Technical University of Munich
TUM School of Engineering and Design

Deep Learning in Synthetic Aperture Radar Tomographic Inversion

Kun Qian

Complete reprint of the dissertation approved by the TUM School of Engineering and Design of the Technical University of Munich for the award of the

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

Chair: Prof. Dr. rer. nat. Niklas Boers

Examiners:

1. Prof. Dr.-Ing. habil. Xiaoxiang Zhu
2. Prof. Dr.-Ing. Yuanyuan Wang
3. Dr. Gianfranco Fornaro, Ph.D.

The dissertation was submitted to the Technical University of Munich on 28 September 2023 and accepted by the TUM School of Engineering and Design on 19 December 2023.

Abstract

Synthetic aperture radar tomography (TomoSAR) is an advanced interferometric SAR (InSAR) that utilizes stacked SAR images and stands as a powerful technique for 3-D imaging as well as deformation monitoring. To fully leverage the power of very high resolution (VHR) SAR data, state-of-the-art TomoSAR algorithms usually resort to compressive sensing (CS)-based sparse reconstruction which enables super-resolution ability and unprecedentedly high location accuracy. However, the sparse reconstruction cannot be solved analytically and usually iterative solvers need to be employed, thus making the CS-based TomoSAR algorithms extremely computationally expensive and not applicable for large-scale processing.

In recent years, great development of deep learning has been witnessed and extensive applications of deep learning have been found in numerous remote sensing fields due to its potent learning capabilities. A well-trained deep neural network can serve as an effective nonlinear function, thus representing complicated mathematical models including the CS problem. Motivated by this fact, the intention of this thesis is to explore the potential of deep learning in efficient and accurate super-resolving TomoSAR inversion. The work presented in this thesis contributes to the field by mainly addressing the following new aspects:

A novel deep neural network specifically designed for TomoSAR inversion called γ-Net is proposed by improving unrolled complex-valued Iterative Shrinkage Thresholding Algorithm (ISTA) network. γ-Net mimics the iterative optimization step in sparse reconstruction with only 10 to 20 layers. Experiments demonstrate that γ-Net approaches the Cramér-Rao Lower Bound while improving the computational efficiency by one to two orders of magnitude comparing to the state-of-the-art CS-based methods and also shows no degradation in the super-resolution power.

To promote sparsity of the output, shrinkage functions are usually employed in the deep neural networks. However, this shrinkage process inevitably results in information loss within the dynamics of the network, diminishing the model performance. To address this, a recurrent neural network (RNN) equipped with innovative sparse minimal gated units (SMGUs) is introduced aiming at mitigating the information loss. ***This novel RNN design capitalizes on integrating historical data into the optimization process***, ensuring comprehensive information preservation up to the final output, thus improving the model performance.

Considering that spaceborne datasets are typically acquired in a repeat-pass mode over different timestamps, sometimes spanning several years, it is essential to account for potential deformations in object estimations. This includes seasonal shifts from thermal dilation or linear motions, such as subsidence. The 4-D imaging technique, which factors in motion history for scatterers, is termed differential TomoSAR (D-TomoSAR). However, the integration of deep learning-based algorithms into D-TomoSAR inversion faces challenges due to the extensive high-dimensional weight matrices they require. These matrices often encompass millions of freely trainable parameters, leading to computational and memory bottlenecks. To address these challenges, ***a novel hybrid approach for efficient and***

precise D-TomoSAR inversion, called HyperLISTA-ABT is proposed. HyperLISTA-ABT draws from the learning structure of the unrolled ISTA network, while its weights are analytically determined based on a minimum coherence criterion, effectively slimming the model to an ultra-light variant with only three hyperparameters.

All the developed algorithms were tested and validated using both simulated laboratory data and extensive processing of real spaceborne data.

Zusammenfassung

Synthetik Apertur Radar Tomographie (TomoSAR) ist eine fortgeschrittene interferometrische SAR (InSAR) Methode, die gestapelte SAR-Bilder nutzt und als leistungsstarke Technik für die 3-D-Abbildung sowie die Deformationsüberwachung dient. Um die Leistungsfähigkeit von SAR-Daten mit sehr hoher Auflösung voll auszuschöpfen, greifen moderne TomoSAR-Algorithmen in der Regel auf Compressive Sensing (CS) basierende spärliche Rekonstruktion zurück, die eine Super-Auflösung und eine beispiellos hohe Lokalisation (also Schätzung der dritten Koordinate "Elevation") ermöglicht. Allerdings kann die spärliche Rekonstruktion nicht analytisch gelöst werden, und in der Regel müssen iterative Lösungsverfahren eingesetzt werden, wodurch die CS-basierten TomoSAR-Algorithmen extrem rechnerisch kostenintensiv werden und für weiträumig Verarbeitung nicht anwendbar sind.

In den letzten Jahren wurde eine erhebliche Entwicklung im Bereich des Deep Learning beobachtet, und zahlreiche Anwendungen des Deep Learning wurden in vielen Fernerkundungsbereichen aufgrund seiner leistungsstarken Lernfähigkeiten gefunden. Ein gut trainiertes tiefes neuronales Netzwerk kann als effektive nichtlineare Funktion dienen und somit komplizierte mathematische Modelle, einschließlich des CS-Problems, darstellen. Angeregt durch diese Tatsache, zielt diese Arbeit darauf ab, das Potenzial des Deep Learning für eine effiziente und präzise Super-Auflösende TomoSAR-Inversion zu erforschen. Die in dieser Arbeit vorgestellten Beiträge betreffen hauptsächlich die folgenden neuen Aspekte:

Ein neuartiges tiefes neuronales Netzwerk speziell für die TomoSAR-Inversion ist vorgeschlagt, das als $\boldsymbol{\gamma}$-Net bezeichnet wird und das fortschrittliche komplexwert Iterative Shrinkage Thresholding Algorithm (ISTA) Netzwerk verbessert. $\boldsymbol{\gamma}$-Net simuliert den iterativen Optimierungsschritt in der spärlichen Rekonstruktion mit nur 10 bis 20 Schichten. Experimente zeigen, dass $\boldsymbol{\gamma}$-Net sich der Cramér-Rao Lower Bound (CRLB) nähert und gleichzeitig die Recheneffizienz im Vergleich zu den modernsten CS-basierten Methoden um 1 bis 2 Größenordnungen steigert und keine Degradation in der Super-Auflösungskraft zeigt.

Um die Sparsamkeit der Ausgabe zu fördern, werden in den tiefen neuronalen Netzwerken normalerweise Schrumpfungsfunktionen verwendet. Dieser Schrumpfungsprozess führt jedoch unweigerlich zu Informationsverlusten in den Dynamiken des Netzwerks und beeinträchtigt die Modellleistung. Um dies zu beheben, ein rekurrentes neuronales Netzwerk (RNN) mit innovativen spärlichen minimalen gesteuerten Einheiten (SMGUs) eingeführt, das darauf abzielt, den Informationsverlust zu verringern. ***Dieses neuartige RNN-Design nutzt die Integration von historischen Daten in den Optimierungsprozess***, um eine umfassende Informationsbewahrung bis zum endgültigen Ausgang sicherzustellen und somit die Modellleistung zu verbessern.

Da satellitengestützte Datensätze in der Regel im wiederholten Pass-Modus über verschiedene Zeitstempel erfasst werden, manchmal über mehrere Jahre hinweg, ist es wichtig, mögliche Deformationen bei Objektschätzungen zu berücksichtigen. Dies beinhaltet saisonale Verschiebungen durch thermische Ausdehnung oder lineare Bewegungen wie

Absenkungen. Die 4-D-Abbildungstechnik, die die Deformationsgeschichte von Objekte berücksichtigt, wird als Differential-TomoSAR (D-TomoSAR) bezeichnet. Die Integration von Deep-Learning-basierten Algorithmen in die D-TomoSAR Inversion stellt jedoch aufgrund der umfangreichen hochdimensionalen Gewichtsmatrizen, die sie benötigen, Herausforderungen dar. Diese Matrizen umfassen oft Millionen von frei trainierbaren Parametern, was zu rechnerischen und Speicherengpässen führt. Um diesen Herausforderungen zu begegnen, ***ein neuartigen hybriden Ansatz für eine effiziente und präzise D-TomoSAR-Inversion***, benannt als HyperLISTA-ABT, ist introduziert. HyperLISTA-ABT basiert auf der Lernstruktur des aufgerollten ISTA-Netzwerks, wobei seine Gewichte analytisch auf Basis eines Mindestkohärenzkriteriums bestimmt werden, wodurch das Modell effektiv auf eine ultraleichte Variante mit nur drei Hyperparametern reduziert wird.

Alle entwickelten Algorithmen wurden sowohl mit simulierten Labor Daten als auch mit extensiver Verarbeitung von echten satellitengestützte Daten getestet und validiert.

Contents

List of Figures

List of Tables

Acronyms

2-D	two-dimension.
3-D	three-dimension.
AIC	Akaike Information Criterion.
ANM	Atomic Norm Minimization.
BIC	Bayesian Information Criterion.
BSS	Blined Source Separation.
CRLB	Cramér–Rao lower bound.
CS	Compressive Sensing.
DEM	Digital Elevation Model.
EVD	Eigenvalue Decomposition.
InSAR	Interferometric Synthetic Aperture Radar.
ISTA	Iterative Shrinkage and Thresholding Algorithm.
LASSO	Least Absolute Shrinkage and Selection Operator.
LISTA	Learned Iterative Shrinkage and Thresholding Algorithm.
LOS	Line-of-Sight.
MDL	Minimum Description Length.
MSE	Mean Square Error.
NLS	Non-linear Least Squares.
PCA	Principle Component Analysis.
PSI	Persistent Scatterer Interferometry.
RIP	Restricted Isometry Property.
RNN	Recurrent Neural Network.
SAR	Synthetic Aperture Radar.
SLC	Single-Look Complex.
SMGU	Sparse Minimal Gated Unit.
SNR	Signal-to-Noise Ratio.

TomoSAR	Synthetic Aperture Radar Tomography.
VAE	Variational Auto-encoder.
VHR	Very High Resolution.

1 Introduction

1.1 Motivation

In earth observation, synthetic aperture radar (SAR) is a crucial remote sensing tool that allows for contact-free retrieval of large-scale information regarding the physical properties of the earth surface by emitting electromagnetic waves towards the target and recording the backscattered echoes. As an active sensor, SAR is capable of penetrating clouds, vegetation canopies, soil, and snow [5] [6]. SAR interferometry (InSAR) uses co-registered SAR stacks to retrieve the 3-D localization as well as motion information of the scatterering objects. It plays an indispensable role in earth observation since it is to date the only method that allows for assessing millimeter-precision ground deformation over a long period on a global scale.

The launch of new SAR sensors, such as TerraSAR-X/TanDEM-X [7] and COSMO-Skymed [8] satellites, marked a significant advancement in SAR remote sensing from space. Beyond their many advantages, including high absolute geometric accuracy, precise orbit determination, and short revisit times, these SAR sensors offer spatial resolutions of up to 1 meter. However, for these very high-resolution (VHR) SAR data, the inherent elevation resolution is approximately 50 times poorer than that in range and azimuth. This discrepancy is attributed to the small elevation aperture resulting from the tight orbit tube of the satellite sensors. To address this extreme anisotropy and fully explore the potential of VHR SAR data, compressive sensing [9] [10] [11] (CS)-based TomoSAR algorithms have been proposed, enabling super-resolution in the elevation direction [12] [13] [14] [15] [16].

While CS-based TomoSAR algorithms deliver superior performance, they suffer from a substantial computational expense. The challenge stems from the inability to analytically solve sparse reconstruction in CS, necessitating the use of computationally expensive iterative solvers. As a result, the application of CS-based TomoSAR algorithms for large-scale processing is notably challenging. Hence, the development of both super-resolution and computationally efficient TomoSAR algorithms continues to be a crucial and compelling area of research.

1.2 Objective

Deep learning, due to its potent learning capabilities, has experienced swift advancements and found extensive applications in various remote sensing fields [17], including SAR data processing [18]. In particular, a deep neural network can serve as an efficient nonlinear function, adeptly representing complex mathematical models, including those encountered in CS problems [19]. Inspired by these developments, this thesis aims to investigate computationally efficient and generic deep learning-based algorithms tailored for super-resolution TomoSAR processing. Therefore, the main objectives of this thesis can be summarized as follows:

Development of novel neural networks specifically designed for TomoSAR inversion

Deep neural networks, as conventionally designed, possess an inherent limitation: they are unable to leverage the sparsity in specific signals, particularly those pertinent to TomoSAR inversion. This limitation not only hampers their efficiency but also precludes their applicability in solving TomoSAR inversion. The goal is to develop a novel class of deep neural networks specifically designed for solving TomoSAR inversion. Such deep learning-based algorithms should improve the computational efficiency while maintain the performance in realm of super-resolution power and location accuracy.

Performance improvement by incorporation of historical information into dynamics of the network.

Deep unfolding [20], as an emerging deep learning technique, allows us design neural networks by unrolling iterative algorithms. Several unrolled neural networks designed for TomoSAR inversion, such as the learned iterative shrinkage thresholding algorithm (LISTA) and its derivatives, utilize shrinkage functions to trim elements of minimal magnitude. However, this shrinkage process inevitably results in information loss within the dynamics of the network, diminishing the model performance. To address this, we aim to propose an advance neural network architecture to make use of historic information in the dynamics of the neural network, thus ensuring comprehensive information preservation up to the final output.

An ultra-light model for multi-component D-TomoSAR processing.

Deep learning-based algorithms, in their current form, are limited to three-dimensional (3-D) reconstruction. The extension of deep learning-based algorithms to four-dimensional (4-D) imaging, i.e., differential TomoSAR (D-TomoSAR) applications, is relatively unexplored, predominantly due to the high-dimensional weight matrices required by the network designed for D-TomoSAR inversion. These weight matrices typically contain millions of freely trainable parameters. Learning such huge number of weights requires an enormous number of training samples, resulting in a large memory burden and excessive time consumption. In light of these challenges, our research direction centers on introducing an ultra-light model offering efficient training. In addition, this model is supposed to ensure swift and accurate inference.

1.3 Outline

The remainder of this thesis is organized as follows. Chapter 2 gives a brief introduction to fundamentals of SAR tomography and deep unfolding. An overview of related algorithms for TomoSAR inversion are introduced and discussed in Chapter 3. Chapter 4 summarizes the main contributions of the attached journal papers from the author. A reconstruction result using the proposed algorithm over a large area is demonstrated in Chapter 5. Chapter 6 draws a conclusion of this thesis and discusses some potential further research directions.

2 Fundamentals

In this chapter, we begin by offering an overview of repeat-pass SAR tomography, covering the fundamentals of SAR geometry and the TomoSAR/D-TomoSAR system model. Additionally, we introduce the concept of deep unfolding, which involves unrolling iterative algorithms to design innovative neural networks.

2.1 Overview of repeat-pass SAR tomography

2.1.1 SAR geometry

Synthetic Aperture Radar is an advanced imaging radar technique that utilizes the relative motion between the antenna and the target to achieve remarkable spatial resolution, surpassing that of conventional beam-scanning radars. SAR allows for high-resolution remote sensing regardless of the flight altitude and enables wide-area coverage. One of its notable advantages is the capability to achieve cross-range resolution that is significantly finer than the beamwidth of the real aperture.

SAR imaging utilizes the motion of the radar antenna to simulate a large "synthetic" antenna or aperture, enabling the generation of high-resolution images through advanced signal processing techniques. The SAR geometry encompasses the spatial configuration of the SAR system, which includes the trajectory of the SAR platform's flight path, the orientation of the radar antenna, and the swath width, defining the ground area imaged by the radar.

The geometry of SAR plays a crucial role in determining the quality of SAR images. The most commonly used SAR geometry is side-looking, where the radar antenna is oriented perpendicular to the flight direction. SAR systems can be categorized as either monostatic, where the radar transmitter and receiver are on the same platform, or bistatic, where the transmitter and receiver are on separate platforms. For a normal monostatic SAR system, the native 3-D reference frame of a SAR sensor comprises the azimuth (flight direction), range (LOS direction), and elevation direction (perpendicular to the azimuth-range plane), as it is shown in Fig. 2.1. Bistatic SAR offers more flexibility in imaging scenarios as the geometry configuration can be arbitrary, enabling diverse and specialized imaging capabilities.

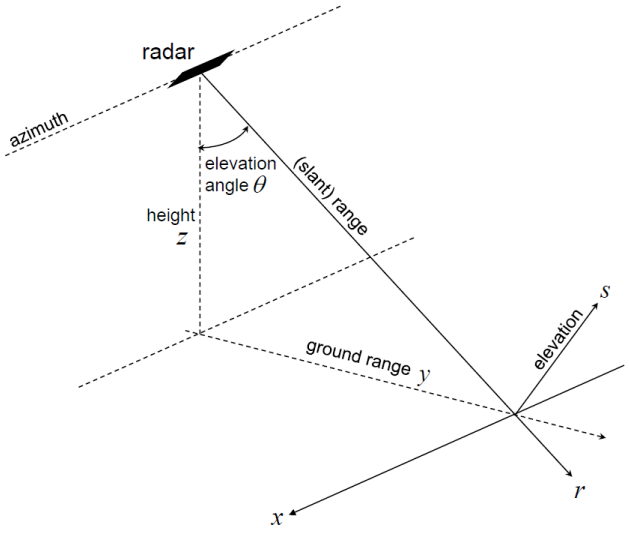


Figure 2.1: Radar side-looking imaging geometry, source in [1]. The flight direction of the sensor is commonly referred to as the "along-track" direction or azimuth (x). The line-of-sight (LOS) direction of the antenna is defined by the elevation angle (θ), often known as the range or slant range (r). The direction perpendicular to the azimuth-range (x-r) plane, known as the cross-range direction, is frequently denoted as the elevation (s). Technically, the third dimension should represent the elevation angle (θ). However, given the considerable range distance, spanning hundreds of kilometers, it is generally acceptable to assume a straight line approximation for small angular variations.

SAR mapping involves three distinct directions, each operating on different principles:

- Azimuth: The synthetic aperture principle is employed, leveraging the movement of the sensor to create a high-resolution image in the azimuth direction.
- Range: This direction is based on radar ranging principles. The transmitter emits pulses, such as chirps, which are reflected back from objects on the ground. By measuring the time it takes for the echoes to return to the receiver, objects at varying distances from the sensor can be resolved, as their echoes exhibit different two-way travel times.
- Elevation: SAR imaging does not provide direct mapping in the elevation direction. Consequently, targets located at the same azimuth-range (x-r) coordinates but with different elevation positions cannot be distinguished from one another.

The conventional two-dimensional backscattering map consists of complex-valued pixels that encompass both amplitude (brightness) and phase information. Disregarding any supplementary phase induced by the scattering process, the non-ambiguous phase ϕ of a pixel with a range distance r and at zero Doppler position x can be written as:

$$\phi = -\frac{4\pi}{\lambda} r \tag{2.1}$$

where λ is the wavelength. The phase ϕ is sensitive to the range (r), because λ is typically in range of centimeters.

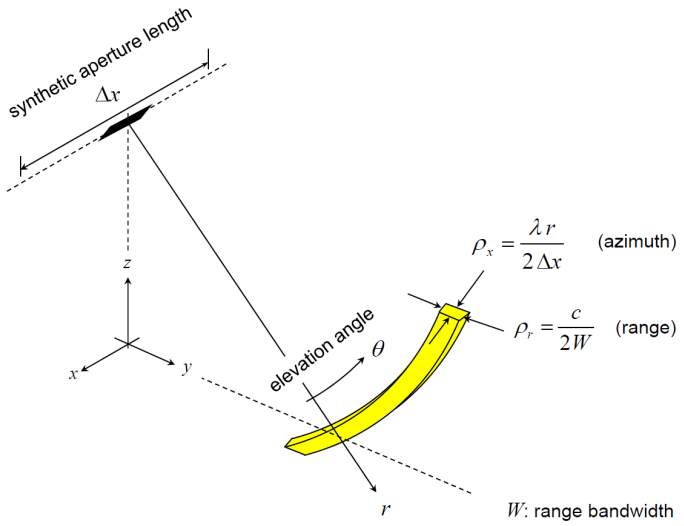


Figure 2.2: 2-D resolution element, source in [1], i.e. the azimuth reolution ρ_x and range resolution ρ_r, of a SAR image. ρ_x is determined by the azimuth synthetic aperture length $\triangle$x and ρ_r is determined by chirp bandwidth emitted by the antenna [2]

.

The three-dimensional resolution element of a conventional Synthetic Aperture Radar (SAR) is depicted in Fig. 2.2. The azimuth resolution, denoted as ρ_x, of the SAR image is determined by the length of the azimuth synthetic aperture, referred to as $\triangle$x ([21] [22]).

$$\rho_x = \frac{\lambda r}{2\Delta x} \tag{2.2}$$

The chirp bandwidth W determines the achievable range resolution ρ_r ([6]):

$$\rho_r = \frac{c}{2W} \tag{2.3}$$

where c is the speed of light.

2.1.2 TomoSAR system model

Conventional SAR imaging projects the 3-D object reflection to the 2-D azimuth-range (x-r) plane [5]. To achieve higher SAR imaging, advanced interferometric SAR (InSAR) techniques are usually required [23]. InSAR techniques utilize two or more complex-valued SAR images acquired from slightly different positions. By leveraging the phase differences present in the images, geometric information about the imaged objects can be extracted [2] [24] [25] [26] [27] [28] [29] and digital elevation models (DEMs) [30]. Hence, InSAR provides us with the ability to access the third dimension, which is elevation $\boldsymbol{s}$ (perpendicular to the azimuth and range plane). However, when using only two acquisitions, it is not possible to distinguish between multiple scatterers present within a single resolution unit along the elevation.

To generate a real 3-D reconstruction, i.e. imaging also in elevation direction [31] [32] [33] [34] [35] [36], TomoSAR constructs a synthetic elevation aperture by combining a stack of N complex repeat-pass SAR acquisitions captured at different times and slightly varied orbit positions (see Fig. 2.3). TomoSAR is occasionally referred to as an extention

of a multi-scatterer persistent scatterer interferometry (PSI) [37] [38], despite the fact that PSI primarily relies on the inversion of phase double-differences between neighboring persistent scatterers.

A single-look complex (SLC) SAR measurement of a certain azimuth-range (x-r) pixel for the n^{th} acquisition at the aperture position b_n can be formulated as the integral of the echoed signal along the elevation direction, with a linear phase term as the weighting factor (we focused on 3-D reconstruction and ignore possible deformation here):

$$g_n = \int_{\Delta s} \gamma(s) \exp\left(-j2\pi\xi_n s\right) \mathrm{d}s \tag{2.4}$$

where $\gamma(s)$ denotes the reflectivity function along the elevation $\boldsymbol{s}$. $\xi_n = \frac{-2b_n}{\lambda r}$ is the spatial frequency. Great details about the deviation of the Eq. (2.4) can be found in [33] [34] [36]. Discretizing the continuous reflectivity function along the elevation $\boldsymbol{s}$ within its extent Δs, we can approximate the TomoSAR imaging model by replacing the integral in Eq. (2.4) with a finite sum. Then, the discrete TomoSAR system model is expressed as:

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{2.5}$$

where $\mathbf{g} = [g_1, g_2, \cdots, g_N]^T$ is the SAR measurement vector and $\mathbf{R} \in \mathbb{C}^{N\times L}$ is the steering matrix with $R_{nl} = \exp\{-j2\pi\xi_n s_l\}$. $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_L]^T$ is the discrete reflectivity profile with $[s_1, s_2, \cdots, s_L]$ being the the discrete elevation positions.

