images, Li et al. [32] unfolded an iterative ADMM solver based on matched filter methods and proposed MF-ADMM-Net. Also, based on the ADMM architecture, Li et al. [33] later combined the sparsity-cognizant total least-square model and proposed the STLS-LADMM-Net for improving the quality of SAR autofocus imaging. In addition, the UESTC team has extended the application of the deep unfolding framework to 3-D millimeter-wave (mmWave) SAR imaging systems. Wang et al. [34] first proposed a two-path iterative framework dubbed TPSSI-Net based on the AMP algorithm architecture. Rather than manually choosing a sparse dictionary, a two-path convolutional neural network is developed and embedded in TPSSI-Net for nonlinear sparse representation in the complex-valued domain. Results in [34] show that TPSSI-Net is capable of yielding favorable 3-D reconstruction performance compared with traditional methods. For the same purpose of improving the performance of 3-D mmWave image reconstruction, Wei et al. [35] proposed SISR-Net, Zhou et al. [36] proposed SAF-3DNet, and Wang et al. [37] proposed LFIST-Net. These deep 3-D SAR imaging methods validate the superiority of deep unfolded networks through simulated and real data, and however, they are all targeting the mmWave SAR under nonsparse scenarios.

Encouraged by deep unfolding, the TomoSAR community also started to design a deep unfolded framework based on CS iterative optimization algorithms in urban areas with sparse scenes. Gao et al. [38] unfolded and mapped vector AMP [39] into a deep network framework for line spectral estimation and applied it to tackle TomoSAR inversion, and Fan et al. [40] quantified the optimal network design and maximum super-resolution (SR) ability of the network. Simulation results validate the superiority of the proposed deep unfolded network based on the backbone of VAMP. Qian et al. [41] developed a sparse unfolding network named $\gamma$-Net to solve the TomoSAR problem by improving the LISTA-CPS. Also, the evaluation shows that the proposed network is able to deliver competitive performance to the state of the art in terms of the SR capability and elevation estimation accuracy. However, although these methods can replace the traditional methods by designing the network with the idea of deep unfolding and learning the parameters, the number of parameters that are needed to be learned is usually very huge. This will lead to huge training computational complexity and enormous dataset size requirement, which is usually unacceptable in SAR applications.

In this article, we propose a novel and efficient deep unfolded network named adaptive threshold ALISTA-based sparse imaging network (ATASI-Net). Based on the success of analytic LISTA (ALISTA) architecture [42], we are able to reduce the number of required training samples and increase the processing speed and improve the reconstruction accuracy. In concise, each layer of ATASI-Net consists of one module and three units in cascade, including precalculation module, error propagation unit, threshold unit, and reconstruction unit. Exploiting the ALISTA, in the precalculation module
The basic principle of TomoSAR is shown in Fig. 1. The elevation direction and achieve the 3-D resolving ability. Observation area to reconstruct the scattering information along images collected from different cross-track angles of the same existing approaches. Section III proposes the details of the structured approach is adopted to output the 3-D point clouds. The main contributions of this article are given as follows. 1) An efficient sparse unfolding network named ATASI-Net is proposed for TomoSAR 3-D imaging based on the ISTA algorithm, in which the L1 regularization is replaced with the log-sum penalty function, enabling the threshold shrinkage to be not only layer-varied but also elementwise. In addition, the final learned thresholds are visualized in the experiments on the measured data, and the threshold segmentation can be combined with the SAR image semantics for mutual feedback. 2) The weight matrix in ATASI-Net is computed via an off-line convex optimization problem based on the ALISTA architecture, leaving only two scalar parameters to data-driven learning, which avoids the parameter fine-tuning stage and reduces temporal and spatial complexity. For SAR scenarios with large data volumes but few training samples, this greatly reduces the training dataset size and time consumption when applied to TomoSAR 3-D imaging. 3) To train the proposed network, we employ simulated data created using preexisting scene geometry parameters. Furthermore, we introduce a novel training approach for real scenes with unknown geometric parameters. This involves using better-focused data obtained from the results as the training set, leading to enhanced conventional reconstruction outcomes. 4) Extensive experiments are carried out on both simulated and real data to demonstrate the superiority of the proposed ATASI-Net. In addition, in our experiments with actual data, we employ the structured 3-D modeling to evaluate the point cloud results. The remainder of this article is organized as follows. Section II introduces the TomoSAR imaging model and existing approaches. Section III proposes the details of the ATASI-Net deep network, as well as its training strategies and discussions. The simulation and real data experimental results are demonstrated in Sections IV and V, respectively. Finally, the conclusion of this article is given in Section VI.

II. PROBLEM SETTING AND UNFOLDED NETWORK OF ISTA

A. TomoSAR Imaging Model

The tomographic SAR technique uses a stack of SAR images collected from different cross-track angles of the same observation area to reconstruct the scattering information along the elevation direction and achieve the 3-D resolving ability. The basic principle of TomoSAR is shown in Fig. 1.

For a single look complex (SLC) SAR image, \( y(x_0, r_0) \) represents the value of an azimuth–range pixel \((x_0, r_0)\). Consider that the same target is observed \(N\) times from slightly different viewing angles to obtain \(N\) SLCs (via an antenna array or multiple flight passes). If the SLCs are perfectly aligned, denote the data acquisition as \( y_n \). Each azimuth–range pixel \((x_0, r_0)\) can be expressed as the integration of the scattering coefficient along the elevation weighted by sinusoids as

\[
y_n(x_0, r_0) = \int y(x_0, r_0, s) \exp\left(\frac{4\pi}{\lambda} \cdot \frac{s b_n}{r_0}\right) ds \tag{1}
\]

where \( y(s) \) represents the scattering coefficient distribution, also known as reflectivity function along elevation \( s \), \( b_n \) is the baseline length of the \(n\)th observation, and \( \lambda \) is the wavelength of the transmitted signal. Assuming that the scatters are sparsely distributed along the elevation direction, we can discretize the continuous function \( y(\cdot) \) into a sparse vector. Along with the additive noise, (1) can be written in a matrix–vector form as follows:

\[
y = R\mathbf{y} + \mathbf{\varepsilon} \tag{2}
\]

where \( y \) is the measurement vector stacked by \( y_n \) and \( R \) plays as an \( N \times M \) mapping matrix with

\[
R_{at} = \exp\left(j4\pi b_n s_l/\lambda r_0\right) \tag{3}
\]

where \( M \) represents the length of the signal matrix \( y \).

Now, TomoSAR inversion boils down to a signal recovery problem, where the goal is to obtain the corresponding scattering parameters such as elevation and reflectivity profile \( \mathbf{y} \) of each range–azimuth cell by solving (2), where \( R \) is usually known from model and \( y \) is the observation.

B. Iterative Shrinkage-Thresholding Algorithm

It is illustrated in [16] that the number of layered scatters, or sparsity, is less than 4 in each range–azimuth resolution cell for TomoSAR in the vast majority of urban areas. Hence, the echo signal along the elevation direction is sufficiently sparse and can be solved via a sparse reconstruction problem (2) using the CS technique. Within the CS framework, \( y \) can be reconstructed by the \( L_0 \)-norm minimization

\[
\min_{\mathbf{y}} ||\mathbf{y}||_0 \quad \text{s.t.} \quad \mathbf{y} = R\mathbf{y} \tag{4}
\]
and in the presence of noise, it can be approximated by
\[ \hat{y} = \arg\min_{y} \frac{1}{2} \|y - R\gamma\|_2^2 + \lambda \|\gamma\|_0. \] (5)

As a convex relaxation, the \( L_0 \)-norm can be replaced by the \( L_1 \) regularizer [43], and \( \gamma \) can be reconstructed in the form of least absolute shrinkage and selection operator (LASSO) problem as
\[ \hat{y} = \arg\min_{y} \frac{1}{2} \|y - R\gamma\|_2^2 + \lambda \|\gamma\|_1 \] (6)

where \( \lambda \) is a regularization parameter that determined by the sparsity of \( \gamma \).