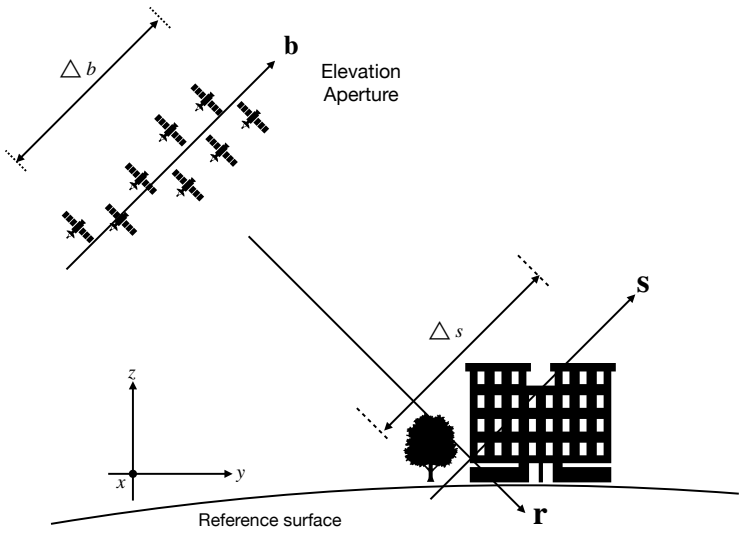


Figure 2.3: The SAR imaging geometry at a fixed azimuth position. The elevation synthetic aperture is built up by acquisition from slightly different incidence angles. Flight direction is orthogonal into the plane.

Now we briefly review some theoretical bounds on the performance of elevation reconstruction. The spread of sensor positions in repeat-pass SAR acquisitions creates an elevation aperture, denoted as Δb (see Fig. 2.3). This synthetic elevation aperture plays a crucial role in determining the Rayleigh resolution, i.e. the width of the elevation point response function. For a given elevation aperture size Δb, the Rayleigh resolution is approximately:

$$\rho_s = \frac{\lambda r}{2\Delta b} \tag{2.6}$$

For the TerraSAR-X [7] [39] [40] data used in this thesis, the Rayleigh resolution is typically in range of tens of meters. This value is significantly coarser compared to the azimuth and range resolutions achieved in the spotlight modes [41], which are typically at the meter or sub-meter levels.

2.1.3 Differential TomoSAR system model

In the case of space-borne SAR, repeat-pass acquisitions are conducted over a time span ranging from several weeks to years, depending on the revisit time of the satellite and the number of stacked images. Consequently, it becomes necessary to consider the long-term motion of the scattering object during the acquisition period. By incorporating a potential motion term, the TomoSAR system model Eq. (2.4) can be extended its differential version [42]:

$$g_n = \int_{\Delta s} \gamma(s) \exp\left(-j2\pi \left(\xi_n s + 2d\left(s, t_n\right)/\lambda\right)\right) \mathrm{d}s \tag{2.7}$$

where g_n is the complex-valued SAR acquisition at a certain azimuth-range pixel at time t_n $(n = 1, 2, \cdots, N)$. $\boldsymbol{\gamma}(s)$ denotes the scatterering profile along the elevation direction with an extent of $\triangle s$. $\xi = 2b_n/\lambda r$ is the elevation frequency proportional to the respective aperture position b_n. $d(s, t_n)$ depicts the LOS motion, which is a function of elevation and time. The LOS motion relative to the master acquisition can be modeled with a linear combination of M base functions $\tau_M(t_n)$

$$d\left(s, t_n\right) = \sum_{m=1}^{M} p_m(s) \tau_m\left(t_n\right) \tag{2.8}$$

where $p_m(s)$ is the corresponding motion coefficient to be estimated. The choice of the base functions $\tau_m(t_n)$ depends on the underlying physical motion processes. Great details about how to choose proper base functions can be found in [43].

Taking multi-component motion into consideration, we generalize Eq. (2.7) as:

$$\begin{aligned} g_n = \int \ldots \iint \gamma(s) \delta\left(p_1 - p_1(s), \ldots, p_M - p_M(s)\right) \\ \exp\left(j2\pi \left(\xi_n s + \eta_{1,n} p_1 + \ldots + \eta_{M,n} p_M\right)\right) ds dp_1 \ldots dp_M \end{aligned} \tag{2.9}$$

The inversion of the system model with multi-component motion retrieves the elevation information as well as the the motion history assigned to each elevation position, even if multiple scatterers are overlaid inside an resolution unit. Therefore, we can acquire a high-dimensional map of scatterers.

2.2 Introduction to deep unfolding (algorithm unrolling)

In the past decade, we have witnessed a profound revolution in deep learning in various fields, including computer vision, speech processing and so on. Significant progress in neural network research has achieved remarkable performance improvements, notably with the introduction of AlexNet [44], which demonstrated groundbreaking results.

Modern neural networks typically employ a hierarchical architecture with multiple layers and a large number of parameters, often numbering in the millions. This design enables them to learn intricate mappings that are challenging to explicitly design. Their adaptability, especially when provided with sufficient training data, allows deep networks to overcome model limitations, particularly in scenarios where accurately characterizing the underlying physical processes is difficult. Another advantage of deep networks is their efficient inference process. The computation through the network layers during inference can be executed rapidly, taking advantage of optimized computational platforms. Additionally, deep networks often have fewer layers compared to the iterations required in iterative algorithms. This characteristic makes deep learning methods computationally advantageous over traditional iterative approaches.

However, it should be noted that the majority of deep learning-based algorithms are purely data-driven and suffer from lacking interpretability. Usually, general network structures are applied and complex mappings are learned through end-to-end training. Hence, it becomes challenging to extract meaningful insights from the model weights, which are typically high-dimensional, and to understand the individual roles of these learned weights. In the field of signal processing, interpretability plays a crucial role in conceptual understanding and identifying system limitations. Therefore, the lack of interpretability can be a significant drawback of conventional deep learning-based methods compared to iterative model-based approaches. In addition to interpretability, generalizability is another issue that usually arises in deep learning. The practical success of deep learning-based models is heavily dependent on high-quality training data. In scenarios where there is no sufficient training samples, such as 3-D reconstruction [45] and medical imaging [46] [47], the performance of deep networks may significantly degrade.

Therefore, it has been constantly pursuing interpretable, generalizable, and high-performance deep learning-based algorithms in the field of signal processing. Considering the strong interpretablity and generalizability of iterative model-based methods, an insightful technique called deep unfolding or algorithm unrolling [20] [48] [3] was introduced to establish connection iterative algorithms and neural networks. As we can see in Fig. 2.4, in the framework of deep unfolding, each iteration of the model-based algorithm is represented as an intermediate layer in the neural network. By concatenating these layers, we can build a deep neural network. Passing data through the network is equivalent to executing a finite number of iterations. Additionally, the fixed system matrices in model-based algorithms are transferred to the network weights, which can be learned through end-to-end training using backpropagation.

2.2.1 Power of deep unfolding

Many recent works [49] [50] [51] [52] indicate that unrolled neural networks are effective in achieving superior performance and high efficiency in many practical domains. Fig. 2.5 offers a comparison of unrolled algorithms, traditional iterative algorithms, and generic neural networks from the perspective of functional approximation.

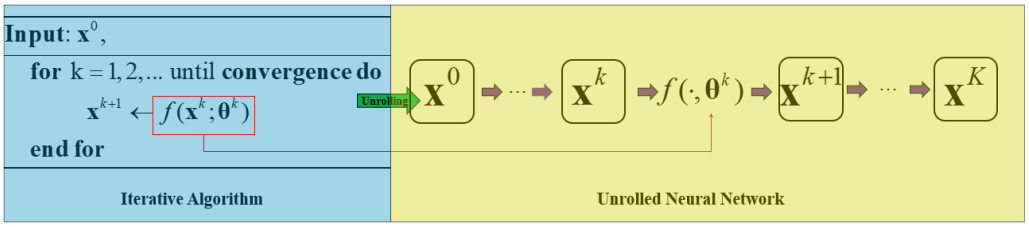


Figure 2.4: Illustration of the general idea of deep unfolding, modified according to [3]. Starting with an abstract iterative algorithm, we map one iteration into a single network layer. By stacking a finite number of these layers together, we form a deep neural network. When we feed the data forward through this K-layer network, it is equivalent to executing K iterations (finite truncation). The parameters $\boldsymbol{\theta}^k$ (k ranging from 0 to $k-1$) are learned by end-to-end training.

Traditional iterative algorithms, though tailored and refined for specific tasks, only cover a relatively narrow subset of the function space, thereby limiting their representational capacity. Nevertheless, they can approximate a given target function with a reasonable degree of accuracy, despite certain performance gaps. More importantly, they tend to generalize effectively within constrained training scenarios. Within the framework of statistical learning, iterative algorithms are viewed as models characterized by high bias but low variance.

In contrast, generic neural networks, with their universal approximation capabilities, are capable of providing a more accurate approximation of the target function. However, the high quantity of parameters they possess constitutes a broad subset within the function space, resulting in an expansive search space during the training phase. This large search space poses substantial challenges for network training. Moreover, the high dimensionality of parameters necessitates extensive training samples, creating potential issues with generalization. As the network size grows, efficiency may also become a concern. In essence, generic neural networks represent models with high variance but low bias.

Table 2.1: Comparison among iterative algorithms, generic neural networks and unrolled neural networks.

Technique	performance	Efficiency	Parameterization	Interpretability	Generalizability
Iterative Algorithms	Low	Low	Low	High	High
Generic Neural Networks	High	High	High	Low	Low
Unrolled Neural Networks	High	High	Middle	High	Middle

Unrolled networks, by augmenting the capacity of iterative algorithms, present a unique advantage. They can approximate the target function more accurately while still covering a relatively compact subset within the function space. This results in a reduced search space, thereby alleviating the demands of the training process and lessening the need for large-scale training datasets. Since iterative algorithms are meticulously constructed based on domain-specific knowledge and already provide reasonable approximations to the target function, enhancing them and training them on real data allows unrolled networks to frequently achieve highly accurate approximations of target functions. Positioned as an intermediate between generic networks and iterative algorithms, unrolled networks generally exhibit relatively low bias and variance simultaneously, presenting a balanced, effective solution.

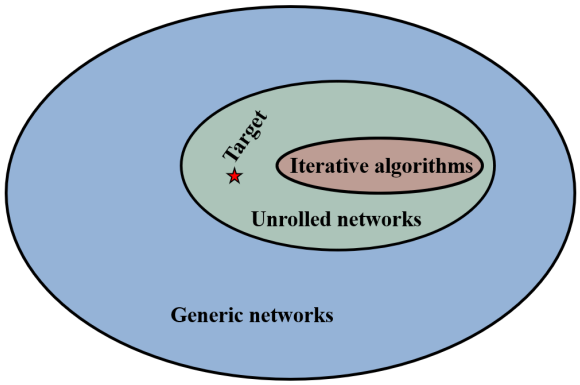


Figure 2.5: The concept of algorithm unrolling can be understood from a functional approximation perspective, as illustrated by the ellipses representing the scope of functions that can be approximated by each method category. Iterative algorithms, marked by their limited representational capacity, often underfit the target function due to a smaller scope of approximation. Unrolled networks, on the other hand, possess a broader representational capacity and typically provide a more accurate approximation of the target function. Generic neural networks have the most extensive representational power, with the capability to approximate a comprehensive scope of functions. However, their generalization capability in practical scenarios often falls short of expectations due to overfitting risks and extensive training requirements. Modified according to [3].

Table 2.1 compares the features of iterative algorithms, generic and unrolled neural networks. The power and advantages of deep unfolding can be summarized as the following key factors:

- Domain Knowledge Incorporation: Deep unfolding enables the explicit incorporation of domain knowledge and prior information into the network architecture. By mapping algorithm iterations to network layers, the network can effectively capture the domain-specific problem structures, constraints, and characteristics. This targeted and tailored approach enhances performance, distinguishing it from generic networks lacking such domain-specific knowledge.

- Parameter efficiency: Unrolled networks typically possess fewer parameters than generic networks, attributed to the direct integration of algorithm structure and operations into the network layers. Leveraging the inherent problem structure, unrolled networks achieve comparable or superior performance with a reduced parameter count, leading to improved computational efficiency.

- Supervised training strategy: Unrolled networks commonly employ paired input-output training, facilitating end-to-end learning. This training approach allows the network to learn optimal parameters directly from problem-specific data, resulting in enhanced performance. The paired input-output training scheme also enables fine-tuning of the network's behavior to align with the desired problem-solving objectives.

- Exploiting iterative optimization: Deep unfolding integrates and encapsulates the iterative steps of model-based algorithms within network layers, effectively merging the advantages of iterative optimization and deep learning. The iterative nature of

these algorithms enables the network to iteratively refine its predictions, resulting in improved performance and convergence.

3 Related works

This chapter provides an overview of significant methods related to TomoSAR inversion. The chapter is divided into two parts. Section 3.1 presents a brief discussion on the limitations of conventional spectral estimation methods and offers a review of the state-of-the-art TomoSAR algorithm based on compressive sensing. In Section 3.2, several data-driven TomoSAR algorithms are introduced.

3.1 State of the art in TomoSAR inversion

3.1.1 Limitation of conventional spectral estimators

Referring to the TomoSAR system model Eq. (2.5), the objective of TomoSAR inversion is to retrieve the reflectivity profile $\boldsymbol{\gamma}$ for each range-azimuth cell from the complex-valued SAR acquisitions $\mathbf{g}$. This involves estimating various scattering parameters, including the number of scatterers, their elevation, reflectivity, and possibly the motion history of the scatterers when temporal baselines are present.

Due to the fact that TomoSAR utilizes only a small angular diversity, spectral estimation [53] [54], instead of back projection [55] [56] [57], is sufficient for solving TomoSAR invesion in cases where the range migration resulting from different viewing angles is significantly smaller than the range resolution [33] [58]. Hence, numerous spectral estimation methods were developed and implemented, including beamforming [59] [35], singular value decomposition (SVD) [34] [58] [36], adaptive beamforming (Capon) [60], multiple signal classification (MUSIC) [61] [62], M-RELAX [63], non-linear least squares (NLS) [64] so on.

Generally, different methods can be employed according to the applications. In this thesis, our main focus is on urban reconstruction and mapping, specifically addressing the task of resolving discrete scatterers with or without motion in urban areas. To achieve this, we recommend utilizing single-looking methods as they allow for maintaining the azimuth-range resolution, which is crucial for obsering the inherent scale of urban infrastructures. However, it is important to acknowledge that single-looking methods have their limitations, including either resolution constraints or high computational costs. Though methods such as SVD and M-RELAX are computationally efficient, they may fall short in accurately modeling and interpreting complex urban environments. These methods struggle to distinguish scatterers that are spaced closer than the Rayleigh resolution, which typically falls within the range of tens of meters for TerraSAR-X data used in this thesis. Moreover, SVD and M-RELAX are susceptible to the sidelobe problem, leading to inaccurate elevation estimates. On the other hand, NLS is a high-resolution algorithm known for its excellent performance. However, it necessitates conducting a multi-dimensional search, which is computationally intensive and impractical for large-scale processing.

3.1.2 State of the art based on compressive sensing

As investigated in [12], for urban scenarios, the echoed signal in very high-resolution (VHR) data is primarily composed of the following contributions:

- The dominant signal contributions in VHR SAR data come from strong returns produced by metallic structures, as well as specular and dihedral or trihedral reflections. These points, which are also utilized in PSI, can be dense in VHR SAR data.
- Weak echos come from horizontal or vertical rough surfaces, such as roads and building walls. These scattering responses are typically in smaller extent compared to the Rayleigh resolution. Consequently, they can be treated as discrete scatterers in the elevation direction, resembling delta functions.
- Reflections from volumetric scatterers, such as vegetation, contribute to a continuous-signal background in the elevation direction. However, these scatterers often exhibit temporal decorrelation, meaning that their response varies over time. As a result, their contribution is considered as noise rather than a coherent signal.

The layover phenomena in urban areas are primarily caused by the roof, facade, and ground, resulting in a limited number of scatterers overlaid along the elevation direction within each resolution unit. To exploit the sparsity of the elevation signal in the object domain, it is recommended to use compressive sensing (CS)-based sparse reconstruction algorithms [9] [10] [11]. Although the ideal solution of the reflectivity profile $\boldsymbol{\gamma}$ can be obtained by solving Eq. (2.5) with L_0 norm minimization, this is a computationally challenging problem since it is NP-hard. Fortunately, the reflectivity profile $\boldsymbol{\gamma}$ is sufficiently sparse, allowing us to approximate the L_0 norm minimization using L_1 norm minimization. Therefore, $\boldsymbol{\gamma}$ can be estimated using L_1 norm regularized least squares, i.e. minimizing the L_1-L_2 mix norm as follows:

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \left\{ \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \right\}, \tag{3.1}$$

where λ (not the wavelength) is a regularization parameter balancing the sparsity and data-fitting terms.

Solving Eq. (3.1) provides a sparse estimate of $\boldsymbol{\gamma}$, but it is susceptible to spike-like artifacts [65] due to the violation of the restricted isometry property (RIP) and incoherence properties required by compressive sensing (CS). To mitigate these artifacts, further model order selection [13] is necessary. Model order selection aims to balance the model fit and complexity and can be formulated as a penalized likelihood problem, given by:

$$\hat{K} = \arg\min_{K} \{ -2 \ln p(\mathbf{g} \mid \hat{\boldsymbol{\theta}}(K), K) + 2C(K) \} \tag{3.2}$$

where $p(\mathbf{g} \mid \hat{\boldsymbol{\theta}}(K)$ is the likelihood describing the model fit and $C(K)$ penalizes the model complexity to prevent over-fitting of the data. K indicates the sparsity, i.e. the number of significant scatterers. The estimate vector of the unknown parameters, such as amplitudes, phases and elevations for the K scatterers, is denoted as $\hat{\boldsymbol{\theta}}(K)$. Assuming that the measurement noise is stationary white Gaussian noise with covariance matrix $\mathbf{C}_{\varepsilon\varepsilon} = \sigma_\varepsilon^{-2}\mathbf{I}$, we can simplify the log-likelihood term as follows:

$$-2 \ln p(\mathbf{g} \mid \hat{\boldsymbol{\theta}}(K), K) = \sigma_\varepsilon^{-2} \|\mathbf{g} - \mathbf{R}\hat{\gamma}\|_2^2 \tag{3.3}$$

As for the penalty term, there are many types in the literature, such as Bayesian Information Criterion (BIC) [66], Akaike Information Criterion (AIC) [67] and Minimum Description Length (MDL). In this thesis, we choose to utilize BIC as the penalty term to account for the model complexity:

$$C(K) = 1.5K \ln N \tag{3.4}$$

Accordingly, model order selection can be formulated as solving the following constraint minimization problem:

$$\hat{K} = \arg\min_{K} \{\sigma_{\varepsilon}^{-2} \|\mathbf{g} - \mathbf{R}\hat{\gamma}\|_2^2 + 3K \ln N\} \tag{3.5}$$

The solution to this minimization problem provides a clean estimate of the reflectivity profile by appropriately estimating the sparsity level $\hat{K}$.

Once the sparsity level K is determined, we can construct an over-determined mapping matrix $\mathbf{R}_s$ of size $N \times \hat{K}$ by selecting the corresponding columns to the $\hat{K}$ scatterers from the original steering matrix $\mathbf{R}$. Then, the final sparse reflectivity profile $\boldsymbol{\gamma}_s$ can be estimated using standard least squares as follows:

$$\hat{\gamma}_s = \left(\mathbf{R}_s^H \mathbf{R}_s\right)^{-1} \mathbf{R}_s^H \mathbf{g} \tag{3.6}$$

3.2 Data-driven algorithms for TomoSAR inversion

CS-based algorithms offer significant advantages in achieving super-resolution in elevation while preserving the full azimuth-range resolution. These algorithms enable substantial improvement in elevation estimation accuracy, approaching the Cramér-Rao Lower Bound (CRLB). Additionally, CS-based methods exhibit robustness against unmodeled phase errors, enhancing their performance and reliability in solving TomoSAR inversion. However, CS-based TomoSAR algorithms are usually impeded by their high computational costs since the L_1-L_2 mix norm minimization in Eq. (3.1) cannot be solved analytically and computationally expensive iterative solvers are required. Therefore, it renders them impractical for large-scale processing in real-world scenarios.

In this section, we delve into innovative data-driven algorithms. By leveraging the inherent patterns and structures present in the TomoSAR data, the data-driven algorithms are designed to lessen the computational load and enhance the efficiency of processing large-scale datasets.

3.2.1 CAESAR: Component extraction and selection SAR

Reducing the computational cost of multi-dimensional optimization is often a challenge. One common strategy to address this issue is by decomposing the optimization problem into several sub-problems. Each of these sub-problems can then be optimized independently. In this vein, an algorithm called CAESAR (Component extrAction and sElection SAR) was introduced in [68] [69] as a data-driven approach for TomoSAR processing. CAESAR utilizes principle component analysis (PCA) to separate the contributions from different scatterers, thereby reducing the computational complexity.

For a K-scatterer profile, the TomoSAR imaging model Eq. (2.5) can be simplified to:

$$\mathbf{g} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_K \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_K \end{bmatrix} + \boldsymbol{\varepsilon} \tag{3.7}$$

where r_k represents the column steering vector that corresponds to the k^{th} scatterer, while γ_k is the complex-valued amplitude of the k^{th} scatterer. Given these definitions, we can express the covariance matrix of the observation $\mathbf{g}$ in the following manner:

$$\begin{aligned} \mathbf{C_{gg}} &= \mathrm{E}\left\{\mathbf{R}\boldsymbol{\gamma}\boldsymbol{\gamma}^H\mathbf{R}^H + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H\right\} \\ &= \mathbf{R}\mathbf{E}\left\{\boldsymbol{\gamma}\boldsymbol{\gamma}^H\right\}\mathbf{R}^H + \mathrm{E}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H\right\}, \end{aligned} \tag{3.8}$$

where the operation $(\cdot)^H$ denotes the conjugate transpose. Under the assumption that the reflectivity profile $\boldsymbol{\gamma}$ is uncorrelated, $\mathbf{E}(\boldsymbol{\gamma}\boldsymbol{\gamma}^H)$ results in a diagonal matrix that represents the expected intensity of individual scatterers. Consequently, we can express the covariance matrix of the observation in a simplified form as follows:

$$\mathbf{C_{gg}} = \sum_{k=1}^{K} \sigma_k^2 \mathbf{r}_k \mathbf{r}_k^H + \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}, \tag{3.9}$$

where σ_k denotes the expected intensity of the k^{th} scatterer, while $\sigma_{\boldsymbol{\varepsilon}}^2\mathbf{I}$ represents the covariance matrix of the noise. For the sake of generality, we may consider the steering vectors $\mathbf{r}_k$ to be normalized.

CAESAR performs PCA on the observation covariance matrix via eigenvalue decomposition (EVD):

$$\mathbf{C_{gg}} = \mathbf{UDU}^H \tag{3.10}$$

where $\mathbf{U}$ and $\mathbf{D}$ denote eigenvectors and the diagonal matrix, respectively. Despite the fact that it is challenging to attribute exact meaning to the eigenvectors, CAESAR approximates the eigenvectors as the steering vectors to the individual scatterers, assuming a strong orthogonality condition among all the scatterers. As a result, the steering vector $\mathbf{r}_k$ can be estimated by:

$$\mathbf{r}_k = \mathbf{U}_i \odot |\mathbf{U}_i|^{-1} \tag{3.11}$$

where $\odot$ is the elementwise product. After estimating the steering vectors, we can estimate the elevation of individual scatterers by executing PSI on each steering vector.

However, achieving strong orthogonality among all scatterers is often impractical. The following are three prevalent factors causing low orthogonality in multi-baseline TomoSAR:

- Insufficient number of stacked SAR acquisitions,
- Close spacing of scatterers, implying that the distance between two scatterers is less than the Rayleigh resolution,
- severe interference due to similar amplitude and phase among the scatterers.

In conditions of low orthogonality, CAESAR encounters phase bias in the extracted steering vectors for individual scatterers, leading to a degradation of performance.

3.2.2 Nonlinear blind source separation

In a more general context, the approach of employing PCA to decompose layover scatterers can be classified as a form of Blind Source Separation (BSS). BSS aims to separate contributions from individual sources without requiring prior knowledge of a mixing matrix. Recognizing the limitation of CAESAR under low orthogonality conditions, we discuss a nonlinear BSS algorithm proposed in [4] to mitigate the phase bias in this section.

To enhance orthogonality, a basic approach involves increasing the dimensionality of the data. However, this strategy is not practical for TomoSAR applications since it is not feasible to increase the number of stacked SAR images. An alternative proposed by [4] involves the use of kernel PCA (KPCA). This method artificially increase the data dimension by projecting the data into a higher-dimensional space using a nonlinear transformation, often referred to as the kernel trick [70]. Subsequently, PCA is carried out in this transformed higher-dimensional space.

Mathematically, the nonlinear transformation can be expressed as follows:

$$\Psi : \mathbb{C}^N \to F, \mathbf{c} \to \Psi(\mathbf{c}), \tag{3.12}$$

where F is the transformed vector space, which can possess an arbitrary dimension, and $\mathbf{c}$ denotes the columns of the observation covariance matrix $\mathbf{C_{gg}}$. Let us denote the transformed covariance matrix in higher dimensional space as $\mathbf{\Psi_c}$, which can be expressed as $\mathbf{\Psi_c} = [\mathbf{\Psi_{c_1}}, \mathbf{\Psi_{c_2}}, \cdots, \mathbf{\Psi_{c_N}}]$. Subsequently, the execution of PCA essentially involves finding the EVD of the transformed covariance matrix $\mathbf{\Psi_c}$. This process can be formally defined as:

$$\mathbf{C}_{\Psi\Psi} = \mathbf{\Psi_c}\mathbf{\Psi_c}^H = \mathbf{U}_{\Psi\Psi}\mathbf{D}_{\Psi\Psi}\mathbf{U}_{\Psi\Psi}^H. \tag{3.13}$$

To clarify, the EVD on $\mathbf{\Psi_c}$ is evaluated indirectly through the kernel trick, rather than being computed explicitly. By employing the kernel trick, the higher-dimensional transformed space can be represented by a certain kernel function of the original data space, i.e.,

$$\kappa\left(\mathbf{c}_i, \mathbf{c}_j\right) = \Psi\left(\mathbf{c}_i\right)^H \Psi\left(\mathbf{c}_j\right), \tag{3.14}$$

where $\kappa(\cdot)$ denotes a kernel function, and $\mathbf{c}_i$ is the i^{th} column of the observation covariance matrix $\mathbf{C_{gg}}$. Then, a kernel matrix $\mathbf{K} \in \mathbb{C}^{N\times N}$ of the transformed data can be defined as:

$$\mathbf{K} = \mathbf{\Psi}_\mathbf{c}^H \mathbf{\Psi_c}, \tag{3.15}$$

where each element of the matrix can be readily calculated using Eq. (3.14). Following this, the EVD of the kernel matrix can be derived as follows:

$$\mathbf{KV} = \mathbf{VS}, \tag{3.16}$$

where $\mathbf{V}$ denotes the eigenvectors and $\mathbf{S}$ represents the diagonal matrix of eigenvalues. When we substitute Eq. (3.15) into Eq. (3.16) and multiply both sides of Eq. (3.16) by $\mathbf{\Psi_c}$, we obtain:

$$\begin{aligned} \mathbf{\Psi}_\mathrm{c}\mathbf{\Psi}_\mathrm{c}^H\mathbf{\Psi}_\mathrm{c}\mathbf{V} &= \mathbf{\Psi}_\mathrm{c}\mathbf{VS} \\ &\Downarrow \\ \mathbf{C}_{\Psi\Psi}(\mathbf{\Psi}_\mathrm{c}\mathbf{V}) &= (\mathbf{\Psi}_\mathrm{c}\mathbf{V})\mathbf{S}. \end{aligned} \tag{3.17}$$

This equation implies that $\mathbf{\Psi}_\mathrm{c}\mathbf{V}$ and $\mathbf{S}$ represent the eigenvectors and eigenvalues of the transformed covariance matrix $\mathbf{C}_{\Psi\Psi}$, respectively. By choosing an appropriate kernel function, $\mathbf{\Psi}_\mathrm{c}\mathbf{V}$ should delineate the space spanned by individual scatterers. Great details about how to choose a kernel function can be found in [4]. Consequently, the data projected onto these eigenvectors would correspond to the steering vectors. By extracting and normalizing the first K eigenvectors of the transformed covariance matrix $\mathbf{C}_{\Psi\Psi}$, we can construct the orthogonal projection basis in the transformed higher-dimensional space, as follows:

$$\mathbf{\Xi} = \mathbf{\Psi_c}\mathbf{V}_{1\sim K}\mathbf{S}_{1\sim K}^{-1/2}, \tag{3.18}$$

where $\mathbf{V}_{1\sim K}$ and $\mathbf{S}_{1\sim K}^{-1/2}$ denote the first K columns of $\mathbf{V}$ and $\mathbf{S}$. The projected data can be obtained as follows:

$$\begin{aligned} \mathbf{Y} &= \mathbf{\Psi}_\mathbf{c}^H\mathbf{\Xi} = \mathbf{\Psi}_\mathbf{c}^H\mathbf{\Psi_c}\mathbf{V}_{1\sim K}\mathbf{S}_{1\sim K}^{-1/2} \\ &= \mathbf{KV}_{1\sim K}\mathbf{S}_{1\sim K}^{-1/2}. \end{aligned} \tag{3.19}$$

After acquiring the projected data $\mathbf{Y}$, the steering vectors of the K scatterers can be estimated by extracting the phase of the first K columns of $\mathbf{Y}$. Similar to CAESAR, standard PSI techniques such as the periodogram or integer least square methods are then utilized to estimate the elevation. It is noteworthy that this extraction process is conducted iteratively. At each iteration, only the dominant scattering contribution is extracted, and subsequently, it is demodulated from the observation covariance matrix. This process continues until no significant scattering is left or when the predefined number of scatterers is reached. Fig. 3.1 [4] presents the workflow of the nonlinear BSS method.

Like CAESAR, the KPCA method also blindly performs layover separation, eliminating the need for explicit inversion of the TomoSAR imaging model, thereby logarithmically

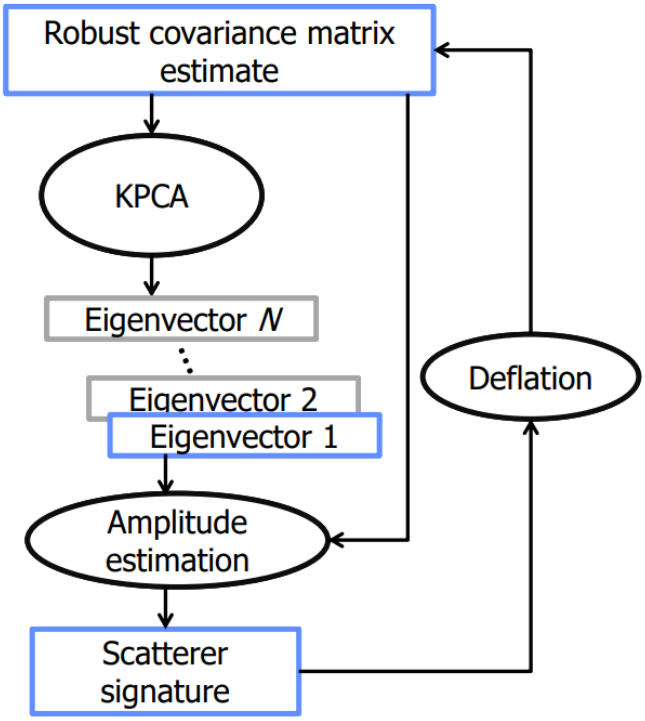


Figure 3.1: The flowchart of the proposed nonlinear BSS algorithm, source in [4].

reducing computational costs. In addition, it demonstrates more robustness and effectively addresses issues arising from low orthogonality. However, its practical application is somewhat limited, as the algorithm is based on the assumption of a fully-coherent signal model that does not align well with real-world scenarios. Furthermore, for fully automatic processing, the KPCA method requires the integration of GIS building footprints or NL-SAR [71], both of which could potentially increase computational costs.

3.2.3 Multi-layer Perceptron

Deep learning, as an emerging data-driven algorithm, has attracted significant interest among researchers and has been extensively studied in recent years. Its broad applications span diverse fields, including image processing [72] [73] [74]. Among the various deep learning models, the Multi-layer Perceptron (MLP), which is a basic neural network model, has also gained popularity due to its learning capabilities and expressiveness. Its potential has been recognized in the field of TomoSAR inversion as well. As a prime example, MLP-based classifiers have been successfully implemented for solving TomoSAR inversion [75], by transforming TomoSAR inversion into a classification. Specifically, Budillon et al. [75] reinterpreted each discretized elevation grid as a unique class. This innovative reconfiguration allowed the elevation estimation to be reformulated as a classification task, aiming to predict the class of the elevation grid housing the scatterer.

The architecture of a representative MLP model is depicted in Fig. 3.2. This model comprises three fully connected (FC) layers followed by a softmax layer. Each FC layer

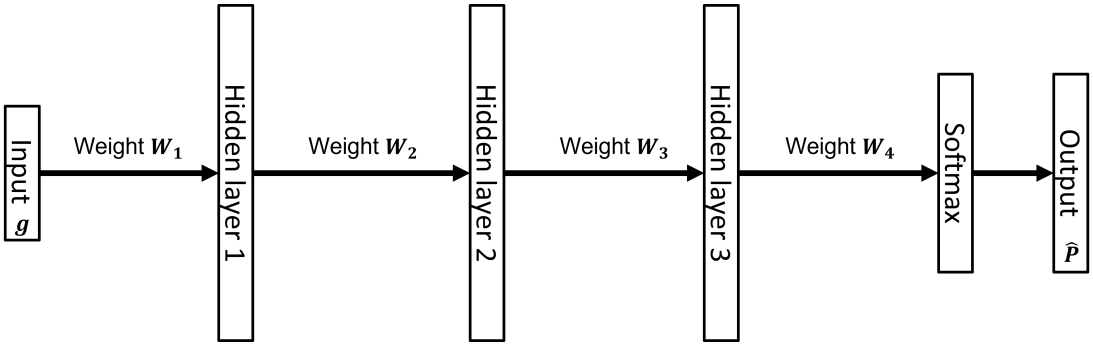


Figure 3.2: The architecure of MLP.

executes the following operation:

$$\mathrm{z}_j^{(l)} = f(\sum_{k=1}^{K} W_{jk}^{(l)} \mathrm{z}_{jk}^{(l-1)} + \mathrm{b}_k^{(l)}) \tag{3.20}$$

where $\mathbf{W}^l$ and $\mathbf{b}^l$ represent the weights and biases to be learned at the current layer. For $l = 1$, $\mathrm{z}^{\mathbf{0}}$ is assumed as the input $\mathbf{g}$. $f(\cdot)$ stands for the activation function that introduces non-linearity into the model. The activation function commonly chosen for this operation is the Rectified Linear Unit (ReLU), which is defined as:

$$f(x) = \max(x, 0) \tag{3.21}$$

Following the fully connected layers, a layer implementing the softmax function is employed to predict the output. This layer produces a probability distribution over the set of possible output classes, making it particularly well-suited for classification tasks. The softmax function is defined as follows:

$$\sigma(\hat{\mathbf{P}})_i = \frac{e^{h_i}}{\sum_{k=1}^{K} e^{h_j}} \tag{3.22}$$

where $\mathbf{h}$ is the output of the FC layers.

The MLP-based classifier is trained in a supervised manner. For a given ground truth reflectivity profile $\boldsymbol{\gamma}$, the SAR observation $\mathbf{g}$ is generated using Eq. (2.5) as the input. Additionally, for each input $\mathbf{g}$, a label vector $\mathbf{P}$, having the same dimensions as $\boldsymbol{\gamma}$, is assigned. All elements in the label $\mathbf{P}$ are set to zero, except for one element, which is given a unitary value and placed at the position corresponding to the scatterer.

To learn the weights and biases in the fully connected layers, a cost function is minimized. Here, we use the cross-entropy loss function, which is a popular choice for classification tasks as it measures the performance of a classification model whose output is a probability value between 0 and 1. It is formally defined as follows:

$$\underset{\boldsymbol{\Psi}}{\text{minimize}}\ \mathcal{L}(\boldsymbol{\Psi}) = -\frac{1}{L}\sum_{j}^{L}[P_j \ln \hat{P}_j + (1 - P_j) \ln\Big(1 - \hat{P}_j\Big)] \tag{3.23}$$

where $\boldsymbol{\Psi}$ denotes the set of all free parameters that need to be learned, encompassing all weights in the FC layers as well as the biases.

After the MLP-based classifier has been sufficiently trained, the elevation of the scatterer can be estimated in subsequent inferences. This is accomplished by locating the position

of the maximum value in the predicted vector $\hat{\mathbf{P}}$. Specifically, the location of the scatterer corresponds to the index of the maximum value in $\hat{\mathbf{P}}$, i.e., argmax($\hat{\mathbf{P}}$).

The MLP-based classifier represents an initial attempt to apply deep learning to tackle the TomoSAR inversion problem. Despite its innovative application, the capabilities of the MLP-based classifier are limited to the detection of a single scatterer because of the nature of its problem formulation. Specifically, the MLP-based classifier is not applicable in realistic 3-D SAR imaging scenarios because it does not have the ability to resolve the issue of layover separation, a critical challenge in real-world applications. This limitation significantly hampers the practical utility of the MLP-based classifier, indicating a need for more sophisticated deep learning methods that can effectively deal with multiple scatterers in the scene.

4 Summary of the work

The objectives of this thesis mentioned in Chapter 1 are addressed in the attached peer-reviewed journal articles. This chapter gives a brief summary of of the attached articles including the following main contributions:

- Development of a novel end-to-end deep learning-based algorithm for efficient and accurate super-resolving TomoSAR inversion.
- Development of an innovative gated unit to incorporate historical information so that information loss caused by sparsity-promoting shrinkage steps can be mitigated.
- Development of a hybrid approach using analytic weight determination for efficient and accurate multi-component D-TomoSAR processing.
- A large-scale processing that showcases the efficiency and effectiveness of the proposed method in real-world application, highlighting the immense potential of the developed algorithms in future global urban mapping using TomoSAR.

4.1 γ-Net: A novel deep unfolded neural network for TomoSAR inversion

In this section, we present $\boldsymbol{\gamma}$-Net, a deep learning-based TomoSAR algorithm, notable for its impressive computational efficiency. To our knowledge, $\boldsymbol{\gamma}$-Net is the first algorithm that solely applies an end-to-end deep neural network principles for super-resolution TomoSAR inversion. Our rigorous and systematic performance evaluations reveal that $\boldsymbol{\gamma}$-Net closely matches the state-of-the-art in terms of performance under nominal conditions. More impressively, it accomplishes this feat while significantly enhancing computational efficiency by two to three orders of magnitude. These findings underscore the viability and efficacy of incorporating deep learning methodologies into super-resolution TomoSAR inversion. Furthermore, they open new avenues for exploration in the domain of efficient, large-scale, and real-world TomoSAR imaging applications.

4.1.1 Iterative shrinkage thresholding algorithm for TomoSAR

As discussed in Section 3.1.2, the discrete reflectivity profile $\boldsymbol{\gamma}$ is assumed to be compressible in urban scenarios. This means that only a few scatterers (usually between 0-4) are overlaid within the same resolution unit. Consequently, we can estimate $\boldsymbol{\gamma}$ within the framework of CS by solving the L_1-L_2 mixed norm minimization, as referred to in Eq. (3.1).

The Iterative Shrinkage-Thresholding Algorithm (ISTA) is commonly employed to iteratively solve the L_1-L_2 mixed norm minimization due to its simplicity. By applying ISTA

to TomoSAR inversion, we can estimate the reflectivity profile $\boldsymbol{\gamma}$ as follows:

$$\begin{aligned} \hat{\boldsymbol{\gamma}}^k &= \eta_{cv-st}(\hat{\boldsymbol{\gamma}}^{k-1} + \beta \mathbf{R}^H \mathbf{b}^{k-1}, \theta^k) \\ &\text{with} \quad \mathbf{b}^k = \mathbf{g} - \mathbf{R}\hat{\boldsymbol{\gamma}}^k, \end{aligned} \tag{4.1}$$

where $\hat{\boldsymbol{\gamma}}_0 = \mathbf{0}$, β is the stepsize, η_{cv-st} is the complex-valued 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 soft-thresholding function η_{cv-st} is defined by

$$\eta_{cv-st}(\hat{\boldsymbol{\gamma}}^k, \theta^k) = \begin{cases} \frac{\hat{\boldsymbol{\gamma}}^k}{|\hat{\boldsymbol{\gamma}}^k|} \max(|\hat{\boldsymbol{\gamma}}^k| - \theta^k, 0) & |\hat{\boldsymbol{\gamma}}^k| \neq 0 \\ 0 & \text{else} \end{cases} . \tag{4.2}$$

During each iteration of the ISTA process, the estimate undergoes a two-step optimization process. First, gradient descent is applied, optimizing the estimate by adjusting it in the direction of steepest descent of the cost function. This iterative adjustment of the estimate brings it closer to the minimum of the function. Following this, the soft-thresholding function is applied. This operation sets coefficients whose absolute values are below a certain threshold to zero, effectively removing elements of the estimate with small magnitude, thus inducing sparsity in the final estimate.

4.1.2 Network architecture: improved ISTA network built by unfolding technique

By substituting $\mathbf{W}_1 = \beta \mathbf{R}^H$ and $\mathbf{W}_2 = \mathbf{I} - \beta \mathbf{R}^{\mathrm{H}} \mathbf{R}$ into Eq. (4.1), we can rewrite Eq. (4.1) as the following form:

$$\hat{\boldsymbol{\gamma}}^k = \eta_{cv-st}\left(\mathbf{W}_1 \mathbf{g} + \mathbf{W}_2 \hat{\boldsymbol{\gamma}}^{k-1}, \theta^k\right), \tag{4.3}$$

By considering the soft-thresholding function in Eq. (4.3) as an activation function, Eq. (4.3) intriguingly resembles the structure of a recurrent neural network (RNN). In this perspective, we can interpret ISTA as an RNN, which is depicted in Fig. 4.1. Inspired

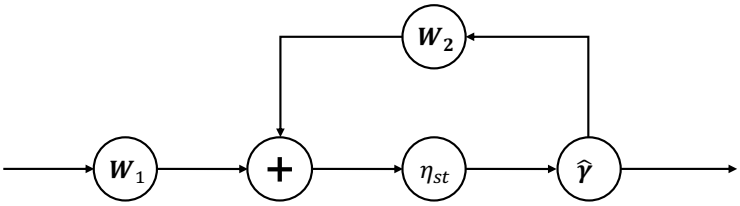


Figure 4.1: RNN structure of ISTA by viewing an iteration of ISTA as a layer of the RNN.

by the underlying link between ISTA and RNN, a learning-based model named Learned ISTA (LISTA) was put forward in [48]. Fig. 4.2 illustrates the learning architecture of a K-layer LISTA. It 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 lies in how their respective parameters are determined. In LISTA, the weight matrices $\mathbf{W}_1^k$, $\mathbf{W}_2^k$, and the threshold θ_i in each layer are not predefined. Instead, these parameters are learned from the training data, enabling the LISTA network to adapt to the specific characteristics of the data. The loss function for LISTA, evaluated over the

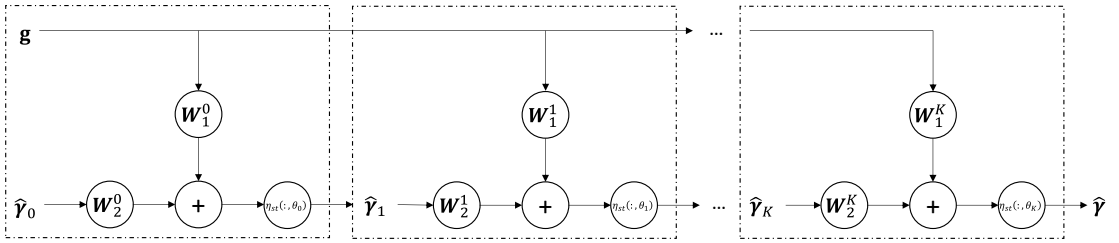


Figure 4.2: Unfolded LISTA architecture. A K-layer LISTA unrolls the RNN and truncates it into K iterations, thereby creating a side-connected feedforward neural network.

training data $\{(\mathbf{g}_i, \boldsymbol{\gamma}_i)\}_{i=1}^{T}$, is defined as the mean square error (MSE) loss, as described below.

$$\underset{\boldsymbol{\Psi}}{\text{minimize}}\ \mathcal{L}(\boldsymbol{\Psi}) = \frac{1}{T}\sum_{i=1}^{T} ||\hat{\boldsymbol{\gamma}}(\boldsymbol{\Psi}, \mathbf{g}) - \boldsymbol{\gamma}||_2^2, \tag{4.4}$$

where T represents the number of samples within the training dataset, while $\boldsymbol{\Psi} = \mathbf{W}_1, \mathbf{W}_2, \boldsymbol{\theta}$ constitutes the set of parameters that the network aims to learn. As demonstrated in numerous recent studies [48, 49, 50, 51, 52], LISTA has the ability to attain a comparable estimation accuracy as ISTA, but with two to three orders of magnitude fewer iterations.