Iterative shrinkage-thresholding algorithm (ISTA) is a widely used sparse reconstruction algorithm. Compared with greedy algorithms such as OMP, ISTA has better robustness against noise and reconstruction performance but is relatively time-consuming. In ISTA, the estimate of \( \gamma \) is achieved in an iterative manner as
\[ \hat{y}_{k+1} = h(\alpha L)\left( y_k + \frac{1}{L} R^T (y - R \hat{y}_k) \right) \] (7)
where \( L \) is a parameter controlling the iteration step size satisfying \( L > \lambda_{\text{max}}(R^T R) \), \( \lambda_{\text{max}}(\cdot) \) denotes the maximum eigenvalue, \( \theta = \alpha/L \) is the threshold parameter, and \( h_\theta(X) \) is the soft-thresholding function defined as
\[ h_\theta(X) = \text{sign}(X) \max(|X| - \theta, 0). \] (8)

However, the parameters \( \alpha \) and \( L \) in the ISTA algorithm are manually chosen, which usually requires a time-consuming fine-tuning process to achieve the best performance. Furthermore, these parameters are not adaptive i.e., they are fixed from one scene to another unless we fine-tune them repeatedly. To overcome the drawbacks of traditional ISTA, Gregor and LeCun [21] unfolded the ISTA into a deep network to leverage the benefits of deep learning, namely, LISTA.

C. Unfolded Deep Networks Based on ISTA

In LISTA, one iteration in (7) can be rewritten in a way of neuron’s activities as one layer of neural network
\[ \hat{y}_{k+1} = h_{\theta_k}(W_1^{k+1} y + W_2^{k+1} \hat{y}_k) \] (9)
where \( W_1^{k+1} = (1/L)R^T \) and \( W_2^{k+1} = I - (1/L)R^T R \) are two sets of parameters that can be trained. Total \( K \) iterations construct a \( K \)-layer neural network. It is a sparse unfolded network that takes advantage of the representation power of deep learning and uses available data to train the parameters, matrix \( W_1^k \) and \( W_2^k \), and threshold \( \theta_k \). Compared with ISTA, LISTA converges faster and produces a better solution as [24] demonstrated.

Although LISTA allows us to learn parameters from data autonomously, the sizes of the weight matrices \( W_1 \) and \( W_2 \) are usually huge in real SAR signal processing applications. Hence, a vast dataset is required to train the enormous amount of parameters, which is usually impractical, and the training process is temporally and spatially complex.

D. Analytic LISTA

Tackling such challenges mentioned above, we propose to introduce the ALISTA [42] to the TomoSAR problem. ALISTA is an improved version of LISTA where both weight matrices are determined off-line via a convex optimization problem, leaving only the step size and threshold parameters to be learned. It significantly simplifies the training stage in both spatial and temporal domains. It has been shown that ALISTA retains the benefits of LISTA in terms of optimal linear convergence rate and achieves a performance comparable to LISTA.

Instead of training \( W_1 \) and \( W_2 \), weight matrix \( W \) in ALISTA is predetermined by solving the following convex optimization problem:
\[ W = \arg\min_{W \in C^{M \times N}} \|W^T R\|_F^2 \text{s.t.} \langle W, i \rangle^T R, i = 1, \ldots, n \] (10)
which plays as a mutual coherence minimizer between \( W \) and \( R \). This is motivated by the tenet in CS that a dictionary with smaller coherence possesses better sparse recovery performance. Also, the equivalence has been proven in [42].

Then, let \( W^k = \mu_k W \), and (9) can be rewritten as follows:
\[ \hat{y}_{k+1} = h_{\theta_k}(\hat{y}_k + \mu_k W^T (y - R \hat{y}_k)) \] (11)
where only two scalar parameters \( \{\theta_k, \mu_k\} \in R \) are learned from end-to-end data. In particular, the number of training parameters is reduced from \( O(KM^2 + K + MN) \) in LISTA down to \( O(K) \) in ALISTA, causing a marked reduction in training burden and required training dataset size.

III. PROPOSED ATASI-NET FOR TomoSAR

A. Adaptive Threshold

ALISTA still uses the traditional threshold shrinkage function where the regularization parameter is fixed in a specific layer. In practice, the scattering intensity of different target structures of SAR images varies greatly, and taking the same threshold value for the whole SAR image will result in chopping off the information, which is inappropriate. In this work, we consider a threshold adaptation approach for sparse recovery by introducing a concave regularizer to promote sparsity. Based on this, we introduce the learnability of deep networks. In the actual data experiment, we segmented and visualized the elementwise threshold learned through the network and verified that it can feed mutually with SAR image semantic information, which validates the rationality of the proposed adaptive thresholds.

We formulate a sparse-promoting problem using concave regularizer \( G(x) \) to approximate the substitution of \( L_0 \) norm in (5)
\[ \gamma = \arg\min_{\gamma} \frac{1}{2} \|y - A\gamma\|_2^2 + \lambda \sum_{i=1}^n G(|\gamma_i|) \] (12)
where matrix \( A \) represents the sensing matrix, which equals \( R \) in (3).

Kim and Park [44], Fan and Li [45], and Zou and Li [46] also used such concave regularizers for sparse recovery problems. According to [46], when the derivative of \( G(x) \),
where \( \lambda \) is equivalent to ISTA with adaptive thresholds

\[
\theta_i^{k+1} = \frac{\zeta}{|y_i^k| + \tau} \\
\gamma_i^{k+1} = h_{\theta_i}(y_i^k - \delta A_i^T (A y^k - y))
\]

(17)

where \( \zeta = \lambda \tau \). The parameters \( \delta \) and \( \lambda \) in (17) need to be adjusted manually.

2) LISTA With Adaptive Threshold: For maximizing efficiency and better recovering results, we introduce learnable parameters to enable automatic hyperparameter tuning as shown in the following equation:

\[
\theta_i^{k+1} = \frac{s^{k+1}}{|y_i^k| + \tau} \\
\gamma_i^{k+1} = h_{\theta_i}(y_i^k - \delta A_i^T (A y^k - y))
\]

(18)

where \( \Theta = \{W_1^k, W_2^k, \zeta^k\} \) are learnable parameters. Their initial values are \( W_{1,i}^{k+1} = 0 \) and \( W_{2,i}^{k+1} = 1 \). Also, (18) is equivalent to LISTA with adaptive elementwise threshold. Following the same derivation from LISTA to ALISTA, according to [42], we can also apply our elementwise adaptive thresholding method to ALISTA.