To apply LISTA to the problem of TomoSAR inversion, we extend LISTA to operate in the complex-valued domain, resulting in the complex-valued LISTA (CV-LISTA) [76]. The architecture of CV-LISTA is identical to that of LISTA, with one key difference: in CV-LISTA, each neuron possesses two channels, corresponding to the real and imaginary parts of a complex number. The modifications to Eq. (4.3) are as follows:

$$\tilde{\boldsymbol{\gamma}}^k = \eta_{cv-st}\left\{\widetilde{\mathbf{W}}_1^k \tilde{\mathbf{g}} + \widetilde{\mathbf{W}}_2^k \tilde{\boldsymbol{\gamma}}^{k-1}, \theta^k\right\} \tag{4.5}$$

where

$$\begin{aligned}
\widetilde{\mathbf{W}}_j^k &= \begin{bmatrix} \mathrm{Re}(\mathbf{W}_j^k) & -\mathrm{Im}(\mathbf{W}_j^k) \\ \mathrm{Im}(\mathbf{W}_j^k) & \mathrm{Re}(\mathbf{W}_j^k) \end{bmatrix}, \\
\tilde{\mathbf{g}} &= \begin{bmatrix} \mathrm{Re}(\mathbf{g}) \\ \mathrm{Im}(\mathbf{g}) \end{bmatrix}, \\
\tilde{\boldsymbol{\gamma}} &= \begin{bmatrix} \mathrm{Re}(\hat{\boldsymbol{\gamma}}) \\ \mathrm{Im}(\hat{\boldsymbol{\gamma}}) \end{bmatrix}
\end{aligned} \tag{4.6}$$

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

In $\boldsymbol{\gamma}$-Net [77], we propose several innovative improvements to enhance the performance of CV-LISTA, primarily in three key aspects.

- We note that the $\mathbf{W}_1^k$ and $\mathbf{W}_2^k$ matrices in the original CV-LISTA formulation are closely correlated. To reduce the amount of redundant trainable parameters in $\boldsymbol{\gamma}$-Net, we introduce a weight coupling structure, as suggested in [78].
- We apply an acceleration technique, known as *support selection*, initially developed for the Least Absolute Shrinkage and Selection Operator (LASSO) method. This technique accelerates the convergence of $\boldsymbol{\gamma}$-Net, significantly improving the speed of its operation.

- We replace the conventional soft-thresholding function with a piecewise linear function. This change addresses an issue in the original CV-LISTA model, where the traditional soft-thresholding function could lead to information loss, resulting in a large reconstruction error and a decreased convergence rate.

These novel improvements will be discussed in more detail in the following sections.

Weight coupling

To move away from the pure "black-box" training method for $\boldsymbol{\gamma}$-Net, we utilize the dependencies among the trainable weights to simplify the CV-LISTA model. This results in our proposed $\boldsymbol{\gamma}$-Net. According to Chen et al. [78], the weights to be learned in each layer $(\mathbf{W}_1^k, \mathbf{W}_2^k)_{k=0}^K$ asymptotically adhere to the following partial weight coupling structure:

$$\mathbf{W}_2^k = \mathbf{I} - \mathbf{W}_1^k \mathbf{R} \ . \tag{4.7}$$

By employing the partial weight coupling structure, we can simplify the k^{th} layer of $\boldsymbol{\gamma}$-Net to:

$$\begin{aligned} \tilde{\boldsymbol{\gamma}}^k &= \eta_{cv-st}\{\tilde{\boldsymbol{\gamma}}^{k-1} + \widetilde{\mathbf{W}}^k(\tilde{\mathbf{g}} - \widetilde{\mathbf{R}}\tilde{\boldsymbol{\gamma}}^{k-1}), \theta^k\} \\ \text{with } \ \widetilde{\mathbf{R}} &= \begin{bmatrix} \mathrm{Re}(\mathbf{R}) & -\mathrm{Im}(\mathbf{R}) \\ \mathrm{Im}(\mathbf{R}) & \mathrm{Re}(\mathbf{R}) \end{bmatrix} \end{aligned} \tag{4.8}$$

where $(\mathrm{Re}(\mathbf{W^k}), \mathrm{Im}(\mathbf{W^k}), \theta^k)$ are the parameters to be learned in the k^{th} layer, and the trainable weight $\mathbf{W^k}$ is initialized using the system measurement matrix $\mathbf{R}$ with $\mathbf{W^k} = \beta\mathbf{R}^H$. The coupled structure contributes to eliminating the number of free parameters to be trained, thus accelerating the training procedure significantly. Although theoretically, the condition in equation (4.7) can only be satisfied for very deep layers, comprehensive simulations as described in [78] show that applying the partial weight coupling structure to every layer does not negatively impact the empirical performance.

Support selection

Beyond the use of the weight coupling structure, we also introduce a unique thresholding scheme in $\boldsymbol{\gamma}$-Net called *support selection*, inspired by the "kicking" strategy in linearized Bregman iteration [79]. This approach involves selecting a certain percentage of entries with the largest magnitudes at each layer of $\boldsymbol{\gamma}$-Net prior to the shrinkage step. These selected entries are deemed as "true support" and are directly fed into the subsequent layer, bypassing the shrinkage step. The remaining entries undergo the standard shrinkage step. Assuming that ρ^k percentage of entries are trusted in the k^{th} layer, the support selection can be formally defined as:

$$\eta_{ss_{\theta^k}^{\rho^k}}(\tilde{\boldsymbol{\gamma}}^k) = \begin{cases} \tilde{\boldsymbol{\gamma}}^k & k \in \mathcal{S}^{\rho^k}(\tilde{\boldsymbol{\gamma}}) \\ \eta_{cv-st}(\tilde{\boldsymbol{\gamma}}^k, \theta_i) & k \notin \mathcal{S}^{\rho^k}(\tilde{\boldsymbol{\gamma}}) \end{cases} , \tag{4.9}$$

where $\mathcal{S}^{\rho^k}(\boldsymbol{\gamma})$ contains the entries with the ρ^k largest magnitudes. It is worth mentioning that the percentage ρ^k is a hyperparameter that requires manual tuning. When we apply the support selection to $\boldsymbol{\gamma}$-Net, then (4.8) is modified as:

$$\tilde{\boldsymbol{\gamma}}^k = \eta_{ss_{\theta^k}^{\rho^k}}\{\tilde{\boldsymbol{\gamma}}^{k-1} + \widetilde{\mathbf{W}}^k(\tilde{\mathbf{g}} - \widetilde{\mathbf{R}}\tilde{\boldsymbol{\gamma}}^{k-1}), \theta^k\} \tag{4.10}$$

Simulations in [78] affirm that the incorporation of support selection improves the convergence rate, both theoretically and empirically. Additionally, it contributes to a reduction in the recovery error, subsequently enhancing the accuracy of estimation.

Piecewise linear thresholding function

The conventional soft-thresholding function simply prunes elements of small magnitude to zero, which inevitably leads to a loss of potentially significant information. To conserve as much of this significant information as possible while still executing the shrinkage step, we replace the soft-thresholding function in $\boldsymbol{\gamma}$-Net with the piecewise linear function $\eta_{pwl}(\tilde{\boldsymbol{\gamma}}, \boldsymbol{\theta})$, as defined below:

$$\eta_{pwl}(\hat{\boldsymbol{\gamma}}) = \begin{cases} \theta_3 \hat{\boldsymbol{\gamma}}, & |\hat{\boldsymbol{\gamma}}| \leq \theta_1 \\ \frac{\hat{\boldsymbol{\gamma}}}{|\hat{\boldsymbol{\gamma}}|}[\theta_4(|\hat{\boldsymbol{\gamma}}| - \theta_1) + \theta_3\theta_1], & \theta_1 < |\hat{\boldsymbol{\gamma}}| \leq \theta_2 \\ \frac{\hat{\boldsymbol{\gamma}}}{|\hat{\boldsymbol{\gamma}}|}[\theta_5(|\hat{\boldsymbol{\gamma}}| - \theta_2) + \theta_4(\theta_2 - \theta_1) + \theta_3\theta_1], & |\hat{\boldsymbol{\gamma}}| > \theta_2 \end{cases} . \tag{4.11}$$

Figure 4.3 provides a comparison between the two functions. It can be observed that, instead of completely pruning elements of small magnitude, the piecewise linear function merely scales them down. This approach effectively mitigates information loss. However, it does result in the final output of $\boldsymbol{\gamma}$-Net not being strictly sparse. Most elements of the final output are not driven to absolute zero, but rather to some extremely small values. Consequently, an additional post-processing step becomes necessary to clear elements with minute magnitudes when employing the piecewise linear function.

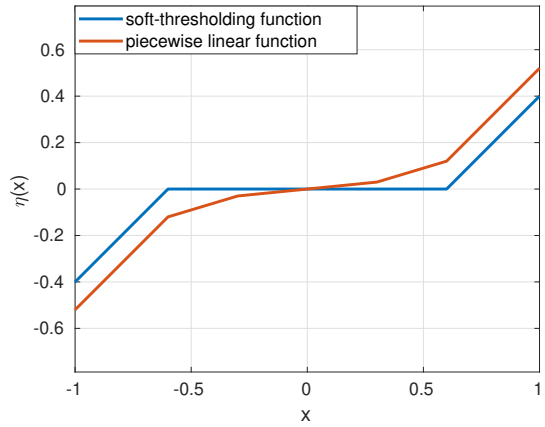


Figure 4.3: Comparison between the piecewise linear function and soft-thresholding function. Instead of pruning the elements with small magnitude, the piecewise linear function just further minifies them, thus possibly avoiding the information loss.

Figure 4.4 presents a comparison of the performance of $\boldsymbol{\gamma}$-Net in terms of the normalized mean square error (NMSE) under two shrinkage functions. It is important to note that the performance of $\boldsymbol{\gamma}$-Net, under different shrinkage functions, was assessed using a set of noise-free data. This was done to ensure that the results reflect the ideal performance.

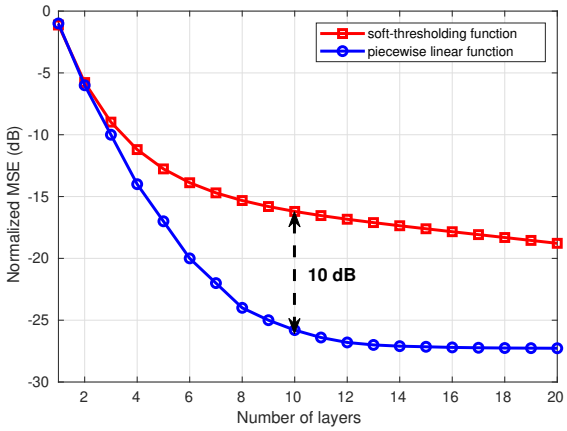


Figure 4.4: Performance of γ-Net using different shrinkage function. The piecewise linear function conduces to faster convergence and improves the estimation accuracy.

The NMSE is defined as follows:

$$\text{NMSE} = \frac{1}{T} \sum \frac{\|\hat{\gamma} - \gamma\|_2^2}{\|\gamma\|_2^2} \tag{4.12}$$

As evident from the figure, γ-Net, when utilizing the piecewise linear function, attains a lower NMSE with the same number of layers. This suggests that the piecewise linear function enhances the accuracy of estimations, or in other terms, it improves the rate of convergence. Specifically, it can be observed that γ-Net requires approximately only 12 layers to attain convergence when using the piecewise linear function. On the other hand, it is quite clear that significantly more layers are required when the traditional soft-thresholding function is employed. This implies that our modification not only enhances the estimation accuracy but also contributes to improving the computational efficiency.

Algorithm summary

After integrating the aforementioned improvements, we achieve the learning architecture of γ-Net, as illustrated in Fig. 4.5.

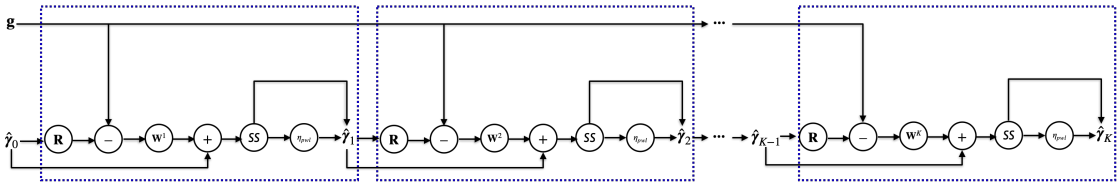


Figure 4.5: Illustration the learning architecture of a K-layer γ-Net. Each block indicates a layer of γ-Net. ***SS*** depicts support selection.

Through γ-Net, a preliminary sparse estimate of the reflectivity profile can be obtained. However, due to the violation of the Restricted Isometry Property (RIP) and incoherence [1], outliers exist in the estimated reflectivity profile of γ-Net. Consequently, further procedures are needed to perform model order selection [58] to precisely suppress the

Algorithm 1 Summary of TomoSAR processing using $\boldsymbol{\gamma}$-Net

Simulate training data
  Sampling the elevation extent
  Generate steering matrix $\mathbf{R}$ with
  $R_{nl} = \exp(-j2\pi\xi_n s_l)$, 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 $\{(\mathbf{g_i}, \boldsymbol{\gamma_i})\}_{i=1}^{T}$
  minimize $\underset{\boldsymbol{\Psi}}{\mathcal{L}(\boldsymbol{\Psi})} = \frac{1}{T}\sum_{i=1}^{T} ||\hat{\boldsymbol{\gamma}}(\boldsymbol{\Psi}, \mathbf{g}) - \boldsymbol{\gamma}||_2^2$
  where $\boldsymbol{\Psi} = [\mathrm{Re}(\mathbf{W}), \mathrm{Im}(\mathbf{W}), \boldsymbol{\theta}]$
for each pixel in the image: **do**
    Preliminary estimate via $\boldsymbol{\gamma}$-Net:
    $\boldsymbol{\gamma} = \boldsymbol{\gamma}$-Net$(\mathbf{g})$
    Model order selection to remove outliers:
    $\hat{P} = \underset{P}{\mathrm{argmin}} \left\{\sigma_\varepsilon^{-2}||\mathbf{g} - \mathbf{R}\hat{\gamma}||_2^2 + 1.5P\ln N\right\}$
    Determine the number of scatterers
    Final estimation of their elevation
end for

undesired outliers and estimate the number and location of scatterers, which are typical steps in TomoSAR.

The proposed super-resolving TomoSAR inversion algorithm combines $\boldsymbol{\gamma}$-Net, model order selection, and re-estimation. The fundamental workflow of the proposed algorithm is detailed in Algorithm 1. The model order selection is based on the Bayesian Information Criterion (BIC) [66], a widely recognized statistical tool for model selection and order estimation. This combination aims to provide an accurate, robust, and efficient approach to TomoSAR inversion.

4.1.3 Experiment and validation

To demonstrate the effectiveness and efficiency of $\boldsymbol{\gamma}$-Net, we conduct a series of experiments utilizing both simulated and real data. The experimental results obtained will offer valuable insights into the performance and potential advantages of $\boldsymbol{\gamma}$-Net in real-world scenarios.

Simulation

The simulation aimed to evaluate the theoretical performance of the proposed $\boldsymbol{\gamma}$-Net with respect to detection power and localization accuracy.

Initially, we compared $\boldsymbol{\gamma}$-Net to SVD-Wiener, a conventional non-superresolving algorithm, in the detection of a single scatterer. As demonstrated in Fig. 4.6, both $\boldsymbol{\gamma}$-Net and SVD-Wiener are capable of detecting the position of the single scatterer. However, $\boldsymbol{\gamma}$-Net reconstructs spectral lines as opposed to a sinc-like point response function, thus mitigating the sidelobe issue. Further statistical details are provided in Table 4.1. The

Cramér–Rao lower bound (CRLB), the estimates mean (μ), and standard deviation (σ) in Table 4.1 are normalized to the Rayleigh resolution. From the table, it is apparent that the proposed algorithm can detect almost all single scatterers at various SNRs. Additional statistics on the mean value μ and standard deviation σ of the estimation error indicate the high estimation accuracy of the proposed algorithm, with the bias nearing zero and the standard deviation approaching the CRLB. This demonstrates the superiority of $\boldsymbol{\gamma}$-Net in detecting and accurately localizing single scatterers even in noisy conditions.

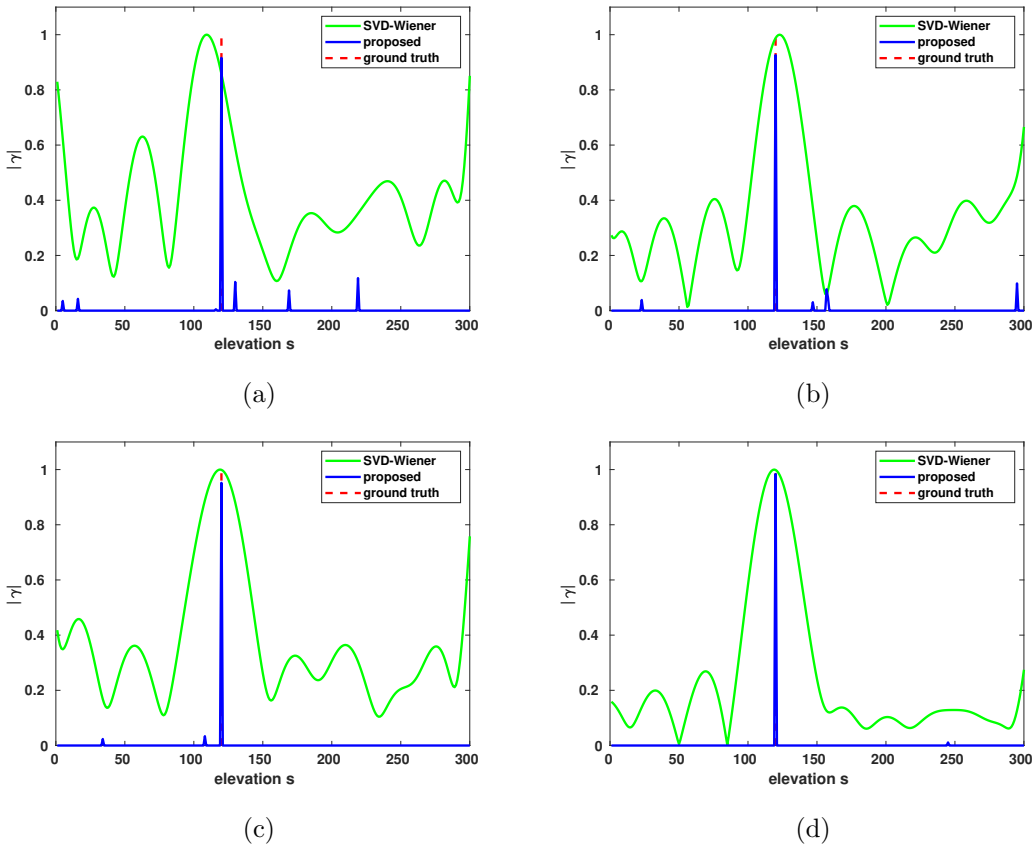


Figure 4.6: Estimated reflectivity profile of simulated data with single scatterer at different SNR. (a) $SNR = 0$dB, (b)$SNR = 3$dB, (c) $SNR = 6$dB, (d)$SNR = 10$dB.

In addition, we performed a well-known benchmark TomoSAR test [58] [12] to evaluate the super-resolution power, in which double scatterers with increasing elevation distance between the two layovered scatterers were simulated, in order to mimic a facade-ground interaction. Then, different algorithms were compared in double scatterers detection under two SNR levels, i.e., SNR$\in \{0, 6\}$ dB, which represent typical SNR levels in a high-resolution spaceborne SAR image. Fig. 4.7 demonstrates some examples of the estimated reflectivity profile at the normalized elevation distance $\alpha = [0.2, 0.5, 1.0]$.

From Fig. 4.7, it is clear that both $\boldsymbol{\gamma}$-Net and SVD-Wiener are capable of distinguishing the overlaid double scatterers in the non-superresolving case, i.e., the last column, when $\alpha = 1.0$. However, compared to SVD-Wiener, $\boldsymbol{\gamma}$-Net provides much higher elevation estimation accuracy. Furthermore, when we bring the double scatterers closer into the

Table 4.1: 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}

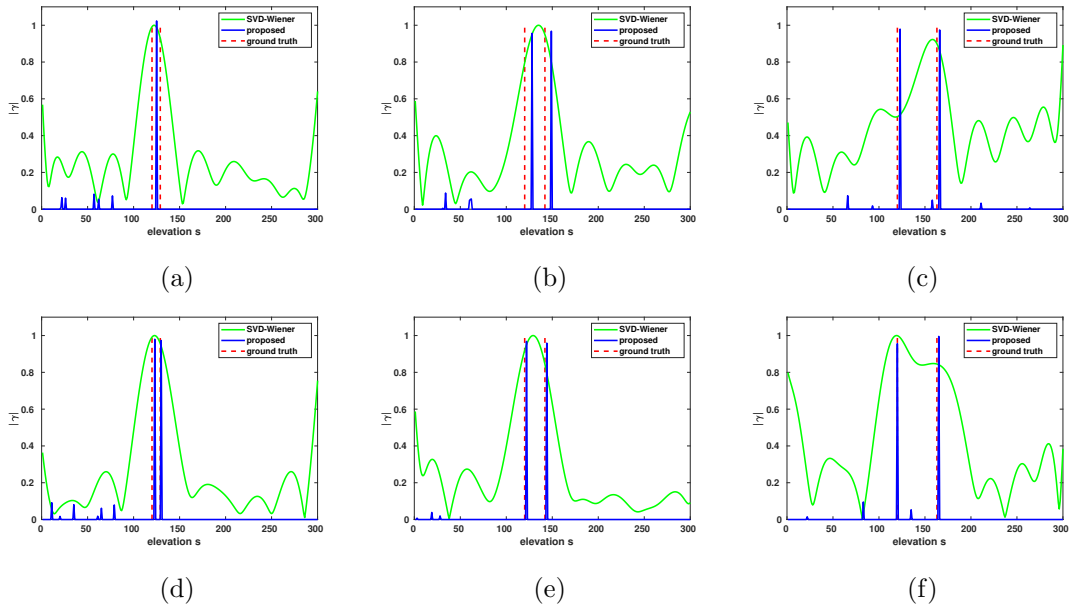
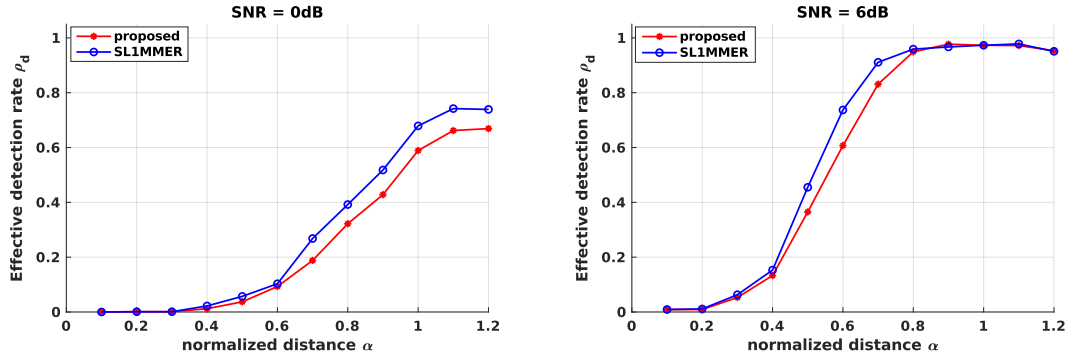


Figure 4.7: Estimated reflectivity profile of simulated data with overlaid double scatterers. From top to bottom, $SNR = 0, 6$ dB. From left to right, the normalized elevation distance $\alpha = 0.2, 0.5, 1$.