B. ATASI-Net

The proposed ATASI-Net, as an enhanced version of unfolded ISTA-based architecture, consists of several well-designed update layers in a cascaded form to pursue a performance boost. Fig. 3 shows the overall architecture of ATASI-Net and illustrates the data flow of each module in the \( k \)th layer. As shown, each updating layer consists of one module and illustrates the data flow of each module in the

\[
\begin{align*}
&\text{Fig. 2. } \text{Canonical } L_0 \text{ sparsity count is better approximated by the log-sum penalty function than by the traditional convex } L_1 \text{ relaxation.}
\end{align*}
\]
2) Error Propagation Unit (D): This unit estimates the residual measurement error. As for the output, given the estimated reflection profile of the previous iteration, $D^k$ can be calculated by

$$D^k = A\gamma^k - y.$$  

(21)

3) Threshold Unit (T): Different from ALISTA, the threshold of the proposed ATASI-Net is not only layer-varied but also elementwise, which improved retention against weak targets and utilized the semantic information. In particular, we introduce trainable parameter $\varsigma^k$ to adjust the step size of the thresholds between adjacent layers. Given the reconstruction signal $z_{i}^{k}$ for the $i$th element, the elementwise shrinkage-thresholding operator $\theta^k_{i}$ can be calculated by (22), and according to [39], $\tau$ is set to 0.01 to achieve an optimal performance

$$\theta^k_{i} = \varsigma^k \frac{1}{|z_{i}^{k}| + \tau}.$$  

(22)

The threshold unit $T^k$ computes the shrinkage threshold $\theta^k$ of the block based on (22).

4) Reconstruction Unit (Z): The roughly estimated signal can be calculated by a form of one-step gradient descent as

$$z^{k+1} = y^k - W^k D^k = y^k - \mu^k W_{i}^k D^k.$$  

(23)

Then, the reconstruction unit $Z^k$ obtains the reconstructed reflection profile $\hat{y}^{k+1}$ by means of an activation layer defined based on an adaptive threshold shrinkage function, which can be calculated by

$$\hat{y}^{k+1} = h_{\theta^{k+1}}(z^{k+1}).$$  

(24)

For elementwise adaptive thresholds, the specific value of reflection profile $\gamma^k_{i}$ in the $i$th azimuth–range cell of the $k$th layer can be expressed as $\gamma^k_{i} = h_{\varsigma^k_{i}}(z^k_{i}) = \text{sign}(z^k_{i}) \max(|z^k_{i}| - \theta^k_{i}, 0)$.

5) Training Strategy and Backpropagation: Similar to the traditional deep networks, ATASI-Net can be trained from end-to-end via backpropagation in a data-driven manner.

a) Loss function: To pursue a high reconstruction quality, mean squared error (mse) is used as the cost function to obtain optimal hyperparameters. Given dataset used for training $\gamma_{\text{label}}^{k}, y^{k}$ and learnable parameter set $\Theta_{\text{para}} = \{\mu_t, \varsigma_t\}_{t=1}^{M}$, the cost function can be defined as

$$L(\Theta) = \frac{1}{N_{\text{train}}} \sum_{\Omega_{\text{max}}} \| \hat{y}(\Theta, y) - y_{\text{label}} \|_2$$  

(25)

where $N_{\text{train}}$ represents the number of trained samples and $\hat{y}(\Theta, y)$ denotes the network output based on the echo signal $y$ and network parameters $\Theta$.

b) Backpropagation and gradient calculation: Similar to typical deep networks, the parameters of ATASI-Net can be optimized via backpropagation with a variety of optimizers, such as stochastic gradient descent (SGD) [52], Adam [53], and Adadelta [54].

In the training process of the model, the observation signal $y$, the measurement matrix $W$, and the reflection profile $\gamma$ are all complex-valued. Therefore, the proposed ATASI-Net is a complex-valued network, even if the trained parameters are real. Inspired by the work in [33], the gradients of $L(\Theta)$ for the parameters $\Theta$ of the $k$th layer can be computed via the formulas as follows:

$$\frac{\partial L}{\partial \Theta_k} = \frac{\partial L}{\partial \Theta_k} \frac{\partial \Theta_k}{\partial \Theta_k}.$$  

(26)

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Also, for the partial derivation of a complex-valued matrix, $O$ can be calculated as follows [55]:

$$\frac{\partial L}{\partial O} = \frac{\partial L}{\partial \text{Re}[O]} + j \frac{\partial L}{\partial \text{Im}[O]}.$$  \hspace{1cm} (27)

It should be pointed out that, in widely used deep learning toolboxes such as Pytorch and TensorFlow, complex numbers are already supported, making it convenient for implementing complex number networks. In this article, the proposed network ATASI-Net is implemented by Tensorflow.

### Algorithm 1 Summary of the Proposed ATASI-Net

1) Obtain training data:
   - Generate reflectivity profile $\gamma$
   - Generate measurement matrix $A$ via (3)
   - Acquire simulation signal $y$ via (2);
   - Finish: The generation of training data $\{(y_i, \gamma_i)_{i=1}^{M \times N}\}$

2) Training of ATASI-Net

   Precomputing $W$ by solving
   $$W = \arg \min_{W \in \mathbb{C}^{M \times N}} \|Wy - Ay\|^2_F \text{ s.t.} (Wz, A, i, i) = 1, \ldots, n)$$
   Over given training samples $\{(y_i, \gamma_i)_{i=1}^{M \times N}\}$
   $$\min L(\Theta) = \frac{1}{N_{\text{train}}} \sum_{i=\text{train}} \|\hat{y}(\Theta, y) - y_{\text{label}}\|^2_2$$
   where $\Theta = \{\mu_k, \varsigma_k\}_{k=1}^K$
   Obtain the optimal parameters $\Theta = \{\mu_k, \varsigma_k\}_{k=1}^K$

3) ATASI-Net for TomoSAR

   Input: observed signal $y$, measurement matrix $A$
   Output: reflectivity profile $\gamma$, threshold $\theta$
   1: Load the optimal parameter set $\Theta = \{\mu_k, \varsigma_k\}_{k=1}^K$, $K$ is the maximum number of layers;
   2: Precompute $W$ via (19)
   3: Set $\tau = 0.005 \times Y_{\text{max}}, \mu^0 = 0.01$; Initialize $\gamma^0, D^0$
   4: while $k \leq K$ do
   5: $W^k = \mu^k W$
   6: $z^k = y^{k-1} - W^{k-1} D^{k-1}$
   7: $D^k = A \gamma^{k-1} - y$
   8: $\theta_k^l = \varsigma^l \frac{1}{|\gamma|^2 + \tau}$
   9: $\gamma^{k+1} = h_{\theta_k^l}(z^{k+1})$
   10: $k = k + 1$
   11: end while

### C. Discussion

The proposed ATASI-Net takes advantage of both the theoretical interpretability of conventional model-driven algorithms and the learnability of deep networks. A modified version of unfolded ISTA is taken as the backbone, which aims to pursue a performance boost. The characteristics of ATASI-Net are discussed as follows.

Compared to the traditional CS algorithm such as OMP and ISTA, the proposed network learns parameters that need to be manually adjusted, such as threshold and step size, by deep learning approach. At the same time, an adaptive threshold is proposed, which enables the parameters to be not only layer-varied but also elementwise. Thus, compared with the traditional threshold shrinkage function, the proposed module has better retention ability of small and weak targets, as well as unveiling and utilizing the semantic information.

Compared with conventional sparse microwave reconstruction methods, ATASI-Net shows superiority in both speed and accuracy. First, ATASI-Net is designed into a feedforward network structure, which is suitable to be accelerated by GPU. Meanwhile, the iterative nature of conventional algorithms always makes them suffer from high data dependency and low parallelism. Second, ATASI-Net learns optimal parameters automatically in an end-to-end manner instead of a huge effort in tuning, which guarantees efficiency and avoids performance decreasing caused by mistuning of parameters. Third, ATASI-Net only learns the threshold and step parameters that are real-valued. Therefore, compared with deep networks that need to learn matrix parameters, we can significantly reduce the number of training parameters and system complexity [the number of parameters to learn is the same as ALISTA: $O(K)$], which in turn reduces the training burden and required training dataset size.