Rayleigh resolution, SVD-Wiener fails to separate them. In contrast, $\boldsymbol{\gamma}$-Net still manages to detect the double scatterers in most cases, exhibiting its super-resolution power.

Subsequently, we compared $\boldsymbol{\gamma}$-Net with the state-of-the-art SL1MMER algorithm [13] focusing on the effective detection rate, which offers a more intuitive comparison of super-resolution power. Fig. 4.8 compares the effective detection rate P_d of SL1MMER and $\boldsymbol{\gamma} - Net$ 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 $\boldsymbol{\gamma} - Net$ and SL1MMER, respectively. As we can see from Fig. 4.8, $\boldsymbol{\gamma} - Net$ has comparable super-resolution power as SL1MMER.

However, SL1MMER requires approximately 10 CPU hours for processing 0.2 million Monte Carlo trails, whereas $\boldsymbol{\gamma}$-Net takes only a few CPU minutes. Fig. 4.9 provides an intuitive view of the time consumption of the two methods for TomoSAR processing.



(a) (b)

Figure 4.8: 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 SL1MMER (dashed circle) with SNR = 0dB and 6dB, N = 25 and phase difference $\triangle\phi = 0$(worst case) under 0.2 million Monte Carlo trials.

Despite the fact that about 9 hours are required for training the model, this fixed cost of model training diminishes with the increasing amount of inference data. In contrast, the time consumption of SL1MMER escalates with the increase in the data volume. In real-world TomoSAR processing, the number of pixels is usually in the tens or even hundreds of millions, thus making the application of SL1MMER or other second-order CS-based methods challenging. The proposed algorithm is able to complete the processing, including the training procedure, within a matter of hours. This superior computational efficiency makes large-scale super-resolving TomoSAR processing feasible and practical, which is a significant advantage of the proposed $\boldsymbol{\gamma}$-Net.

In addition, a further simulation was carried out to verify the performance of the proposed $\boldsymbol{\gamma}$-Net at limited number of baselines. We simulated data with only 6 baselines according to a real TanDEM-X images stack we have. The baseline ranges from -565.5m to 373.2m. Fig. 4.10 compares the performance of the two algorithms at limited number of measurements. As one can see, in the noisy case, i.e. SNR = 0dB, the two algorithms have similar performance. However, with the increase of the SNR level, the proposed algorithm outperforms SL1MMER by a fair margin. To be specific, the proposed algorithm outperforms SL1MMER by about 20% effective detection rate in moderate super-resolving cases at 6dB 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 ρ_s.

Real data

In experiment with real data, we worked with a stack of six pairs of TanDEM-X high resolution staring spotlight images, with each pair acquired in pursuit monostatic mode [80], with a slant-range resolution of 0.6m and an azimuth resolution of 0.25m. An optical image of the test site from Google Earth and the SAR mean intensity image are showed in Fig 4.11. The yellow arrow indicates the range direction. Preprocessing, such as multiple SAR images co-registration and phase calibration were carried out using the DLR's integrated wide area processor (IWAP) [81].

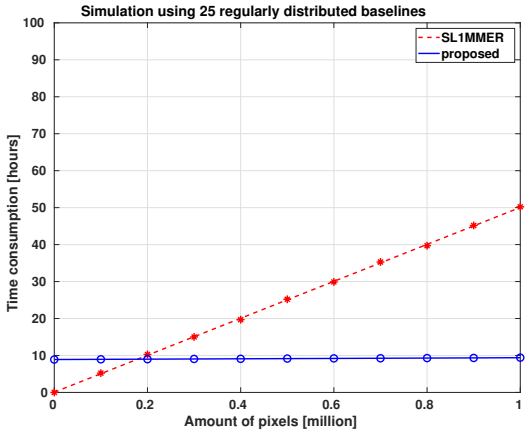
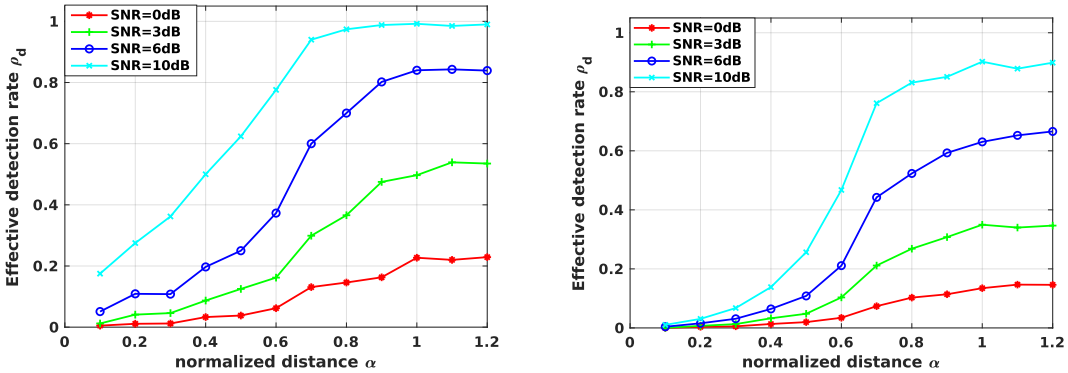


Figure 4.9: Comparison of time consumption between the proposed algorithm and SL1MMER. The training time will be affected only by the size of training data and the number of training epochs we set. The time consumption of SL1MMER escalates with the increasing number of data, whereas the inference time 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



(a) γ-Net (b) SL1MMER

Figure 4.10: Effective detection rate P_d as a function of the normalized elevation distance between double scatterers simulated with 6 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) the proposed algorithm, (b) SL1MMER.

Fig. 4.12 demonstrates the reconstruction result of γ-Net. As we can see in Fig. 4.12, γ-Net is able to detect dense double scatterers, which contribute to significant information increment and complete the structure of the individual building. Fig. 4.12(c) and Fig. 4.12(d) demonstrate the separation of the top and bottom layer of detected double scatterers. As we can see, most reflections from roof and facade are overlaid at the top of the building and γ-Net is able to separate these close layover.

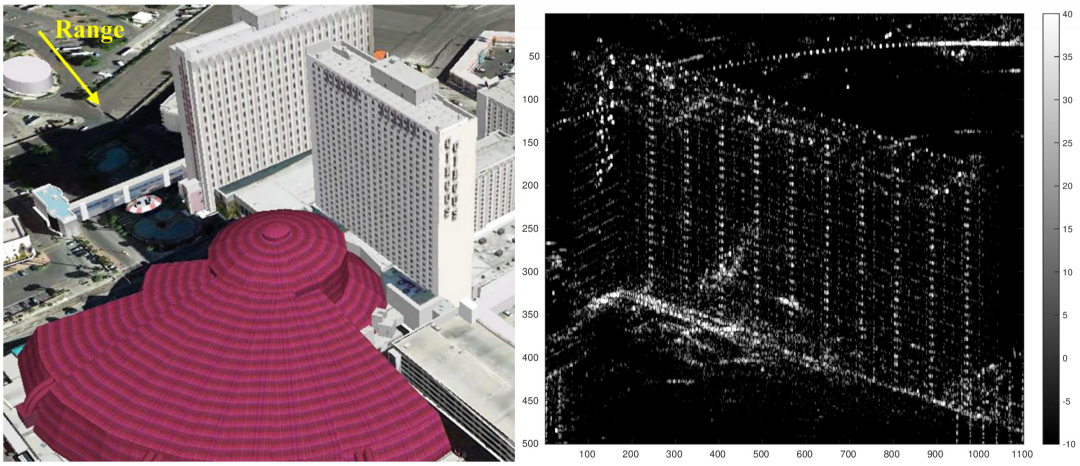
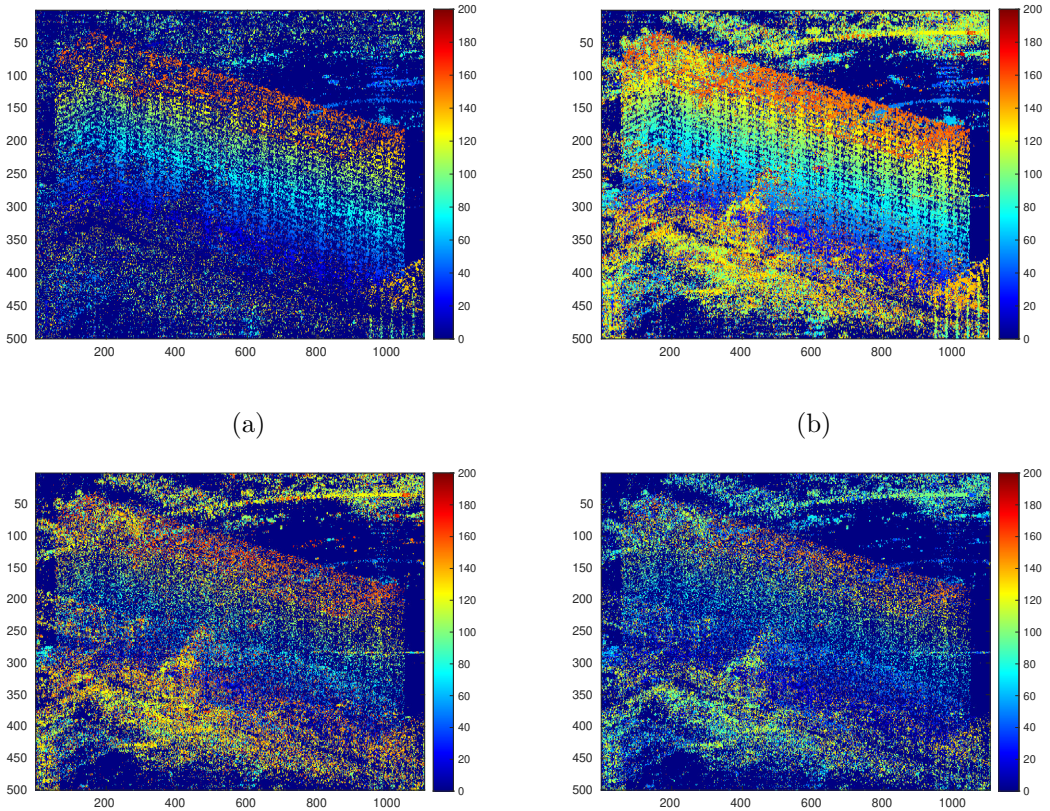


Figure 4.11: Test site. Left: optical image from Google Earth, right: SAR mean intensity image



(a) (b)

(c) (d)

Figure 4.12: Reconstructed and color-coded elevation estimated by γ-Net. (a) detected single scatterer, (b) detected single scatterer + top layer of detected double scatterers, (c) top layer of detected double scatterers, (d) bottom layer of detected double scatterers.

4.2 Complex-valued sparse minimal gate units (CV-SMGUs): Incorporation of historic information

To promote sparsity, γ-Net employs shrinkage operations to discard elements of lower magnitude. However, this approach can precipitate unavoidable loss of information within the network dynamics. Specifically, weak echoes originating from poorly reflective scatterers may be suppressed or even completely annihilated. Consequently, this process can result in a notable diminishment of the model performance. The challenge, therefore, lies in devising a method that promotes sparsity while minimally impacting the fidelity of these weaker signals. In this section, we propose an innovative architecture for recurrent neural networks (RNNs) that includes Sparse Minimal Gated Units (SMGUs) to address the issue of information loss. This novel structure takes advantage of the inherent historical data processing characteristic of RNNs. Through the SMGUs, we introduce a method of integrating historical information into the optimization process, thereby ensuring comprehensive retention and utilization of all pertinent information to the final output.

4.2.1 False detection caused by information loss

As depicted in Fig. 4.5, despite the modifications, γ-Net inherits the learning architecture with LISTA and thus is susceptible to the same limitations. More specifically, in the γ-Net architecture, the current layer only uses the information in the output from the previous layer, resulting in a final output that relies solely on the information from the second last layer. Due to this design, any crucial information that gets pruned away by shrinkage steps in the intermediate layers cannot contribute to the final output, leading to possible significant reconstruction errors. As an example, Fig. 4.13 illustrates an instance from our experiments where the γ-Net fails to accurately detect the presence of two scatterers. The loss of pertinent information in the intermediate layers, which cannot be recovered, likely contributes to this incorrect result.

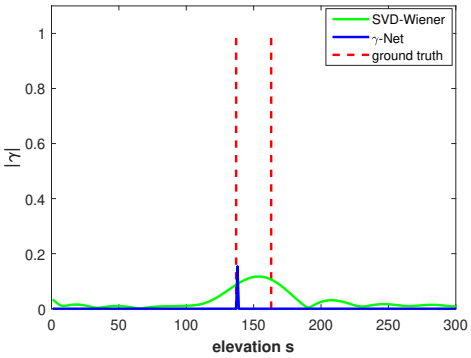


Figure 4.13: An example of unsuccessful detection of double scatterers caused by information loss. γ-Net detects one of the double scatterers with very high localization accuracy but fails to find the other one.

In Fig. 4.13, a scenario was simulated where two scatterers, assumed to have identical phase and amplitude, were placed 0.6 Rayleigh resolution units apart - a super-resolution regime. The SNR level was set at 6dB. As a general observation, the inability to resolve overlaid double scatterers should produce a reflectivity profile characterized by a single

dominant amplitude peak located between the true elevation positions of the scatterers. This behavior is illustrated by the non-superresolving algorithm SVD-Wiener [58], as shown in Fig. 4.13. In contrast, $\boldsymbol{\gamma}$-Net was able to detect one of the scatterers with high localization accuracy but failed to detect the second one. This anomaly prompted us to hypothesize that the unsuccessful separation of double scatterers may be attributed to the information loss induced by the shrinkage steps in $\boldsymbol{\gamma}$-Net. Further investigation of the intermediate layers in $\boldsymbol{\gamma}$-Net revealed a gradual diminishment of information pertaining to the second scatterer after each shrinkage step. By the second last layer, the information associated with the second scatterer was completely lost. As a result, the final output of $\boldsymbol{\gamma}$-Net, i.e., the estimated γ, was devoid of any information regarding the second scatterer, leading to its detection error.

4.2.2 Incorporation of historic information

In the realm of optimization, numerous studies have demonstrated and established that the incorporation of historical information can substantially enhance algorithmic performance [82] [83] [84]. Building on these pivotal insights, researchers introduced an adaptive ISTA in [85] that integrates and makes use of historical information through the introduction of two adaptive momentum vectors, $\mathbf{f}$ and $\mathbf{i}$, into each iteration of ISTA. This approach can be mathematically represented as follows:

$$\begin{aligned} \bar{\mathbf{c}}^k &= \mathbf{W}_2\hat{\boldsymbol{\gamma}}^{k-1} + \mathbf{W}_1\mathbf{g} \\ \mathbf{c}^k &= \mathbf{f}^k \odot \mathbf{c}^{k-1} + \mathbf{i}^k \odot \bar{\mathbf{c}}^k \\ \hat{\boldsymbol{\gamma}}^k &= \eta_{st}\left(\mathbf{c}^k\right) \end{aligned} \tag{4.13}$$

where η_{st} indicates the conventional soft-thresholding function. In comparison to ISTA, which can be equivalently expressed as $\hat{\boldsymbol{\gamma}}^k = \eta_{st}\left(\bar{\mathbf{c}}^k\right)$ using the same notation, the adaptive ISTA accounts for both current and prior information. More specifically, in the t^{th} iteration of the adaptive ISTA, the estimate is derived through a linear combination of the historical information, $\mathbf{c}^{k-1}$, obtained from the previous iteration and the current information, $\bar{\mathbf{c}}^k$. These two sources of information, historical and current, are weighted by the adaptive momentum vectors $\mathbf{f}^k$ and $\mathbf{i}^k$ respectively. As a result, the final estimate generated by the adaptive ISTA accumulates historical information, each component of which is weighted differently according to its corresponding $\mathbf{f}^k$ and $\mathbf{i}^k$.

4.2.3 Network architecture

One primary challenge associated with the adaptive ISTA is determining the two momentum vectors in each adaptive ISTA iteration, namely $\mathbf{f}^k$ and $\mathbf{i}^k$. Currently, there is no analytical method to determine these vectors. Commonly, they are selected via tediously hand-craft tuning, which is not only time-consuming but also cannot guarantee optimal performance. To resolve this challenge better leverage the power of incorporating the historical information for solving TomoSAR inversion, we propose a novel gated RNN structure termed the sparse minimal gated unit (SMGU) based on the valuable works in [86] [87]. This structure offers two significant improvements. Firstly, it couples the two momentum vectors into a single gate, thereby simplifying the architecture [88] [89] [90]. Secondly, it parameterizes the gate, allowing its learning from data. This eliminates the need for time-consuming manual tuning, enhancing the efficiency of the process.

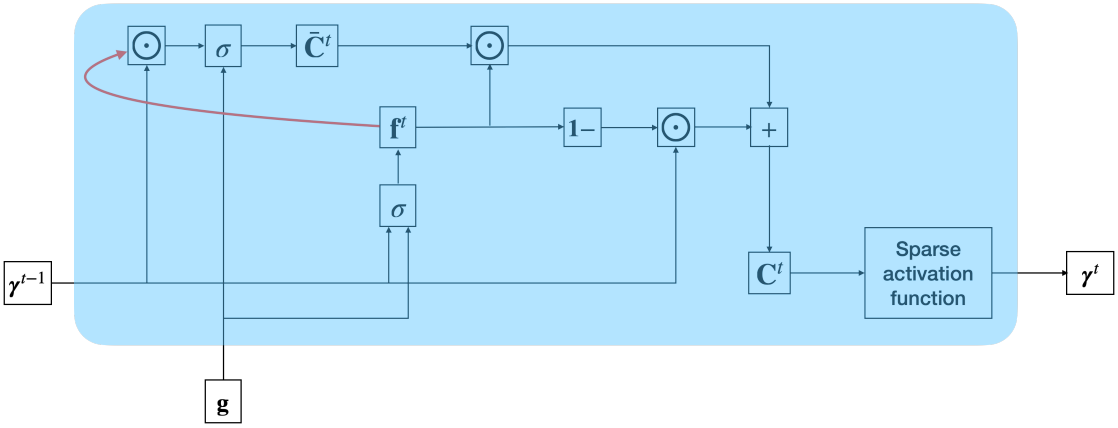


Figure 4.14: Structure of the proposed SMGU. **f** indicated the only gate in each SMGU.

Fig. 4.14 illustrates us the learning architecture and the detailed equations for defining SMGU are given by:

$$\begin{aligned}
\mathbf{f}^k &= \sigma\{\mathbf{W}_{f2}^k \hat{\boldsymbol{\gamma}}^{k-1} + \mathbf{W}_{f1}^k \mathbf{g}\} \\
\bar{\mathbf{c}}^k &= \mathbf{W}_2(\mathbf{f}^k \odot \hat{\boldsymbol{\gamma}}^{k-1}) + \mathbf{W}_1 \mathbf{g} \\
\mathbf{c}^k &= (1 - \mathbf{f}^k) \odot \hat{\boldsymbol{\gamma}}^{k-1} + \mathbf{f}^k \odot \bar{\mathbf{c}}^k \\
\hat{\boldsymbol{\gamma}}^k &= \eta_{dt}\left(\mathbf{c}^k\right)
\end{aligned} \tag{4.14}$$

where η_{dt} is the double hyperbolic tangent function to promote sparsity and defined as:

$$\eta_{dt}(\hat{\gamma}, s, \theta) = s \cdot [\tanh(\hat{\gamma} + \theta) + \tanh(\hat{\gamma} - \theta)] \tag{4.15}$$

where s and θ denote two trainable parameter. It is noteworthy that the double hyperbolic tangent function can be regarded as a smooth and continuously differentiable alternative to the conventional soft-thresholding function, offering two primary advantages. Firstly, its second derivative sustains for a long span, which helps alleviate the gradient vanishing problem induced by the cell recurrent connection [91]. Secondly, it can effectively imitate the soft-thresholding function within the $[-\theta, \theta]$ interval. This functionality allows for a more seamless and efficient integration into the network, improving the overall performance. Fig. 4.15 demonstrates an example of the double hyperbolic tangent function and compares it to the soft-thresholding function.

As depicted and formulated in the Sparse Minimal Gated Unit (SMGU), the following operations are conducted in the k^{th} layer of a gated Recurrent Neural Network (RNN) constructed with SMGUs. Initially, the forget gate $\mathbf{f}^k$ is computed. The short-term response $\bar{\mathbf{c}}^k$ is then derived by combining the input data $\mathbf{g}$ with the "forgotten" portion of the output from the previous layer, expressed as $\mathbf{f}^k \odot \hat{\boldsymbol{\gamma}}^{k-1}$. Subsequently, the new hiddne state of the current layer, $\mathbf{c}^t$, is formulated by combining portions of $\hat{\boldsymbol{\gamma}}^{k-1}$ and the short-term response $\bar{\mathbf{c}}^k$, respectively weighted by $(1 - \mathbf{f}^k)$ and $\mathbf{f}^k$. In this formulation, we can see that the SMGU is able to simultaneously execute a two-fold task with only one forget gate. On the one hand, SMGU allows a compact representation by enabling the hidden state $\mathbf{c}^{(t)}$ to discard irrelevant or redundant information. On the other hand, SMGU is capable of controlling how much information from the previous layer to take over.

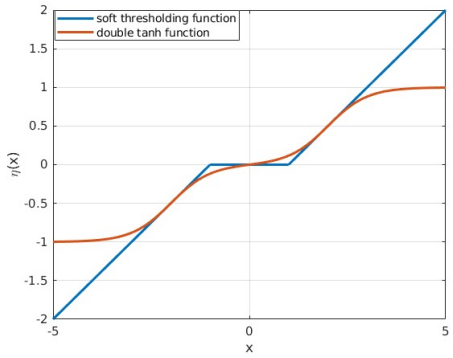


Figure 4.15: Comparison of double hyperbolic tangent function $\eta_{dt}(\cdot)$ and soft-thresholding function. $\eta_{dt}(\cdot)$ effectively imitates the soft-thresholding function within the interval of $[-\theta, \theta]$.

Ultimately, the sparse activation function, represented by the double hyperbolic tangent function, is applied to the current hidden state $\mathbf{c}^t$ to induce shrinkage and thresholding, thereby enhancing the sparsity of the output.

The extension of the SMGU into the complex domain can be conducted in a manner analogous to the one presented in Eq. (4.5), by assuming that each neuron has two channels representing the real and imaginary parts of a complex number. Moreover, the complex-valued version of the double hyperbolic tangent function is defined as:

$$\eta_{cv-dt}(\hat{\boldsymbol{\gamma}}^k, s, \theta) = \begin{cases} \frac{\hat{\boldsymbol{\gamma}}^k}{|\hat{\boldsymbol{\gamma}}^k|} s \cdot [\tanh(|\hat{\boldsymbol{\gamma}}| + \theta) + \tanh(|\hat{\boldsymbol{\gamma}}| - \theta)], & |\hat{\boldsymbol{\gamma}}_i| \neq 0 \\ 0 & \text{else} \end{cases} . \tag{4.16}$$

4.2.4 Experiment and validation

In this section, we present experimental results that highlight the performance enhancements brought by the integration of historical information into our proposed CV-SMGUs, as compared to $\boldsymbol{\gamma}$-Net.

Simulation

The simulation was conducted using a TomoSAR benchmark test [58] [12] at two SNR levels, i.e., SNR $\in 0, 6$ dB. The aim of this simulation was to examine the super-resolution power and estimation accuracy by analyzing the effective detection rate. In Fig. 4.16, the effective detection rate P_d of the proposed CV-SMGUs and $\boldsymbol{\gamma}$-Net is compared, which is presented as a function of the normalized distance α. By comparing the effective detection rates for the two methods at various α levels, we can gain insights into their respective super-resolution capabilities and the effectiveness of the proposed CV-SMGUs. The results of these comparisons are discussed in the subsequent sections. As illustrated in Fig. 4.16, CV-SMGUs demonstrates superior performance compared to γ-Net at both tested SNR levels. More specifically, CV-SMGUs can achieve an effective detection rate that is approximately 10-20% higher than it of $\boldsymbol{\gamma}$-Net in moderate super-resolution cases when the SNR is at 6 dB. In the more challenging scenario where the SNR is at 0 dB, which represents a noisier environment, CV-SMGUs gradually approach an effective detection

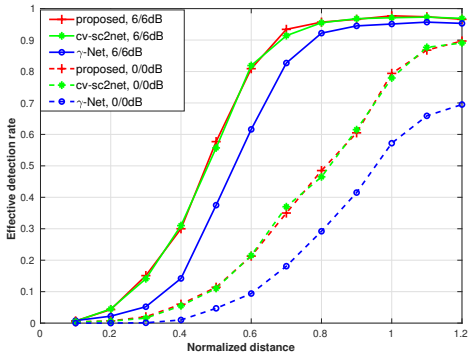


Figure 4.16: Effective detection rate of the proposed algorithm, CV-sc2net and γ-Net as a function of the normalized elevation distance between the simulated facade and ground with SNR = 0dB and 6dB under 0.2 million Monte Carlo trials.

rate of about 90% as the normalized distance increases. On the other hand, γ-Net only reaches a maximum effective detection rate of around 70% under similar conditions. These results clearly demonstrate the improvement of super-resolution power of CV-SMGUs. By integrating historic information into the estimation process, CV-SMGUs are able to achieve higher effective detection rates, even in noisy conditions.

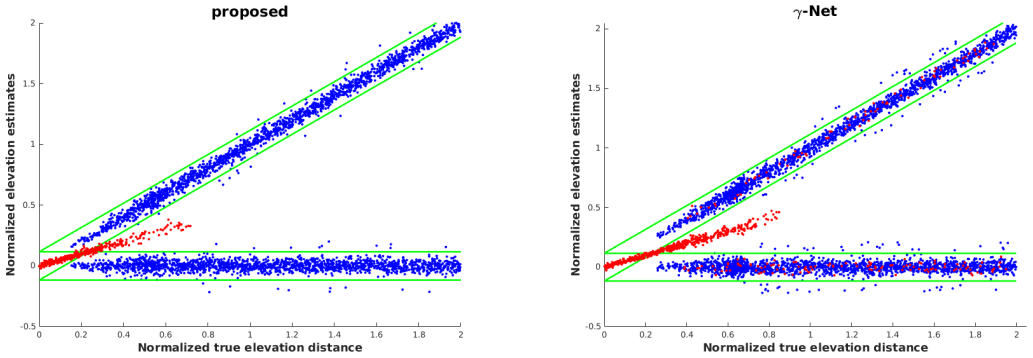


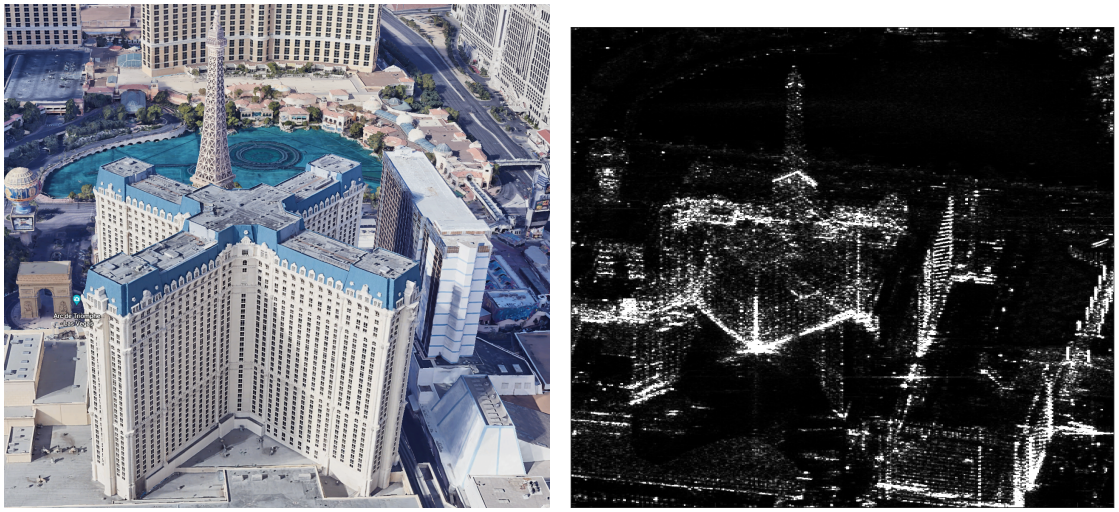
Figure 4.17: Normalized estimated elevation of facade and ground of increasing elevation distance, with SNR=6dB and N=25. The double scatterers were simulated to have identical phase and amplitude. The true positions are a horizontal line referring to the ground and a diagonal line referring to the scatterers at variable elevation. The green lines depict true positions ± 3 times CRLB of elevation estimates for single scatterers. Red dots represent samples detected as single scatterers. Blue dots indicate detected overlaid double scatterers.

To better manifest how CV-SMGUs outperform γ-Net, we simulated 2000 samples containing double scatterers with increasing scatterers distance at 6 dB SNR. We made a scatter plot of their elevation estimates and color coded the points by the detector decision in Fig. 4.17. The x-axis refers to the true normalized elevation distance of the scatterers. The y-axis shows their normalized elevation estimates. The ideal reconstruction would be a horizontal and a diagonal straight line, which represent the ground truth of the simulated ground and facade. The green lines refer to ground truth ±3 times CRLB of single scatterer elevation estimate. The blue dots indicate the detected double

scatterers, whereas the red dots represent the samples were detected as single scatterers, meaning the second scatterer were lost in the network output. Fig. 4.17 clear shows that (1) γ-Net experiences much more red dots locate within ± 3 times CRLB w.r.t the ground truth, meaning it occasionally can only detect one of the double scatterers but is able to estimate its elevation with high precision. We ascribe this problem to the information loss caused by the learning structure of γ-Net. In the contrary, the proposed algorithm utilizes CV-SMGUs to preserve full information, thus avoiding discarding any significant information; (2) the proposed algorithm is able to resolve double scatterer at much smaller scatterers distance. Specifically, the proposed algorithm starts to separate double scatterer from about 0.15 Rayleigh resolution, whereas γ-Net can only detect double scatterer only after about 0.3 Rayleigh resolution.

Real data

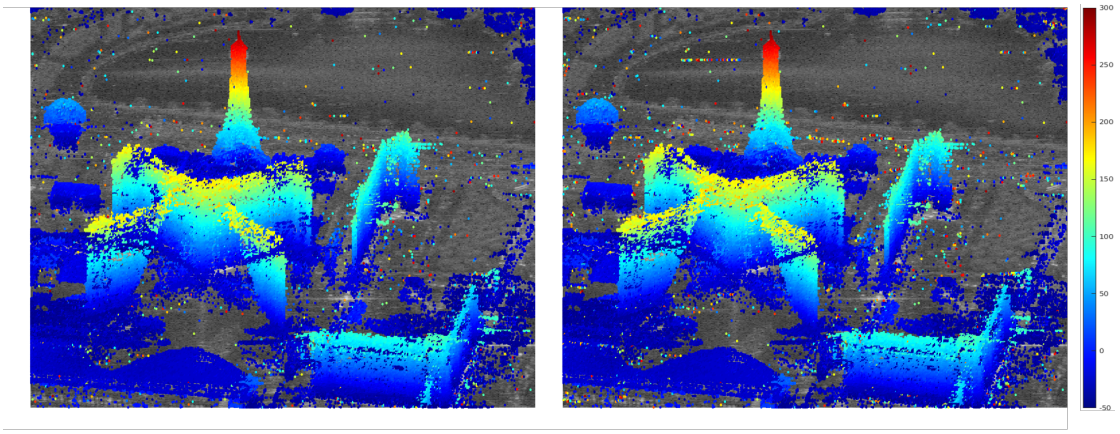
In the real-world data experiment, we tested our CV-SMGUs model on a stack of SAR images from the TerraSAR-X high-resolution spotlight images. The stack consists of 50 images, all centered over the city of Las Vegas, specifically focusing on the Paris Hotel. The TerraSAR-X images offer a slant-range resolution of 0.6 meter and an azimuth resolution of 1.1 meter. Fig. 4.18 provides a visual representation of our testing site. It includes an optical image sourced from Google Earth and a SAR mean intensity image that indicates the distribution and intensity of radar reflections over the test site.



(a) (b)

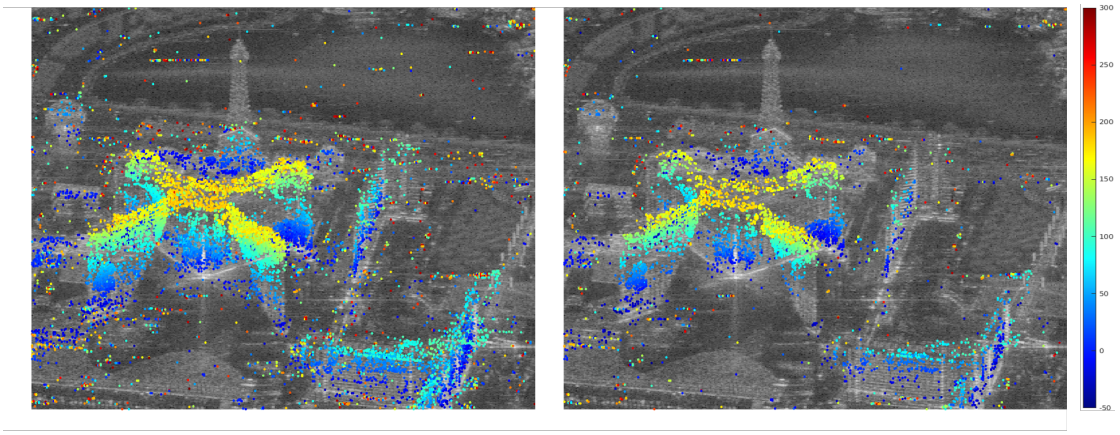
Figure 4.18: Test site of Paris hotel. (a): optical image from Google Earth, (b): SAR mean intensity image

The reconstruction results of the test site are demonstrated in Fig. 4.19 and compared to the results derived by γ-Net. In Fig. 4.19, (a) and (b) illustrate color-coded elevation of single scatterers detected by both algorithms. (c)-(f) depict the reconstruction of detected double scatterers of the both algorithm. The double scatterers are separated into the top and bottom layer according to their elevation estimates and the top and bottom layers are demonstrated separately. By comparing the reconstruction results of the both algorithms, we can see that the proposed algorithm detects the double scatterers with a



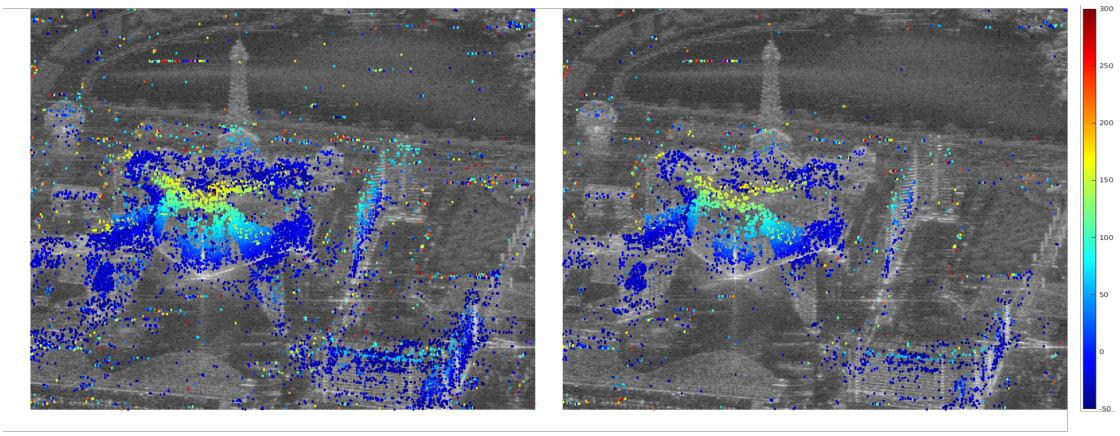
(a), single scatterer detected by the proposed algorithm

(b), single scatterer detected by γ-Net



(c), top layer of double scatterers detected by the proposed algorithm

(d), top layer of double scatterers detected by γ-Net



(e), bottom layer of double scatterers detected by the proposed algorithm

(f), bottom layer of double scatterers detected by γ-Net

Figure 4.19: Reconstructed and color-coded elevation of detected scatterers. From left to right: Elevation estimates derived by the proposed algorithm and γ-Net, respectively. From top to bottom: Color-coded elevation of detected single scatterers, top layer of detected double scatterers and bottom layer of detected double scatterers, respectively.

higher density, indicating that the proposed algorithm has stronger super-resolution power. Closer inspection of the reconstruction of double scatterers shows that serious layover exists on the top of the cross building. Moreover, the elevation estimates of detected double scatterers indicate that the top layer is mainly caused by reflections from building roof and building facade, whereas the bottom layer is composed of scatterers on the ground or lower infrastructures.

To provide a more intuitive comparison of the super-resolution power of both algorithms, we summarized the scatterers detection of both algorithms in Table 4.2. As it is shown in Table 4.2, most pixels are detected as 0 scatterers by the two algorithms because the fountain and many low infrastructures in the test site exhibit no strong scattering, which can be seen in Fig. 4.18(b). Comparing to γ-Net, the proposed algorithm detected less single scatterers (33.30%), but more double scatterer. Comparison between the double scatterers detected by both algorithms shows that the proposed algorithm is able to detect 95.2% of the double scatterers detected by γ-Net. Moreover, it detects 50% more double scatterers than γ-Net.

Table 4.2: Percentage of scatterers detection for CV-SMGUs and γ-Net.

Algorithm	Percentage of detection as		
	0 scatterer	1 scatterer	2 scatterers
CV-SMGUs	62.01 %	33.30 %	4.69 %
γ-Net	61.06 %	35.83 %	3.11 %

Further investigation was conducted to inspect the improvement of double scatterer detection. The histogram of detected double scatterers' elevation difference from the proposed algorithm and γ-Net is shown in Fig. 4.20. In the non-super-resolution region, especially when the distance between double scatterers is larger than twice Raleigh resolution, the two algorithms have comparable performance of double scatterers detection. However, in the super-resolution region, the proposed algorithm delivers obviously stronger resolution ability.

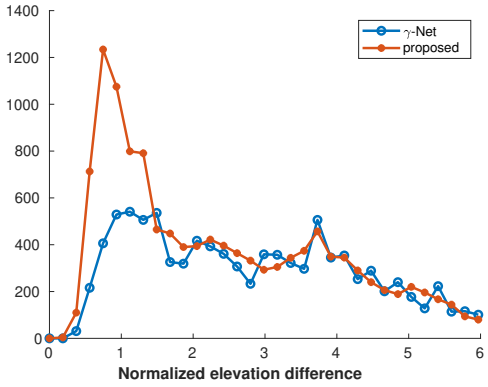


Figure 4.20: Histogram of the elevation distance between the detected double scatterers from the proposed algorithm and γ-Net. The proposed algorithm shows significantly more detection in the super-resolution region.

4.3 A hybrid approach for multi-component D-TomoSAR inversion

Deep learning-based TomoSAR algorithms have made impressive success. Nevertheless, current deep learning-based TomoSAR algorithms are limited to 3-D reconstruction. The extension of such deep learning-based algorithms differential TomoSAR (D-TomoSAR) applications, especially for multi-component D-TomoSAR, is mainly obstructed by the need for training high-dimensional weight matrices. The network designed for D-TomoSAR inversion typically necessitates millions of freely trainable parameters. The learning process for such an enormous quantity of weights demands a considerable number of training samples, leading to a substantial memory load and extensive time consumption. In this section, we propose a hybrid approach named HyperLISTA-ABT for accurate and efficient multi-component D-TomoSAR processing. The weights in HyperLISTA-ABT are analytically determined in adherence to a minimum coherence criterion, which significantly reduces the model to an ultra-lightweight version equipped with only three hyperparameters. Furthermore, HyperLISTA-ABT improves the global thresholding method by employing an adaptive blockwise thresholding scheme. This strategy incorporates block-coordinate techniques and implements thresholding within local blocks, enabling the retention of weak expressions and local features throughout each shrinkage step, thereby enhancing model performance.

4.3.1 Limitation of deep learning-based algorithms applying to high-dimensional D-TomoSAR inversion

By employing $\boldsymbol{\gamma}$-Net as an illustrative example, we will explain the challenges associated with implementing deep learning-based algorithms for solving D-TomoSAR inversion. Reflecting on the learning architecture of $\boldsymbol{\gamma}$-Net, we note that each layer requires the training of a weight matrix, $\mathbf{W}^k$, of dimension $N \times L$. Here, N is dictated by the number of stacked SAR acquisitions and is typically fixed within a reasonable range. Conversely, L is governed by the number of grid cells after discretization.

In 3-D reconstruction scenarios, L usually falls within the hundreds for spaceborne sensors, given that only elevation needs to be discretized. Thus, the product $N \times L$ remains in the thousands. However, this number increases dramatically in D-TomoSAR cases when multi-component motion terms, usually linear and periodic motions, are taken into consideration. The training of the network then becomes conversely a challenge due to the tremendous amount of free trainable parameters.

For instance, when we consider two motion terms, i.e., linear and seasonal motion, the value of L will be determined by the product of the discretized grid numbers along each direction $L = L_s \times L_v \times L_a$, where L_s, L_v, and L_a indicate the discretization levels for elevation, linear motion, and seasonal motion, respectively. Even with a conservative discretization level for TerraSAR-X image stacks—200 for L_s, 50 for L_v, and 50 for L_a—the product yields a value of L equal to 0.5 million. This indicates millions of parameters to be learned for each weight matrix. Such large weight matrices bring two unavoidable disadvantages. Firstly, the model requires an enormous number of training samples, rendering the training procedure highly inefficient. Secondly, training a model of such magnitude demands a significant volume of GPU memory, a requirement typically unmet by consumer-level GPUs.

4.3.2 HyperLISTA-ABT: An ultra-light model

In order to circumvent the onerous model training necessitated by large weight matrices, an analytical weight optimization strategy has been proposed in [92] [93], which is based on the principle of coherence minimization. It enables the analytic determination of weights within an unrolled neural network designed specifically for sparse recovery applications, such as LISTA and its variants. Accordingly, the weight matrix can be computed in the following manner:

$$\begin{aligned}\hat{\mathbf{W}} &= \arg\min_{\mathbf{W}} \mu(\mathbf{W}, \mathbf{R}) \\ &= \arg\min_{\mathbf{W}} \inf_{\mathbf{W}\in\mathbb{C}_{N\times L}} \max_{i\neq j} \mathbf{W}_{:,i}^T \mathbf{R}_{:,j} \\ &\text{s.t. } \forall i \in \{1,\dots,L\} : \mathbf{W}_{:,i}^T \mathbf{R}_{:,i} = 1 \end{aligned} \tag{4.17}$$

This approach effectively alleviates the burden of extensive model training, paving the way for more efficient computational processes.

The analytic method of weight determination significantly trims the number of trainable parameters, necessitating the learning of only the stepsize and thresholds. This methodology, however, was taken a step further with the proposition of HyperLISTA in [93]. HyperLISTA reconceptualizes the remaining trainable parameters as instance-adaptive parameters, further condensing them into just three hyperparameters, formulated as follows:

$$\begin{aligned}\boldsymbol{\gamma}_{k+1} = \eta_{\theta^k}^{p^k}(\boldsymbol{\gamma}_k + \boldsymbol{W}^H(\boldsymbol{g} - \boldsymbol{R}\boldsymbol{\gamma}_k) \\ +\beta^k(\boldsymbol{\gamma}_k - \boldsymbol{\gamma}_{k-1}))\end{aligned} \tag{4.18}$$

where

$$\theta^k = c_1 \left\| \boldsymbol{R}^+ (\boldsymbol{R}\boldsymbol{\gamma}_k - \boldsymbol{g}) \right\|_1 \tag{4.19}$$

$$\beta^k = c_2 \left\| \boldsymbol{\gamma}_k \right\|_0 \tag{4.20}$$

$$p^k = c_3 \min\left(\log\left(\frac{\left\| \boldsymbol{R}^+ \boldsymbol{g} \right\|_1}{\left\| \boldsymbol{R}^+ (\boldsymbol{R}\boldsymbol{\gamma}_k - \boldsymbol{g}) \right\|_1} \right), L \right) \tag{4.21}$$

where c_1, c_2, and c_3 indicate the three hyperparameters to be tuned.

While HyperLISTA exhibits exceptional efficiency and performance as demonstrated in [93], it unfortunately succumbs to the same issue of information loss when applied to TomoSAR processing, a problem outlined in a previous section. Due to the high level of parameterization, it becomes impossible to utilize gated units such as those in CV-SMGUs to address this loss of information. Upon closer examination through experimentation, it became apparent that this information loss could be traced back to the global thresholding scheme. This scheme applies the same threshold value uniformly across all entries in the signal, leading to a situation where a threshold that effectively captures strong spikes might unintentionally suppress weaker, but nonetheless significant, spikes in the reflectivity profile.

To cope with the aforementioned issue, we propose HyperLISTA-ABT, which is an improvement of the original HyperLISTA by incorporation of an adaptive blockwise thresholding (ABT) scheme that explores a local thresholding strategy. The advantages of HyperLISTA-ABT is three-fold.

- It carries out thresholding within each local block, thereby facilitating a more refined thresholding process, and possibly preserving weak reflections from dark scatterers. This method allows for better capture of the diverse range of spike magnitudes encountered in the signal, thereby improving the accuracy and reliability of the reflectivity profile characterization.

- HyperLISTA-ABT proves to be more efficient since it only updates one block of variables at a time, rather than updating all variables simultaneously. As a result, HyperLISTA-ABT is better suited for our large-scale and high-dimensional application.

- HyperLISTA-ABT shrinks the block size in a layerwise manner, which contributes to a better fine-focusing capability.