### IV. SIMULATION EXPERIMENTS

In the experimental part, simulation and real data are used to illustrate the efficiency and superiority of the proposed network. The simulation experiments include scattering point simulation and 3-D building model simulation. The settings of these two simulations are different, and the parameters of the simulation experiment of 3-D building facade are the same as those of the actual scene.

#### A. Experiment Settings

The training samples in our simulation include two types: one is the single scatterer sample with only one target point in the same range–azimuth cell and the other is the double scatterer sample with overlapping mask in the same range–azimuth cell.

For a single scatterer, a backscattering coefficient is a complex number, which can be expressed as $\gamma = A \exp(j \phi)$, where the amplitude $A$ is randomly distributed from 0 to 4 following Rayleigh distribution and the scattering phase $\phi$ is randomly distributed from 0 to $2\pi$ following uniform distribution. With the provided parameters and the known imaging geometry, the measurement matrix $D$ can be determined following (3). For double scatterers, the generation of the two single scatterers is identical to the previous step. In addition, for double scatterers, we also vary the elevation distance between the two single scatterers from 0.1 to 1.2 times of Rayleigh resolution. The echo signal is simulated at 11 different SNR levels between [0, 30 dB] with additive white Gaussian noise. The number of discretized grids in each range–azimuth cell along the s-direction is fixed to 300 in the range from 0 to 300 m with 1-m sampling. To avoid the off-grid bias, we assume that all scatterers locate on-grid.

#### B. ATASI-Net Configuration and Training

All the experiments are conducted on a platform with 4.70 GHz Intel Core i7-12700H CPU and NVIDIA
TABLE I

| Parameter         | Scatter Simulation | 3D Building Simulation | SARMV3D1.0 |
|-------------------|--------------------|------------------------|------------|
| Carrier frequency | 14.25 GHz          | 14.25 GHz              | 14.25 GHz  |
| Wavelength λ      | 0.021 m            | 0.021 m                | 0.021 m    |
| Antenna interval Δd| 0.084 m            | 0.084 m                | 0.084 m    |
| Array number n    | 8                  | 8                      | 8          |
| Range R₀          | 3200 m             | 1200 m                 | 1184-1304 m|
| Average incident angle θ₀ | 45°               | 30°                    | 24.9-34.5° |

RTX 3060 GPU (6-GB memory). During the training stage, the ATASI-Net is optimized with an Adam optimizer. The learning rate is set to 0.001, which decays dynamically according to the loss descent rate in the training process.

C. Scatterer Simulation Experiment

The simulation parameters of the scatterer experiment were set, as shown in Table I. For uniform linear arrays, the expected Rayleigh elevation resolution ρs can be calculated by ρs = λr/2nΔd, which is equal to 50 m in this simulation. The training dataset of this scatterer simulation experiment is 20,000, half of which are single scatterers and the other half are double scatterers, and the training data were generated as described above.

In the training procedure, we gradually increased the number of layers from 2 to 15 to determine an optimal network structure and compared conventional ALISTA with the proposed ATASI-Net in terms of the normalized mse (NMSE), which is defined as follows:

\[
NMSE = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{samples}}} \left( \frac{\| \hat{y}_i - y_i \|^2}{\| y_i \|^2} \right). \tag{28}
\]

The validation dataset contains 2000 samples simulated in the presence of noise with SNR = 30 dB using the same settings, as shown in Table I. Fig. 4 shows the performance of ALISTA and the proposed network. Closer inspection of Fig. 4 shows that the proposed ATASI-Net has smaller NMSE, better network performance, and faster convergence speed than ALISTA with the same layer number. Meanwhile, it can be seen that when the proposed network has more than ten layers, the performance payback of layer number will be diminished, as well as the training burden will be increased. Therefore, the ten-layer network is used in our experiment.

D. Double Scatterer Simulation Analysis

For the double scatterer simulation experiment, we implemented four different algorithms, including SVD, OMP, ALISTA, and the proposed ATASI-Net, for comparison. It is worth noting that the four methods represent three types of paradigms, including traditional spectrum estimation, the conventional CS-based algorithm, and the deep-network-based method. By changing the SNR and the distance between the two scatterers, six groups of 1000 samples each were used to analyze the simulation experiments of the double scatterers with different SNRs and different spacing. We employed a normalized distance α to represent the distance between the two scatterers, which is defined as follows:

\[
α = \frac{d_s}{ρ_s}. \tag{29}
\]

where \(d_s\) is the actual distance between the two scatterers along the elevation direction.

Fig. 5 shows the reconstruction results of different algorithms under different SNRs and α’s.

As we can see from Fig. 5, when α = 1.2, all four algorithms can successfully reconstruct the scatterers. However, when α = 0.3, the traditional CS-based greedy algorithm and the SVD spectrum estimation method cannot distinguish the targets with close distance, and only the sparse unfolded network ALISTA and the proposed ATASI-Net can distinguish them. In addition, when the SNR is high (20 dB), all four methods can obtain relatively accurate target information for the two scatterers in the same range–azimuth cell. Also, when the SNR is reduced to 10 dB, although all four methods can reconstruct the position information of the two scatterers in the elevation, the traditional spectrum estimation and the CS-based method have some errors in the estimation of reflection intensity, and the method based on the sparse unfolded deep network ALISTA and the proposed ATASI-Net can distinguish them. In particular, when the SNR is 0 dB, the traditional CS-based greedy algorithm and the spectrum estimation method of SVD lose their antinoise ability and cannot estimate the target position and scattering information. Moreover, the sparse unfolded deep network ALISTA has a certain deviation in the estimation of the target position along the elevation,
which is within an acceptable range. The accuracy of the proposed ATASI-Net for the position estimation of the two scatterers is similar to that of ALISTA, but the antinoise performance of the proposed ATASI-Net is better than those of the other three methods, because the adaptive threshold unit can filter the noise without affecting the retention of the target.

As our proposed ATASI-Net is unfolded based on the ISTA algorithm without a priori sparsity setting, there will be more than targets misestimated scatterers due to the interference of noise, which can be seen in Fig. 5(a) and (d) when SNR = 0 dB. This can be eliminated by subsequent sparsity constraints or by threshold filtering in the actual processing.

Then, we give the estimated position error and estimated reflectivity error at different SNRs. For each experiment, the Monte Carlo experiments with five thousand trials were executed. Also, the experimental results are shown in Figs. 6 and 7.

The estimated position error can be calculated by

$$\Delta s = \sqrt{\frac{1}{2N} \sum_{n=1}^{N} \left[ (\hat{s}_{n1} - s_{n1})^2 + (\hat{s}_{n2} - s_{n2})^2 \right]}$$

(30)

where \(N\) is the number of Monte Carlo simulations, \(\hat{s}_{n1}\) and \(\hat{s}_{n2}\) are the estimated positions of the two scattering points, \(s_{n1}\) and \(s_{n2}\) are the true values of the positions of the two scattering points, and \(\hat{\sigma}_{n1}\), \(\hat{\sigma}_{n2}\), \(s_{n1}\), and \(s_{n2}\) are the values normalized with Rayleigh resolution.

The estimated reflectivity error can be calculated by

$$\Delta |\sigma| (\text{dB}) = 20 \log \left( 1 + \frac{1}{\sqrt{\frac{1}{2N} \sum_{n=1}^{N} \left( (|\hat{\sigma}_{n1}| - 1)^2 + (|\hat{\sigma}_{n2}| - 1)^2 \right)} \right)$$

(31)

where \(|\hat{\sigma}_{n1}|\) and \(|\hat{\sigma}_{n2}|\) are the amplitude estimates of the two scattering points and \(|\sigma_{n1}|\) and \(|\sigma_{n2}|\) are the true values of the amplitudes of the two scattering points.