According to [94] [95], the update rules of HyperLISTA-ABT after applying block coordinate techniques can be written as:

$$\begin{aligned}\boldsymbol{\gamma}_{i_p}^{k+1} = \eta_{\theta_{i_p}^k}(\boldsymbol{\gamma}_{i_p}^k + \boldsymbol{W}_{i_p}^T(\boldsymbol{y} - \boldsymbol{R}_{i_p}\boldsymbol{\gamma}_{i_p}^k) \\ +\beta_{i_p}^k(\boldsymbol{\gamma}_{i_p}^k - \boldsymbol{\gamma}_{i_p}^{k-1}))\end{aligned} \tag{4.22}$$

where i_p is the index of the updated block. To clarify, in HyperLISTA-ABT, we remove the support selection scheme and just use the conventional soft-thresholding function. The threshold $\theta^{(k)}$ and the factor $\beta^{(k)}$ are determined for each block as well:

$$\theta_{i_p}^k = c_1 \left\| \boldsymbol{R}_{i_p}^+ \left(\boldsymbol{R}_{i_p}\boldsymbol{\gamma}_{i_p}^k - \boldsymbol{g} \right) \right\|_1 \tag{4.23}$$

$$\beta_{i_p}^k = c_2 \left\| \boldsymbol{\gamma}_{i_p}^k \right\|_0 \tag{4.24}$$

where $c_1 > 0$, $c_2 > 0$, and $c_3 \in (0, 1)$ are the three hyperparameters. Notably, c_3 is a latent hyperparameter and plays a crucial role in controlling the blocksize despite it not explicitly appearing in the formula. In our application, we usually initialize the blocksize according to the grid number within half of the Rayleigh resolution. The block is chosen with a random variants scheme where i_p follows the probability distribution given by:

$$P_{i_p} = \frac{L_{i_p}}{\sum_{j=1}^{J} L_{i_p}}, \quad i_p = 1, \ldots, J \tag{4.25}$$

where J is the number of blocks and $L_{i_p} = ||R_{i_p}^T R_{i_p}||$. All the hyperparameters c_1, c_2, and c_3 can be selected using the grid search method. In grid search, a coarse grid is first applied to find an interest region, and then this is zoom-in with a fine-grained grid. The hyperparameters are determined by minimizing the normalized mean square error (NMSE) over the simulated ground truth.

With the blockwise threshlding scheme, local features and weak expressions can be possibly retained. This is due to the fact that many elements of the entries are not strictly driven to zero but to some extremely small value, thus making the output not strictly sparse. Therefore, a post-processing is usually required to clean the output and make it sparse. The framework of the proposed HyperLISTA-ABT is summarized in the following table.

Algorithm 2 Summary of HyperLISTA-ABT

Generate steering matrix R for given baselines
  Analytic weight optimization **W** according to Eq. (4.17)
Tuning of hyperparameters
  Simulate ground truth of reflectivity profile $\boldsymbol{\gamma}$ [77]
  Simulate noise-free SAR acquisitions $\mathbf{g} = \mathbf{R}\boldsymbol{\gamma}$
  Grid search to determine the hyperparameters by
  minimizing NMSE over simulated data
Inference
  Init: $\boldsymbol{\gamma}^{(0)} = \mathbf{R}^H\mathbf{g}$ and blocksize B_1
  for $k = 1, 2, \cdots, K$ **do**
    Determine the number of blocks J_k
    based on the blocksize B_k
    for $i_p = 1, 2, \cdots, J_k$ **do**

$$\begin{aligned}
\boldsymbol{\gamma}_{i_p}^{(k+1)} &= \eta_{\theta_{i_p}^{(k)}}(\boldsymbol{\gamma}_{i_p}^{(k)} + \boldsymbol{W}_{i_p}^T(\boldsymbol{y} - \boldsymbol{R}_{i_p}\boldsymbol{\gamma}_{i_p}^{(k)}) \\
&\quad + \beta_{i_p}^{(k)}(\boldsymbol{\gamma}_{i_p}^{(k)} - \boldsymbol{\gamma}_{i_p}^{(k-1)})) \\
\theta_{i_p}^{(k)} &= c_1 \left\| \boldsymbol{R}_{i_p}^{+} \left(\boldsymbol{R}_{i_p}\boldsymbol{\gamma}_{i_p}^{(k)} - \boldsymbol{g} \right) \right\|_1 \\
\beta_{i_p}^{(k)} &= c_2 \left\| \boldsymbol{\gamma}_{i_p}^{(k)} \right\|_0
\end{aligned}$$

    end for
    Update blocksize with $B_{k+1} = c_3 \cdot B_k$
  end for
Output clean-up
Model order selection and final estimation

4.3.3 Experiment and validation

In this section, we demonstrate the efficacy and efficiency of the HyperLISTA-ABT algorithm for multi-component D-TomoSAR processing through a series of experiments conducted on both simulated and real-world data sets.

Simulation

In the simulation, we carried out experiments based on TomoSAR inversion to exhibit the improvements introduced by HyperLISTA-ABT over the original HyperLISTA [93] and to compare its performance with existing state-of-the-art compressed sensing (CS)-based and deep learning-based methods. Since the use of existing deep learning-based algorithms is not feasible for D-TomoSAR cases, as explained in the previous section, the focus of our simulation was solely on TomoSAR inversion for 3-D reconstruction.

In the first experiment, we aimed to investigate the performance improvement offered by HyperLISTA-ABT over the original HyperLISTA by inspecting their effective detection rates at different amplitude ratios, which is demonstrated in Fig. 4.21.

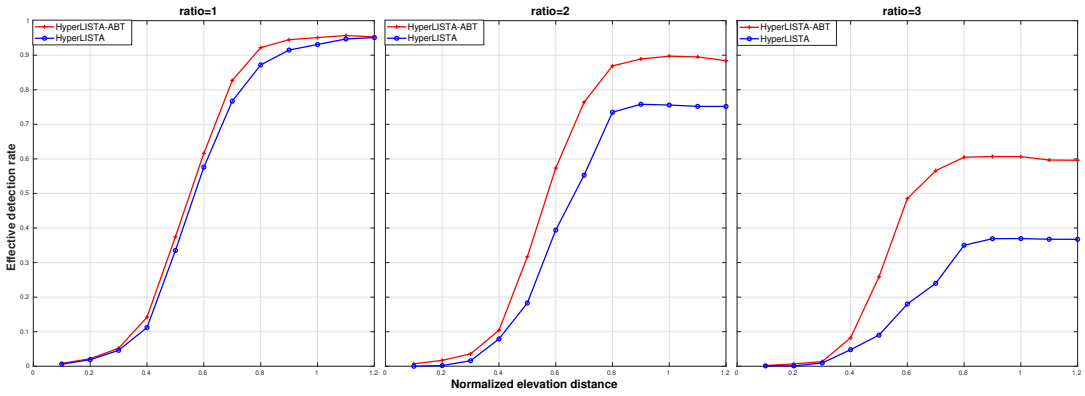
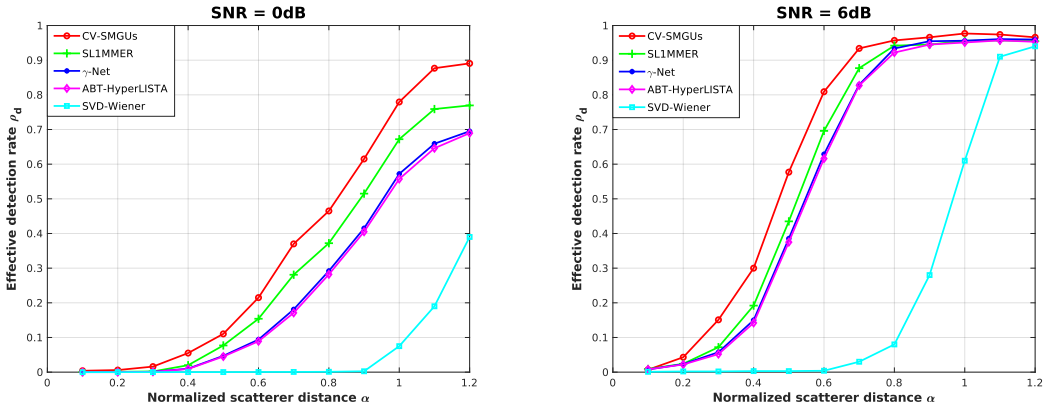


Figure 4.21: Effective detection rate of HyperLISTA-ABT and the original HyperLISTA with respect to the normalized elevation distance at different amplitude ratios. The overlaid double scatterers were set to have an identical phase and the SNR level was 6dB. HyperLISTA-ABT significantly outperformed HyperLISTA at high amplitude ratios between the scatterers.

The results demonstrate that HyperLISTA-ABT achieved a significantly higher effective detection rate than the original HyperLISTA. Both algorithms (in fact, all other methods) experience performance degradation with respect to an increase in amplitude ratio. This is attributed to two main factors. Firstly, dark scatterers experience a large bias in their elevation estimates at high amplitude ratios due to their elevation estimates approaching the more prominent ones. Consequently, many detections of double scatterers will not be recognized as effective due to the large elevation estimate bias. Secondly, the energy of dark scatterers is close to the noise level at high amplitude ratios. This makes it particularly challenging for HyperLISTA, which employs a global thresholding scheme, to detect the local features of dark scatterers. Further elaborating, when high-intensity scatterers are present in the signal, their strong energy can overshadow the low-energy regions where dark scatterers are located. This overshadowing effect can lead to the suppression or even annihilation of the weaker expressions associated with the dark scatterers. Consequently, the presence of these strong intensity scatterers can mask or obscure the signals originating from the dark scatterers, making their detection and characterization challenging. In contrast, HyperLISTA-ABT conducts thresholding in each local block, which can allow retaining local information and, thus, it can detect dark scatterers. This results in a higher effective detection rate at high amplitude ratios.

In addition, we compared HyperLISTA-ABT with other state-of-the-art TomoSAR algorithms, which are deep learning-based algorithms γ-Net [77] and CV-SMGUs [96], as well as the traditional CS-based method SL1MMER [13] with second-order optimization. To highlight the super-resolution ability of these methods, we also involved a conventional spectral estimator SVD-Wiener [58] as a baseline in the comparison. he comparison results are demonstrated in Fig. 4.22.

From the comparison results, we can see that all the methods except the conventional spectral estimator SVD-Wiener showed a great super-resolution power. The proposed HyperLISTA-ABT delivered almost the same super-resolution ability as γ-Net and approached the performance of SL1MMER in both scenarios. When focusing solely on the effective detection rate, it was challenging to proclaim a clear advantage of the proposed



(a) (b)

Figure 4.22: Detection rate P_d as a function of the normalized elevation distance between the simulated facade and ground with SNR = 0 dB and 6 dB, N = 25, and phase difference $\triangle\phi = 0$ (worst case) under 0.2 million Monte Carlo trials.

HyperLISTA-ABT method over the existing state-of-the-art approaches. In fact, when comparing it to CV-SMGUs, we could observe a slight underperformance. However, all the state-of-the-art methods come with a relatively high computational cost. Both γ-Net and CV-SMGU require tailored training according to the baseline distribution of the stack. SL1MMER is a model-based algorithm, thus needs no training, yet requires significantly computational time for solving the L1-norm minimization.

Table 4.3: Comparison of the number of required training samples and time consumption for processing 0.2 million Monte Carlo trials with each algorithm. The training time of HyperLISTA-ABT indicates the combined duration of both the analytic weight optimization process and the tuning of hyperparameters. It provides a measure of the overall time required for these essential steps.

Algorithm	number of training samples	training time	inference time	total time consumption	transferability
CV-SMGUs	4 million	$\approx$ 10 hours	$\approx$ 0.25 h	$\approx$ 10 h	low
γ-Net	3 million	$\approx$ 8 hours	$\approx$ 0.2 h	$\approx$ 8 h	low
SL1MMER	-	-	$\approx$ 20 h	$\approx$ 20 h	high
HyperLISTA-ABT	-	$\approx$ 0.5 hour	$\approx$ 0.25 h	$\approx$ 1 h	medium

We tested and recorded the time consumption of different algorithms for processing the 0.2 million Monte Carlo trials as well as the requirements for training data. The results are summarized in Table 4.3. To clarify, all inference was conducted using a local CPU for a fair comparison. As can be seen in Table 4.3, it took about ten hours for the deep learning-based algorithms to process 0.2 million Monte Carlo trials, which was predominantly the training time. In addition, a large amount of training samples was essential as well. For SL1MMER, it took about 20 hours for the processing since the iterative second-order optimization is computationally expensive. Further inspecting the table, we can see that HyperLISTA-ABT showed similar efficiency in the inference as the other deep learning-based algorithms. However, HyperLISTA-ABT required no training data and it took much less time for the training. In total, HyperLISTA-ABT speeded up

the processing by about one order of magnitude compared to the other algorithms tested in the experiment.

Upon evaluating the performance and efficiency, it was observed that HyperLISTA-ABT achieved comparable performance to existing state-of-the-art methods while significantly improving the computational efficiency by approximately one order of magnitude. This is especially advantageous in the multi-component D-TomoSAR case. The application of the aforementioned deep learning-based algorithms and SL1MMER are very limited in the D-TomoSAR case due to the need of time-consuming model training and the heavy computational expense. On the contrary, the application of HyperLISTA-ABT can be easily extended to computationally efficient D-TomoSAR processing. Therefore, HyperLISTA-ABT is a more applicable approach for the large-scale processing of real data.

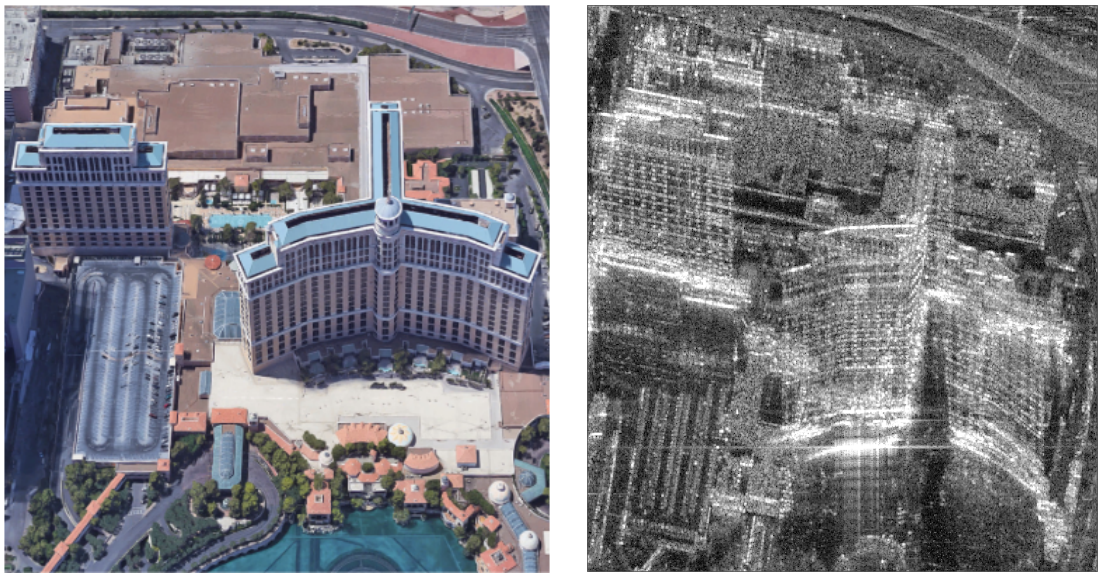
Furthermore, HyperLISTA-ABT demonstrates superior transferability compared to deep learning-based algorithms. Deep learning models are typically trained to fit specific baseline configurations, such as a fixed number of SAR acquisitions and a specific baseline distribution. While they may exhibit satisfactory generalizability to small baseline discrepancies [77] [96], directly applying a trained deep learning model to a new data stack with a different number of acquisitions or a completely different baseline distribution is not feasible. In such cases, time-consuming retraining of the model becomes necessary, resulting in low transferability.

In contrast, HyperLISTA-ABT offers better transferability. Although it requires analytical optimization of the weight matrix for each new data stack, the efficiency of the analytical optimization process allows for scalability and improved transferability. This finding highlights the potential of HyperLISTA-ABT in enabling global urban mapping using TomoSAR, as it can be effectively applied to diverse data stacks with varying acquisition configurations and baseline distributions.

Bellagio hotel

For the real data experiment, the absence of the ground truth means direct verification of accuracy is not possible. However, utilizing the same dataset as in [97] provides us an chance to compare our results with the findings from SL1MMER, thus evaluating the effectiveness of the proposed HyperLISTA-ABT in real-world scenarios. The datastack was composed of 29 TerraSAR-X high-resolution spotlight images covering the Bellagio Hotel in Las Vegas. The slant-range resolution was 0.6m and the azimuth resolution was 1.1m. The elevation aperture size of about 270m resulted in the inherent elevation resolution ρ_s to be about 40m, i.e. approximately 20m resolution in height since the incidence angle here was 31.8°. An optical image and the SAR mean intensity image of the test site are shown in Fig. 4.23. The dataset with fine spatial resolution allows for evaluation on accurately reconstructing topographies with complex scattering mechanisms. The primary goal here is to verify whether our method can reproduce or surpass the quality of reconstructions generated by SL1MMER.

As for the D-TomoSAR system model, a time wrap operation assuming only sinusoidal seasonal motion was adopted as in [43] because no long-term linear motion was observed during the acquisition period of the test area. The results depicted in Fig. 4.24 and Fig. 4.25 underline the effectiveness of the proposed HyperLISTA-ABT in processing real-world TomoSAR data. Several key takeaways can be gathered from the presented results:



(a) (b)

Figure 4.23: Test site of Bellagio hotel. (a): optical image from Google Earth, (b): SAR mean intensity image

- **Reasonable Elevation Estimation**: We can observe a smooth and consistent gradation in the elevation estimates from the ground to the top of the building in Fig. 4.24(a), which is indicative of the capability to capture the inherent vertical structure of urban buildings, thereby ensuring the reasonable 3-D reconstruction.
- **Comparable Performance with SL1MMER**: The close resemblance between the results of HyperLISTA-ABT and SL1MMER demonstrates that HyperLISTA-ABT is able to achieve performance at par compared to SL1MMER, thus establishing the credibility of HyperLISTA-ABT is real-world TomoSAR processing.
- **Layover Separation Ability**: Fig. 4.25 delves deeper into one of the most challenging aspects of TomoSAR processing: the layover effect, particularly prevalent in urban areas with high-rises. The detection of double scatterers demonstrates the layover separation ability of HyperLISTA-ABT. As can be seen, the top layer was mainly caused by signals from the roof and facade of the high rise building while the bottom layer was caused by signals from the ground structures.

Table 4.4: Percentage of scatterers detection for HyperLISTA-ABT and SL1MMER.

Algorithm	Percentage of detection as		
	0 scatterer	1 scatterer	2 scatterers
HyperLISTA-ABT	48.48 %	44.09 %	7.43 %
SL1MMER	49.41 %	43.63 %	6.96 %

We also conducted some numerical comparisons of both algorithms. First, we compared the percentage of pixels detected as zero, one, and two scatterers by both algorithms.

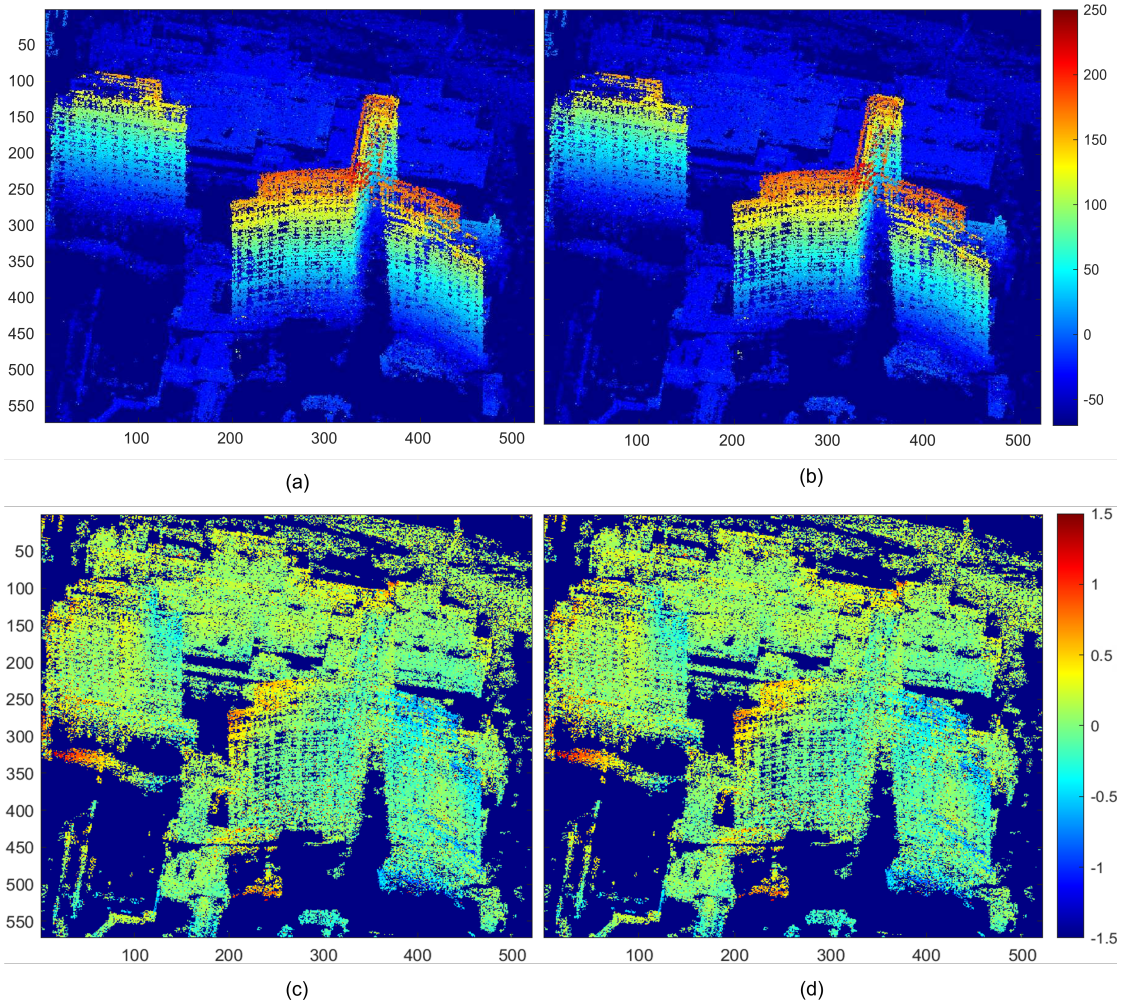


Figure 4.24: Color-coded reconstruction results of the test site. (a) Elevation estimates using HyperLISTA-ABT in meters, (b) elevation estimates using SL1MMER in meters, (c) estimated amplitude of seasonal motion using HyperLISTA-ABT in centimeters, (d) estimated amplitude of seasonal motion using SL1MMER in centimeters.