Since the goal of TomoSAR is to have a good elevation estimate, to more clearly illustrate the performance of the proposed ATASI-Net, we defined the detection success rate (DSR) according to the literature [33], [42], which is used to quantify the accuracy of the estimated positions. The DSR should satisfy that the estimated elevation of both the two
scatterers should be within ±3 times Cramer–Rao lower bound (CRLB) with respect to the ground truth and within ±0.5 times $d_s$ with respect to their true elevation. For the latter condition, which needs to be satisfied, it mainly acts as a constraint when the two scatterers are in close proximity.

According to [42], the CRLB $\sigma_d$ of the two scatterers can be calculated by

$$\sigma_d = c_0 \cdot \sigma_s$$

where

$$\sigma_s = \frac{\lambda R_0}{4\pi \sqrt{2N \cdot SNR \cdot \sigma_b}}$$

is the CRLB of the elevation estimates of single scatterer. $\sigma_b$ is the standard deviation of the elevation aperture sample positions, and for uniform linear array, $\sigma_b = \rho_s/(12)^{1/2}$ by the empirical formula

$$c_0 = \sqrt{\max\left\{\frac{40\alpha^{-2}(1 - \alpha/3)}{9 - 6(3 - 2\alpha) \cos(2\Delta\varphi + 2\pi \alpha (-\frac{1}{N})) + (3 - 2\alpha)^2}, 1\right\}}$$

where $\Delta\varphi$ is the phase difference between the two scatterers.

We compared the DSRs of different SNRs, $\alpha$’s, amplitude ratios, and phase differences. Moreover, for each experiment, the Monte Carlo experiments with five thousand trials were executed. First, we fixed the normalized distance $\alpha$ with 0.6, gradually increased the SNR from $-5$ to 20 dB, and compared the performance of the CS-based method, the sparse unfolded network ALISTA, and our proposed ATASI-Net; the results are shown in Fig. 8. It can be seen that the proposed ATASI-Net has stronger robustness and the antinoise performance is better than the other two methods. Second, we compared the DSRs of different algorithms under different $\alpha$’s with SNR = 10 dB. As can be seen from Fig. 9, the method based on the sparse unfolded network is superior to the traditional CS-based algorithm in the reconstruction of position information of the target with a short distance.

Third, we evaluated the performance of the proposed ATASI-Net at different amplitude ratios of the double scatterers with fixed SNR = 10 dB and $\alpha = 0.6$. As the amplitude ratio of the two scatterers increases, the scatterer with a smaller amplitude becomes less prominent. Therefore, at high amplitude ratios, the DSR will be decreased. However, our proposed ATASI-Net has a higher DSR for double scatterers with different amplitude ratios, as shown in Fig. 10.

Fourth, the phase difference between the double scatterers was changed in the simulation experiments to further verify the performance of the proposed ATASI-Net. Fig. 11 shows the DSR when the SNR = 10 dB and the normalized distance $\alpha = 0.6$. The double scatterers in the simulation were set to have the same amplitude. We can see that the DSR of the double scatterers is above 50% under different phase differences.

E. 3-D Building Simulation

The parameters of the 3-D building simulation experiment were set, as shown in Table I. These parameters were approximately the same as the parameters of the real data.

In this experiment, a 3-D scatterer model of a building with the roof, wall, and ground was constructed to visually compare the imaging performance of different algorithms. Fig. 12 shows the 3-D views of the simulated target. We compared the OMP algorithm, ISTA algorithm, ALISTA, and the proposed ATASI-Net. It is of note that the first three methods represent...
three different algorithms: the greedy-based traditional CS algorithm, the $L_1$ regularization-based traditional CS algorithm, and the sparse unfolded deep network. By changing the SNR, roof:wall:ground RCS ratios, and the number of targets (NoT), we compared the different methods regarding their qualitative visual effect and quantitative evaluation indicator. The specific experimental details are given as follows.

1) Visual Comparisons: In this section, we mainly compare various sparse reconstruction algorithms in terms of imaging quality by varying the SNR and the roof:wall:ground RCS ratio and qualitatively demonstrate the effectiveness of the proposed improvements in terms of visual effects.

First, the roof:wall:ground RCS ratio was fixed to 2:2:1 and NoT was fixed to 7000, while the SNRs were different. The results obtained are shown in Fig. 13. Noticeably, both ALISTA and the proposed ATASI-Net were capable of reconstructing the 3-D target precisely when SNR = 30 dB. Nevertheless, the reconstruction quality decreased as SNR decreased. Deep network-based methods appeared to yield images with intact profiles compared with conventional sparse reconstruction algorithms, such as OMP and ISTA. This is expected because network-based methods automatically learn the optimal parameters instead of requiring manual tuning, which avoids the performance deterioration caused by parameter mistuning. However, when the SNR decreased to 10 dB, OMP, ISTA, as well as ALISTA yielded images with noisy backgrounds. In contrast, the proposed ATASI-Net was still able to produce a well-reconstructed image due to its elementwise threshold.

After that, the SNR was fixed to 30 dB, and the images obtained with the different algorithms were compared by applying various roof:wall:ground RCS ratios. As a result, when the RCS ratios differed greatly, the outcome of the reconstruction worsened. It can be seen that in the reconstruction results, the reconstruction completeness and density of the roof were greater than that of the wall in the same resolution cell, which was due to the fact that scatterer targets with smaller RCS were ignored, resulting in different point cloud densities in different regions. Among the four algorithms, the proposed ATASI-Net produced the most compelling imagery results in all cases, demonstrating its superiority. The visual results shown in Fig. 14 are also consistent with the numerical evaluations in the scatterer simulation experiment, which also proves the robustness of ATASI-Net and its elementwise adaptive threshold.

To summarize, this section first demonstrates the noise immunity and robustness of the proposed method by varying different SNRs. Then, by varying different roof:wall:ground RCS ratios, it is validated that the traditional method based on fixed threshold will chop off the weak targets directly when the RCS differences between different targets are large, resulting in the loss of weak targets, while the proposed elementwise adaptive threshold method can preserve the weak targets to a certain extent, thus verifying that the proposed ATASI-Net has a better performance.

2) Numerical Analysis: In addition to visual assessment, we also compared the performance of the different algorithms through numerical analysis. The following evaluation metrics were used: peak signal-to-noise ratio (PSNR), normalized averaged root-mean-square error (NRMSE), and structural similarity index for measuring (SSIM). The first two evaluation indices measure the accuracy of the reconstructed scattering intensity, and the last evaluation index measures the overall reconstruction effect resulting from the spatial similarity. The different evaluation indicators are defined as follows.

The definition of PSNR is

$$\text{mse} = \frac{||\hat{X} - X_{\text{ref}}||^2_F}{\text{num}(X_{\text{ref}})}$$

$$\text{PSNR} = -10\log_{10}(\text{mse}(\hat{X}, X_{\text{ref}}))$$

where $\hat{X}$ is the image reconstructed by the network, $X_{\text{ref}}$ is the label image, and num$(X_{\text{ref}})$ indicates the total pixel number in the image $X_{\text{ref}}$. In addition, the definition of SSIM is

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\mu_y^2 + \mu_y^2 + c_1(\sigma_x^2 + \sigma_y^2 + c_2)}$$

$$c_1 = (k_1L)^2, \quad c_2 = (k_2L)^2$$

where $\mu_x$, $\mu_y$, and $\sigma_x$, $\sigma_y$, are the reconstructed image, reference image, and their mean values, respectively; $\sigma_{xy}$ are the standard deviations of $x$ and $y$, respectively; and $\sigma_{xy}$ represents their cross covariance.

Moreover, $L$ stands for the dynamic range of the pixel values. By default, $k_1$ and $k_2$ are set to 0.01 and 0.03, respectively. SSIM is set between 0 and 1. The larger SSIM is, the smaller the gap between the reconstructed image and the reference image, that is, the better the reconstruction quality.

Here, the RCS ratio and NoTs were fixed at 2:2:1 and 7000, respectively. The robustness of the different methods was evaluated with SNR values ranging from 0 to 30 dB, with incremental steps of 10 dB. The comparison results are shown in Table II, and the best evaluations are marked in boldface. The indices show that ATASI-Net achieved the best scores in most conditions, validating the robustness of the proposed method.

Table III shows the reconstruction results obtained with the different algorithms, applying different roof:wall:ground RCS ratios, with SNR being 30 dB and NoTs being fixed to 7000. As expected, for the other three methods with traditional fixed thresholds, as the RCS ratio increased, the difference in scattering intensity between the target points increased, and thus, the indices deteriorated. In contrast, in our proposed method, the proposed ATASI-Net produced the most compelling imagery results in all cases, demonstrating its superiority. The visual results shown in Fig. 14 are also consistent with the numerical evaluations in the scatterer simulation experiment, which also proves the robustness of ATASI-Net and its elementwise adaptive threshold.

Fig. 12. 3-D building scatterer model.
ATASI-Net, the thresholds showed an elementwise adaptation so that the weak target points were not directly removed by threshold shrinkage, thus improving the retention of weak target points to some extent.

Afterward, we experimentally compared the performance of the four kinds of algorithms when applying different NoTs. The RCS ratio and SNR were set to 2:2:1 and 30 dB, respectively. The NoTs were changed to the range of 1000–9000 dB.

### Table II

| SNR  | OMP PSNR | NRMSE | SSIM | IST PSNR | NRMSE | SSIM | ALISTA PSNR | NRMSE | SSIM | Proposed PSNR | NRMSE | SSIM |
|------|----------|-------|------|----------|-------|------|-------------|-------|------|---------------|-------|------|
| 30   | 25.739   | 0.261 | 0.909| 32.429   | 0.117 | 0.946| 37.283   | 0.076 | 0.991| 40.148       | 0.061 | 0.990|
| 20   | 24.251   | 0.344 | 0.867| 32.214   | 0.227 | 0.855| 37.034   | 0.083 | 0.910| 40.106       | 0.056 | 0.959|
| 10   | 22.483   | 0.676 | 0.731| 31.586   | 0.468 | 0.744| 36.257   | 0.272 | 0.840| 40.074       | 0.203 | 0.904|
| 0    | 21.179   | 1.063 | 0.522| 30.773   | 1.079 | 0.524| 35.068   | 0.664 | 0.647| 40.065       | 0.529 | 0.768|
| Avg. | 23.413   | 0.586 | 0.757| 31.751   | 0.473 | 0.767| 36.411   | 0.274 | 0.847| 40.099       | 0.212 | 0.905|

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
TABLE III

COMPARISONS OF PSNR (dB), NRMSE, AND SSIM WITH DIFFERENT ROOF:WALL:GROUND RCS RATIOS

| RCS     | OMP | IST | ALISTA | Proposed |
|---------|-----|-----|--------|----------|
|         | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM |
| 1:1:1   | 27.173 | 0.227 | 0.922 | 33.852 | 0.101 | 0.957 | 38.624 | 0.062 | 0.996 | 40.739 | 0.063 | 0.994 |
| 2:2:1   | 25.739 | 0.261 | 0.909 | 32.429 | 0.117 | 0.946 | 37.283 | 0.076 | 0.991 | 40.148 | 0.061 | 0.990 |
| 4:2:1   | 20.479 | 0.334 | 0.827 | 27.633 | 0.153 | 0.803 | 31.646 | 0.129 | 0.836 | 37.518 | 0.082 | 0.902 |
| 9:3:1   | 11.127 | 0.347 | 0.512 | 17.941 | 0.186 | 0.469 | 20.559 | 0.155 | 0.505 | 30.294 | 0.106 | 0.747 |
| Avg     | 21.130 | 0.292 | 0.793 | 27.964 | 0.139 | 0.794 | 32.028 | 0.106 | 0.832 | 37.175 | 0.078 | 0.908 |

TABLE IV

COMPARISONS OF PSNR (dB), NRMSE, SSIM, AND RUNTIME (S) WITH A DIFFERENT NoT

| NoTs | OMP  | IST  | ALISTA | Proposed |
|------|------|------|--------|----------|
|      | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | PSNR | NRMSE | SSIM | Runtime | PSNR | NRMSE | SSIM | Runtime |
| 1000 | 35.428 | 0.217 | 0.943 | **0.121** | 39.937 | 0.093 | 0.963 | 0.526 | 42.339 | 0.048 | 0.994 | 0.223 | 43.837 | 0.046 | 0.994 | 0.242 |
| 2000 | 33.774 | 0.221 | 0.936 | **0.212** | 38.241 | 0.098 | 0.967 | 0.716 | 41.173 | 0.054 | 0.992 | 0.267 | 42.559 | 0.051 | 0.992 | 0.283 |
| 3000 | 29.636 | 0.239 | 0.927 | 0.397 | 36.456 | 0.104 | 0.955 | 0.924 | 39.836 | 0.062 | 0.990 | 0.219 | 40.736 | 0.057 | 0.989 | 0.279 |
| 4000 | 25.739 | 0.261 | 0.909 | 0.584 | 32.429 | 0.117 | 0.946 | 1.109 | 37.283 | 0.076 | 0.991 | 0.245 | 38.148 | 0.061 | 0.990 | 0.316 |
| 5000 | 16.238 | 0.307 | 0.873 | 0.894 | 27.533 | 0.152 | 0.904 | 1.664 | 33.492 | 0.091 | 0.933 | 0.242 | 35.169 | 0.078 | 0.924 | 0.289 |
| Avg  | 28.163 | 0.249 | 0.918 | 0.442 | 34.920 | 0.113 | 0.947 | 0.984 | 38.825 | 0.066 | 0.980 | 0.239 | 40.090 | 0.059 | 0.978 | 0.282 |

Fig. 15. Real data images of urban area. The area in the white box is the target area of this experiment. (a) SAR image. (b) Optical image.

with an incremental step of 2000. A comparison of the PSNR, NRMSE, SSIM, and runtime scores is shown in Table IV.

The results of the runtime comparison show that as the NoT points increased, the runtime of the traditional CS-based algorithm also increased. Thus, the time-space complexity of the solution was high for the actual large-scale scenario. For the network-based methods, once the pretraining was completed, little computational time was required compared with conventional CS-driven algorithms. This may be explained by the pretraining process shifting the computational burden to the learning stage of the parameters, consequently yielding high reconstruction efficiency.

V. REAL DATA EXPERIMENT

In this section, we analyze the applicability of the proposed ATASI-Net for TomoSAR imaging with real radar data. The visual reconstructions based on both the traditional CS algorithm and the proposed method were compared, and the 3-D reconstruction results were expressed as point clouds. At the same time, structured modeling of the 3-D point clouds was carried out to achieve a thorough comparison. In addition,
Fig. 17. Point cloud reconstruction results of the last building were selected for structural modeling. (a) and (d) Last building point cloud results in front view. (b) and (e) Last building point cloud results in side view. (c) Structured modeling of the point cloud results based on the traditional CS algorithm. (f) Structured modeling of the point cloud results based on the proposed ATASI-Net.

we sliced the reconstruction results of the ground obtained with both the traditional CS algorithm and ATASI-Net and compared them. Moreover, we used the portion that exhibited the greatest focusing effect in the reconstruction from the traditional CS algorithm as the training set. This portion corresponded to points in the last building, and the aim was to evaluate the reconstruction results obtained under different labels. The specific experimental details are given as follows.

A. Dataset

In this section, we adopt real SAR data from the SARMV3D1.0 dataset [56]. SARMV3D1.0 is an airborne array interferometric SAR system and the data were obtained from an urban community in Wanrong County, Yuncheng City, Shanxi Province, China, by the Aerospace Information Research Institute, Chinese Academy of Sciences (AIRCAS). The array Interferometric SAR (InSAR) system has eight channels. Table I describes the scenario parameters. The optical and SAR images of the area, in which the imaging targets were residential 18-floor buildings, are shown in the white box of Fig. 15. The height of the building is 54 m. The generation of ATASI-Net model in this experiment was based on a previous 3-D building simulation experimental dataset, which shared the same scene parameters.

B. Comparison of Reconstruction Performance

The estimated 3-D imaging point clouds leveraging the CS-based method and proposed ATASI-Net are given in Fig. 16, and Table V provides numerical comparisons. Both methods successfully reconstructed the 3-D information of the target scene, and the relative heights of the buildings were also accurate. As there is an unavailability of tangible 3-D models or light detection and ranging (LiDAR) data pertaining to this particular region, the validation of the proposed technique was limited to a comparison of building heights. The reconstructed building height of the proposed methodology measured 53.6 m, aligning with the actual height and endorsing the efficacy of the approach. In addition, the reconstruction quality of the proposed ATASI-Net was substantially greater than that of the traditional CS method, both in terms of the focusing effect of the point clouds and the antinoise ability.

The 3-D entropy index in Table V can be defined as follows:

$$P(i, j) = \frac{f(i, j)}{(N_r \times N_a \times N_z)}$$

$$3D_{Entropy} = - \sum_{i=0}^{255} \sum_{j=0}^{255} P(i, j) \cdot \ln(P(i, j))$$

where $(i, j)$ is the combination of the pixel gray-level value $i$ $(0 \leq i \leq 255)$ and the local mean of neighbor domain $j$ $(0 \leq j \leq 255)$ to which the pixel belongs. $f(i, j)$ is the statistical quantity of $(i, j)$. Also, $N_r$, $N_a$, and $N_z$ represent the sampling number in the range direction, azimuth direction, and elevation direction, respectively. Entropy is a statistical measure of randomness, which can be used to characterize the

| Method   | Maximum Height (m) | 3D Entropy | Runtime (s) |
|----------|--------------------|------------|-------------|
| CS-based | 51.9               | 0.3780     | 68.964      |
| Proposed | 53.6               | 0.2204     | 0.813       |
texture of an image. Here, entropy was adopted to evaluate the focusing quality of the imaging results, with a smaller entropy score indicating higher image quality.

As shown in Table V, ATASI-Net yields lower 3-D entropy scores in comparison with traditional CS-based methods, which also indicates that the proposed ATASI-Net is capable of precisely reconstructing images with less clutter and higher focusing quality. In addition, ATASI-Net runs much faster compared with conventional CS-driven algorithms because the pretraining processing shifts the computational burden to the parameter learning stage and consequently results in high reconstruction efficiency.

The generated point cloud data are often applied to the structural modeling of engineering. To further illustrate the reconstruction effect of the two points clouds, we adopted an alpha-shape-based [57] structural modeling algorithm and selected the last building for structural modeling. The results are shown in Fig. 17. The 3-D modeling revealed that the point cloud data based on the traditional CS algorithm could not yield a proper model of the building structure because it has more stray points resulting in poor focus, and the roof and ground cannot be reconstructed well, whereas the proposed ATASI-Net successfully attained the 3-D modeling of the roof as well as the concave and convex prism of the building.

C. Ground Layer of the Traditional CS-Based Method, Conventional ALISTA, and the Proposed ATASI-Net Model

To further illustrate the effectiveness of the proposed algorithm, we examined the scattering information of the ground section in the reconstructed image. The corresponding results are shown in Fig. 18, where the red box highlights the ground details around the building, such as vehicles and fences, which could not be recovered satisfactorily by the traditional CS method and the conventional ALISTA-Net, as confirmed by the optical images shown in Fig. 15. However, due to the elementwise adaptive thresholding feature of our proposed method, it was able to capture the details of some weaker objects on the ground and consequently enhance the overall recovery of details. In addition, our ATASI-Net showed better focusing and reconstruction performance than the traditional CS algorithm.

D. Training Labels

Considering the situation that the imaging parameters of the system cannot be obtained or lost, we hope to realize that it is not necessary to know the parameters of the real imaging scene. Moreover, it is not necessary to generate the training set through simulation first and then use this training set for network training. Therefore, based on the results of the traditional CS method, we used some of the target points of the CS reconstruction as the training labels, without first

Fig. 18. Slices of ground target reconstruction results. (a) Original SAR image. (b) Traditional CS-based algorithm reconstructs result slice. (c) ALISTA-Net reconstructs result slice. (d) Proposed method reconstructs result slice. The proposed method has a superior reconstruction capability for weak targets. The ground targets in the red box can be seen in the optical map as fences and cars, for which the proposed method has better reconstruction.

Fig. 19. Comparison of reconstruction results of different training tags. (a) Result of traditional CS algorithm reconstruction, the chosen training set was the area in the last building. (b) Reconstruction result with the selected area based on (a) used as training label. Using the reconstruction obtained with the traditional CS algorithm as the training label improved the result obtained.
establishing the simulation scene. The selected area exhibited a more accurate reconstruction effect and fewer stray points, and the chosen training set was the last building shown in Fig. 16(b). As expected, using the reconstruction obtained with the traditional CS algorithm as the training label improved the result obtained, resulting in fewer stray points and better focus as shown in Fig. 19. Thus, by using the result obtained with the traditional method as the training label, the step of reconstructing the simulation scene can be omitted, and the 3-D reconstruction process is simplified to a certain extent.

E. Threshold-Semantic Feedbacks

Since the proposed method has the characteristic of element-wise adaptive threshold, the threshold learned by the proposed method was visualized, and the threshold was divided into three categories according to the size. This threshold was rendered with different colors, and the results are shown in Fig. 20. Pink, blue, and white represent the background, roof, and facade, respectively. The classification result of the threshold segmentation is roughly consistent with the actual semantic meaning and exhibits an adequate mutual feed with the semantic meaning. Therefore, in future studies, we will consider introducing semantics into 3-D reconstruction as a prior condition to achieve more accurate 3-D reconstructions and more effective restoration of information such as trees and other targets that are difficult to distinguish around the building.

VI. CONCLUSION

In this article, we propose a novel efficient deep unfolded network, named ATASI-Net, for solving the TomoSAR inversion. The network is designed as a combination of the CS-based iterative algorithm with a data-driven deep learning method, and the parameters are optimized by end-to-end training, thus avoiding complex tuning. The architecture of ATASI-Net is constructed using the iterative shrinkage-thresholding solver. In addition, the precomputation module replaces the learning of matrix parameters by solving a convex optimization problem so that only scalar parameters are used for learning, substantially improving the training efficiency of the network. Moreover, the threshold update unit in ATASI-Net introduces adaptive thresholds, making them not only layer-varied but also elementwise. Extensive experiments, including scattering point simulation experiments, 3-D building simulation experiments, and a SARMV3D1.0 real dataset experiment, demonstrate the superiority of the proposed ATASI-Net compared with the conventional CS-based algorithm and the deep-network-based method. Moreover, for real scenes with conventional reconstruction results, we propose using the data showing greater focus as the training set because this enhances the reconstructions in the absence of geometric parameters from real scenes. In addition, we compare the results of the point cloud reconstruction, using structured modeling, which is intuitive and can easily postprocess the point clouds. At the same time, we visualize the adaptive thresholds, and the results obtained are able to generate feedback in combination with semantics, which provides practical support for the introduction of semantics as prior constraints to achieve better reconstruction performance in the future.

REFERENCES

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, “A tutorial on synthetic aperture radar,” IEEE Geosci. Remote Sens. Mag., vol. 1, no. 1, pp. 6–43, Apr. 2013.
[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] X. X. Zhu and R. Banlur, “Very high resolution spaceborne SAR tomography in urban environment,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 12, pp. 4296–4308, Dec. 2010.
[4] V. Cazcarra-Bes, M. Pardini, M. Tello, and K. P. Papathanassiou, “Comparison of tomographic SAR reflectivity reconstruction algorithms for forest applications at L-band,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 1, pp. 147–164, Jan. 2020.
[5] G. D. Martino, A. Idice, D. Riccio, and G. Ruelle, “A novel approach for disaster monitoring: Fractal models and tools,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 6, pp. 1559–1570, Jun. 2007.
[6] V. Klemm and A. J. Laub, “The singular value decomposition: Its computation and some applications,” IEEE Trans. Autom. Control, vol. AC-25, no. 2, pp. 164–176, Apr. 1980.
[7] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Trans. Antennas Propag., vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[8] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” Proc. IEEE, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
[9] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[10] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[11] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” Appl. Comput. Harmon. Anal., vol. 26, no. 3, pp. 301–321, May 2009.
[12] N. Parikh and S. Boyd, “Proximal algorithms,” Found. Trends Optim., vol. 1, no. 3, pp. 127–239, Jan. 2014.

[13] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” Proc. Nat. Acad. Sci. USA, vol. 106, no. 45, pp. 18914–18919, 2009.

[14] 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.

[15] A. Budillon, A. Evangelista, and G. Schirinzi, “SAR tomography from sparse samples,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2009, pp. 865–868.

[16] X. X. Zhu and R. Bamler, “Tomographic SAR inversion by L1-norm regularization-the comprehensive sensing approach,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 10, pp. 3839–3846, Oct. 2010.

[17] X. X. Zhu and R. Bamler, “Super-resolution power and robustness of compressive sensing for spectral estimation with application to space-borne tomographic SAR,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 1, pp. 247–258, Jan. 2012.

[18] S. He et al., “SAR tomography imaging based on generalized orthogonal matching pursuit—The case study of pangu 7 star hotel in Beijing,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2016, pp. 6656–6658.

[19] D. Han, L. Zhou, Z. Jiao, B. Wang, Y. Wang, and Y. Wu, “Efficient 3D image reconstruction of airborne TomoSAR based on back projection and improved adaptive ISTA,” IEEE Access, vol. 9, pp. 47399–47410, 2021.

[20] J. Hershey, J. Le Roux, and F. Weninger, “Deep unfolding: Model-based inspiration of novel deep architectures,” 2014, arXiv:1409.2574.

[21] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in Proc. 27th Int. Conf. Mach. Learn., 2010, pp. 399–406.

[22] M. Borgerding, P. Schniter, and S. Rangan, “AMP-inspired deep networks for sparse linear inverse problems,” IEEE Trans. Signal Process., vol. 65, no. 16, pp. 4293–4308, Aug. 2017.

[23] Y. Yang, J. Sun, H. Li, and Z. Xu, “ADMM-CSNet: A deep learning approach for image compressive sensing,” IEEE Trans. Pattern Anal. Mach. Intell. vol. 42, no. 3, pp. 521–538, Mar. 2020.

[24] X. Chen, J. Liu, Z. Wang, and W. Yin, “Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds,” 2018, arXiv:1808.10038.

[25] C. A. Metzler, A. Mousavi, and R. G. Baraniuk, “Learned D-AMP: Principled neural network based compressive image recovery,” 2017, arXiv:1704.06625.

[26] S. Deshmukh, A. Dubey, and R. Murch, “Unrolled optimization with deep learning for compressed sensing,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 4706116.

[27] Y. Mai, S. Zhang, W. Jiang, C. Zhang, Y. Liu, and X. Li, “ISAR imaging of target exhibiting micro-motion with sparse aperture via model-driven deep network,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5110212.

[28] S. Wisdom, T. Powers, J. Pitton, and L. Atlas, “Deep recurrent NMF for speech separation by unfolding iterative thresholding,” in Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA), Oct. 2017, pp. 234–258.

[29] H. V. Luo, B. Joukovsky, and N. Deligiannis, “Designing interpretable recurrent neural networks for video reconstruction via deep unfolding,” IEEE Trans. Image Process., vol. 30, pp. 4099–4113, 2021.

[30] W. Pu, “Deep learning algorithm for SAR autofocus,” in Proc. 34th Gen. Assem. Sci. Symp. Int. Union Radio Sci. (URSI GASS), Aug. 2021, pp. 1–4.

[31] W. Pu, “Deep SAR imaging and motion compensation,” IEEE Trans. Image Process., vol. 30, pp. 2232–2247, 2021.

[32] M. Li et al., “Target-oriented SAR imaging for SCR improvement via deep MF-ADMM-Net,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5223314.

[33] M. Li, J. Wu, W. Huo, Z. Li, J. Yang, and H. Li, “STLS-LADMM-Net: A deep network for SAR autofocus imaging,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5226914.

[34] M. Wang et al., “TPPSS-Net: Fast and enhanced two-path iterative network for 3D SAR sparse imaging,” IEEE Trans. Image Process., vol. 37, pp. 7317–7332, 2021.

[35] S. Wei et al., “Learning-based split unfolding framework for 3-D mmW radar sparse imaging,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5229317.
Zhe Zhang (Member, IEEE) received the B.Sc. degree in information engineering from Xi’an Jiaotong University, Xi’an, China, in 2008, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences (IECAS), Beijing, China, in 2015. From 2015 to 2020, he was a Post-Doctoral Research Scientist with George Washington University, Washington, DC, USA, and George Mason University, Fairfax, VA, USA. He has been a Professor and a Ph.D. Supervisor with the Aerospace Information Research Institute, Chinese Academy of Sciences (AIRCAS), and the Suzhou Aerospace Information Research Institute, Chinese Academy of Sciences, 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.

Xiaolan Qiu (Senior Member, IEEE) received the B.S. degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree in signal and information processing from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2009. Since 2009, she has been working at the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. Her research interests include synthetic aperture radar (SAR) imaging, SAR simulation, and SAR 3-D reconstruction.

Silin Gao received the B.S. degree from the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China, in 2020. She is currently pursuing the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing.

Her research interests include sparse signal processing and synthetic aperture radar (SAR) imaging.

Yue Wang (Senior Member, IEEE) received the Ph.D. degree in communication and information system from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is currently an Assistant Professor with the Department of Computer Science, Georgia State University, Atlanta, GA, USA. Previously, he was a Research Assistant Professor with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, USA. His general research interests include machine learning, signal processing, wireless communications, and their applications in cyber-physical systems. His specific research focuses on compressive sensing, massive multi–input multi–output (MIMO), millimeter-wave communications, wideband spectrum sensing, cognitive radios, direction of arrival (DoA) estimation, high-dimensional data analysis, and distributed optimization and learning.