Compared to SL1MMER, we found that HyperLISTA-ABT detected more pixels as coherent sactterers. This does not necessarily mean that HyperLISTA-ABT had a better detection ability since there was no ground truth. We believe HyperLISTA-ABT detected more scatterers because HyperLISTA-ABT tends to maintain weak signals, which could be reflections of dark scatterers but also outliers caused by noise interference. The false detection of noise as coherent scatterers causes a speckle-like noise in the reconstruction result. Model order selection and post-processing techniques like spatial filtering can further mitigate such outliers.

For further evaluation, we compared the elevation estimates differences of scatterers detected by both algorithms. A histogram of the elevation estimates differences is shown in Fig. 4.26. It can be observed that most of the elevation estimates differences were within 1 meter. This observation indicates that both algorithms yielded comparable results in terms of elevation estimation, instilling confidence in their reliability and reasonableness. Furthermore, this similarity in estimation accuracy suggests that HyperLISTA-ABT per-

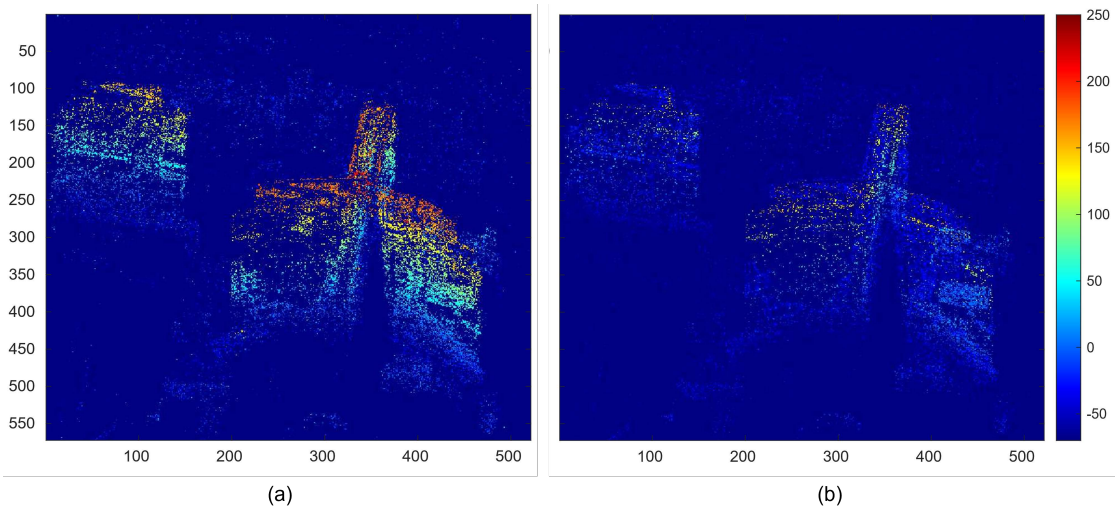


Figure 4.25: Color-coded elevation estimates of the top and bottom layers of detected double scatterers using HyperLISTA-ABT. (a) Top layer, mostly caused by reflections from the building roof and facade, (b) bottom layer, mostly caused by reflections from low infrastructures and the ground.

formed on par with SL1MMER. Moreover, it is worth mentioning that it took more than three weeks for SL1MMER to finish the D-TomoSAR processing over the test site, whereas it only took HyperLISTA-ABT several hours to complete the processing, thus highlighting the great superiority of HyperLISTA-ABT in computational efficiency and potential in global tasks.

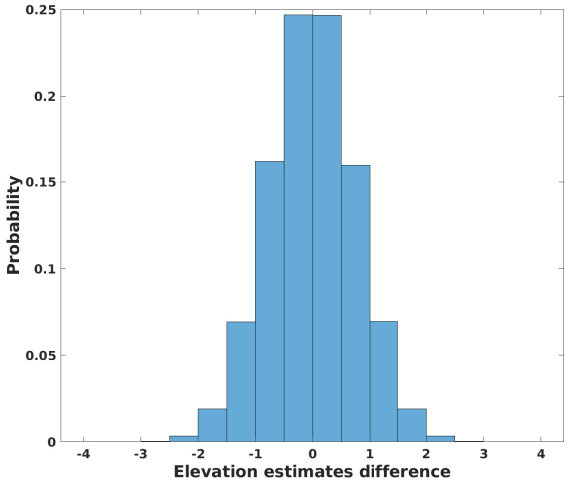


Figure 4.26: Histogram of elevation estimates differences between HyperLISTA-ABT and SL1MMER.

Muti-component D-TomoSAR processing

In this experiment, we applied HyperLISTA-ABT to TerraSAR-X high-resolution spotlight data over a large area surrounding the convention center in Las Vegas. The stack was composed of 29 acquisitions covering a time period from July 2009 to June 2010, during which the test area was undergoing a pronounced subsidence centered at the convention center. Therefore, the acquisitions were characterized by a multi-component nonlinear motion combining linear and thermal-dilation-induced seasonal motion. Fig. 4.27 shows us an optical image and the SAR mean intensity map. The red box indicates the "epiccenter" undergoing subsidence around the convention center.

Figure 4.27: Demonstration of the large test area. (a) Optical image from Google Earth, (b) SAR mean intensity map in dB. The red box in (b) indicates the area undergoing subsidence.

Fig. 4.28 illustrates us the reconstructed elevation estimates as well as the estimated amplitude maps of the two different motions. As we can see from the surface model generated from the elevation estimates in Fig. 4.28(a), we can capture the shapes of individual buildings and the surrounding infrastructures, like roads, at a detailed level. In addition, Fig. 4.28(b) shows that clear deformation caused by seasonal motion can be observed in the metallic building structures since they were affected by thermal dilation more seriously compared to surrounding infrastructures. Furthermore, as illustrated in Fig. 4.28(c), it could be observed that the magnitude of the linear subsidence increased as the scatterer getting closer to the "epicenter". These results are consistent with the fact, thus validating the effectiveness of HyperLISTA-ABT for multi-component nonlinear motion estimation and giving confidence that HyperLISTA-ABT can be applied in large-scale D-TomoSAR processing.

A large-scale demonstration

In this experiment, we demonstration the reconstruction result over a large-scale scene in Munich with the size of 21 km $\times$ 10 km using the proposed HyperLISTA-ABT. The data stack is composed of five co-registered Tandem-X bistatic interferograms, whose most pertinent parameters are presented in Table 4.5.

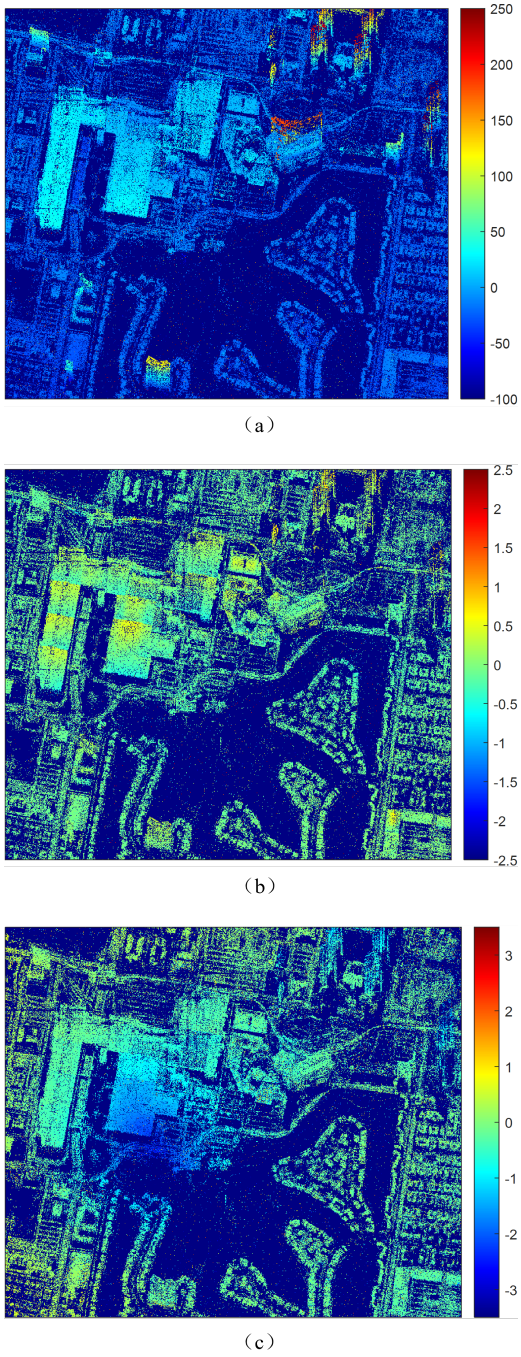


Figure 4.28: Demonstration of color-coded elevation estimates and estimated amplitude of multi-component motion. (a) Elevation estimates in meters, (b) estimated amplitude of seasonal motion in centimeters, (c) estimated amplitude of linear motion in centimeters/year.

Table 4.5: Parameters of Tandem-X acquisitions of Munich

Name	Value
Distance from the scene center	698 km
Wavelength	3.1 cm
Incidence angle at scene center	50.4°
Maximal elevation aperture	187.18 m
Number of interferograms	5
Slant range resolution	1.8 m
Azimuth resolution	3.3 m

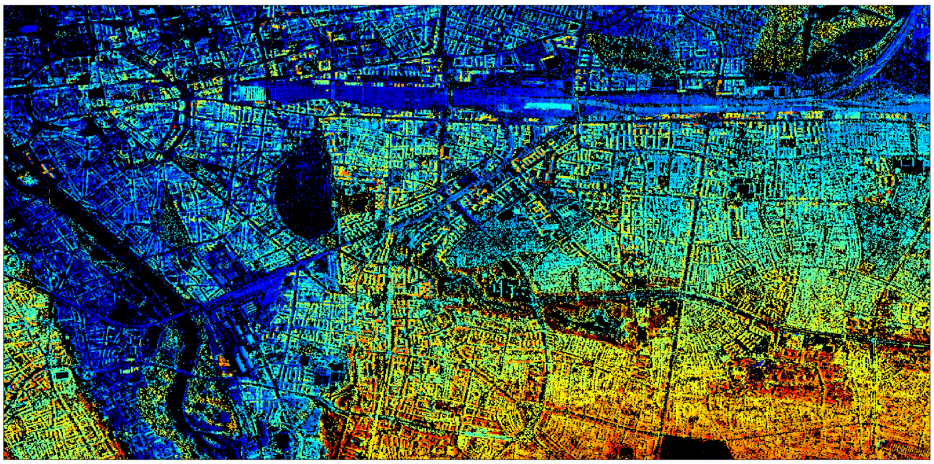
Because of the lack of available ground truth, the result derived from the random block proximal gradient (RBPG) algorithm [98] [99] serves as the benchmark. The RBPG algorithm is currently regarded as the state-of-the-art for efficient and accurate TomoSAR processing for large-scale areas.

The SAR mean intensity map for the test area is showcased in Fig. 4.29(a). Subsequent figures, namely Fig. 4.29(b) and Fig. 4.29(c), offer color-coded elevation reconstructions derived from HyperLISTA-ABT and RBPG, respectively. The black regions within these reconstructions signify scatterers that are non-coherent. The two regions outlined by red boxes in Fig. 4.29(a) indicate specific regions of interest for closed view of detailed building reconstruction, enabling a detailed comparative analysis of building reconstructions by the two algorithms.

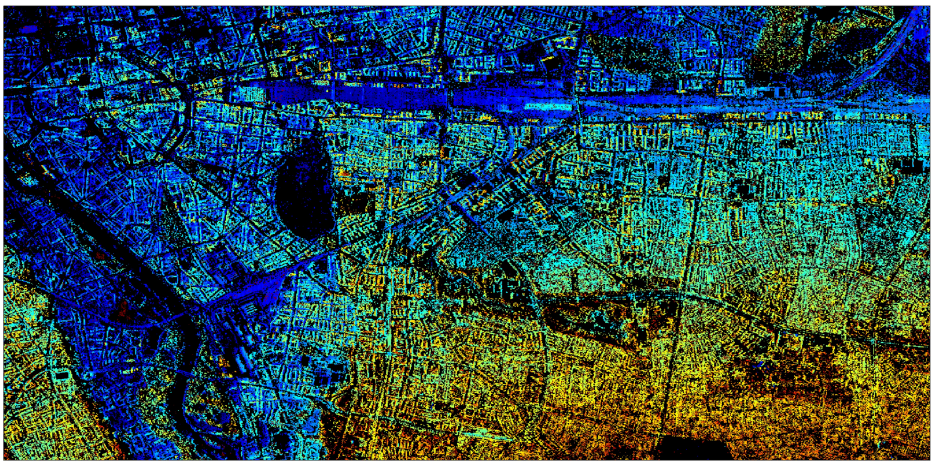
When comparing Fig. 4.29(b) to Fig. 4.29(c), it is evident that both algorithms yield reconstructions of notable quality. The color-coded elevation maps, generated by both HyperLISTA-ABT and RBPG, give valuable insights into the elevation features of the test area. However, upon a more detailed examination, the HyperLISTA-ABT-derived results exhibit superior completeness. This improved clarity and resolution can be further observed in Fig. 4.30 and Fig. 4.30.

Moreover, the computational efficiency of HyperLISTA-ABT stands out prominently even benchmarked against the RBPG, which is already known as an efficient algorithm. The remarkable difference in processing times — one CPU hour for HyperLISTA-ABT compared to the more than three CPU hours for RBPG — signifies a two-fold or greater acceleration in processing speed. The great comuptational efficiency coupled with the promising reconstruction results demonstrated earlier, underscores the immense potential of HyperLISTA-ABT, when we anticipate the future of global urban mapping via TomoSAR.

(a)



(b)



(c)

Figure 4.29: Visual comparison of color-coded reconstruction results over Munich,Germany. Scene size: 21 km × 10 km, north = top. The red bounding boxes indicate two specific regions of interest (ROI) over the area of European bureau of patent and Munich central station. (a) SAR mean intensity image, (b) color-coded reconstruction result derived by HyperLISTA-ABT, (c) color-coded reconstruction result derived by RBPG.

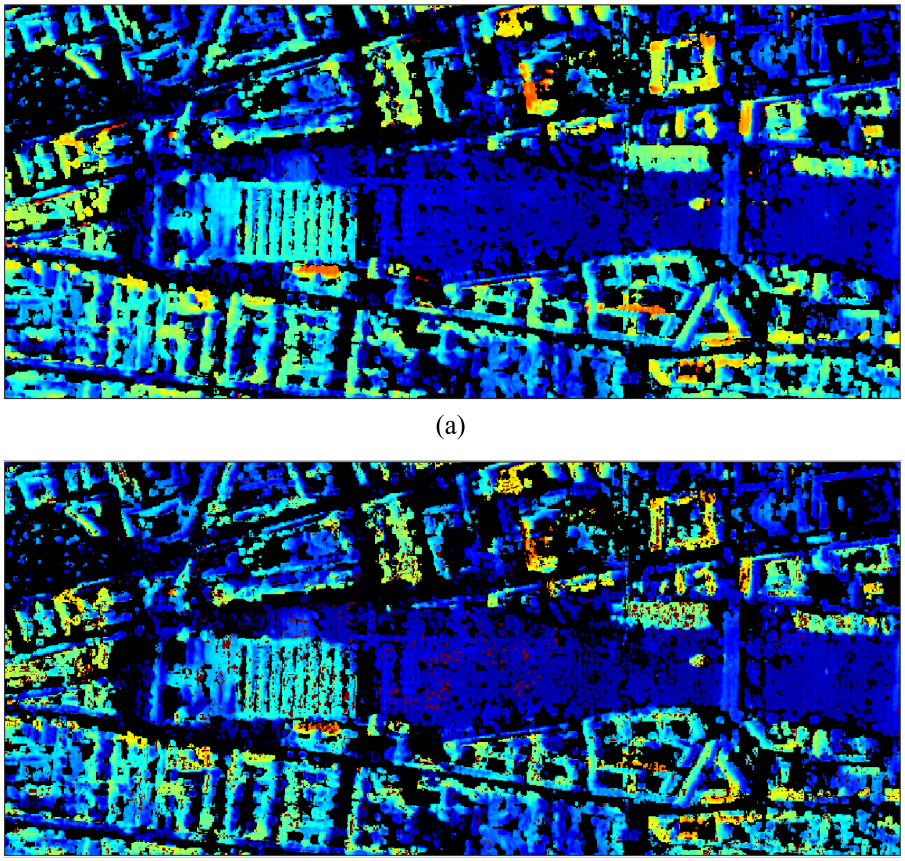


Figure 4.30: Visual comparison of color-coded reconstruction results of the ROI near Munich central station. (a) color-coded reconstruction result derived by HyperLISTA-ABT, (b) color-coded reconstruction result derived by RBPG.

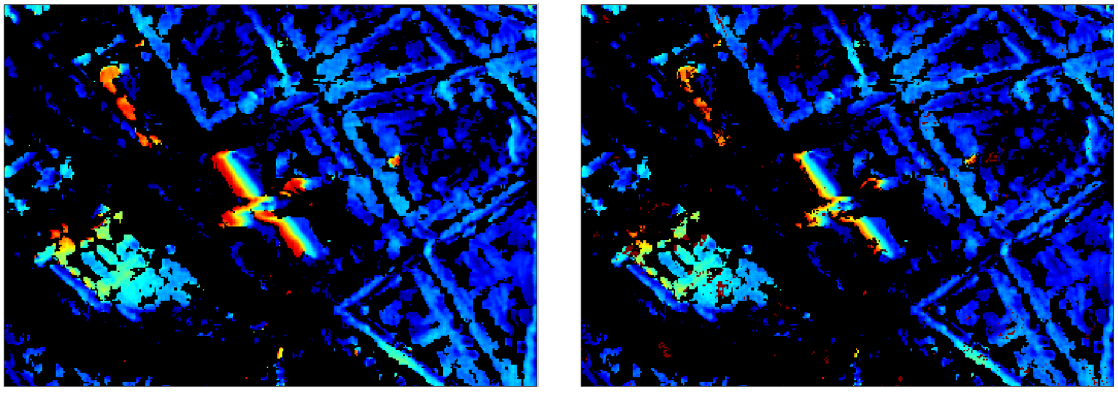


Figure 4.31: Visual comparison of color-coded reconstruction results of the ROI over the area of European bureau of patent. (a) color-coded reconstruction result derived by HyperLISTA-ABT, (b) color-coded reconstruction result derived by RBPG.

5 Conclusion and Outlook

5.1 Conclusion

In a summary, this thesis explores and investigates modern deep learning techniques for the task of efficient and accurate TomoSAR inversion with the potential for practical applications. It delves deep into challenges, presents innovative solutions, and envisions a future where deep learning-based TomoSAR inversion is not just a possibility but a widely accepted norm. With reference to the work presented in this thesis, the following conclusions can be drawn:

- γ-Net is able to efficiently solve TomoSAR inversion. Compared to CS-based TomoSAR algorithms, γ-Net improves the computational efficiency by 1 to 2 orders of magnitude while showing no significant performance degradation with respect to super-resolution power and location accuracy.
- At limited number of SAR acquisitions, deep learning-based algorithms, e.g. γ-Net, outperform CS-based algorithms by a fair margin.
- The impressive computational efficiency and elevation reconstruction results achieved by γ-Net not only validate the concept but also underscore the great potential of leveraging deep learning-based algorithms in large-scale TomoSAR processing.
- The proposed novel gated unit, sparse minima gated unit (SMGU), effectively incorporates historical information into dynamics of the neural network. To be specific, important information will be automatically accumulated while useless or redundant information will be forgotten in the dynamic of the network.
- SMGU adeptly addresses information loss caused by the shrinkage operation, ensuring comprehensive information delivery to the network output while simultaneously promoting output sparsity.
- CV-SMGU, as the extension of SMGU to the complex-valued domain and the first attempt to bridge gated RNN and TomoSAR inversion, showcases superior performance for TomoSAR processing on both laboratory and real-world data, surpassing the state-of-the-art methods.
- An ultra-light model called HyperLISTA-ABT is proposed to address the gap in applying deep neural networks to solve multi-component D-TomoSAR processing.
- According to the minimum coherence criteria, the weights in HyperLISTA-ABT can be determined in an analytical way thus avoiding time-consuming training procedure and requirement of numerous training samples.
- Laboratory experiments for 3-D reconstruction confirmed the efficiency of HyperLISTA-ABT in estimation. Moreover, tests on real data over a large area

demonstrated that HyperLISTA-ABT can reconstruct high-quality 4-D point clouds, making it an efficient and accurate algorithm for future large-scale or even global D-TomoSAR processing.

5.2 Outlook

According to the studies of deep learning-based algorithms for TomoSAR inversion presented in this thesis, a few topics for the future research are outlined in the following:

Deep learning-based gridless approach

Most current TomoSAR algorithms assume that the scatterers are positioned on predefined, discrete elevation grids. However, this discretization might not always align with the actual physical world, where scatterers can exist at any elevation, unconstrained by the pre-imposed grid points. This mismatch between the assumed grid and the actual scatterer location introduces the "off-grid" bias in the elevation estimate. As a gridless CS approach, atomic norm minimization (ANM) [100] [101] [102] [103] can avoid the "off-grid" bias but requires enormous computing resources. This is where the transformative power of deep learning, combined with deep unfolding, could potentially revolutionize the TomoSAR inversion paradigm. By leveraging the representational capabilities of deep networks, it might be possible to mimic the precision of gridless approaches like ANM while significantly reducing the computational overhead.

Deep learning for interferometric phase filtering

As it was comprehensively investigated in [13], the product of the number of SAR acquisitions and SNR determines the quality of TomoSAR reconstruction. However, the number of SAR acquisitions is limited (3-5) for most cities. Therefore, phase filtering is essential to improve the SNR level. The state-of-the-art denoising method is based on non-local means filtering [104] [105] [106], while it suffers from computational expense especially when dealing with high-resolution SAR data. Considering the strong representational and computational power of deep learning, it is promising to develop deep leaning-based methods for interferometric phase filtering. The research focus can be:

- By training models on pairs of noisy and non-local means filtered SAR images, the network can learn the transformations implicitly performed by the non-local means filter, thus efficiently mimicking the operation of non-local means filter.
- With the assistance of generative models, such as variational auto-encoder (VAE) [107], it is possible for us to learn the distribution of real InSAR data in an unsupervised manner. Thereafter, sampling from learned distribution generates new and filtered interferograms.

Incorporation of spatial context

Similar to CS-based TomoSAR algorithms, current deep learning-based algorithms process each range-azimuth resolution cell independently, leading to an isolated 1-D inversion. While this simplifies the inversion process, it neglects the spatial context and relationships

between neighboring resolution cells. This limitation often results in loss of weak scatterers and lack of target continuity. Incorporating spatial context into deep learning-based TomoSAR inversion should improve the performance. By leveraging modern deep learning techniques that recognize and utilize spatial relationships, such as convolutional and recurrent layers, as well as attention mechanisms, it is possible to achieve more accurate, continuous, and interpretable TomoSAR reconstructions.

Deep learning-based TomoSAR in forested areas

Forested regions, characterized by multi-layered vertical structures, result in volumetric scattering mechanisms and make the reflectivity profiles non-sparse. Therefore, CS-based sparse reconstruction is unfortunately not suitable for solving TomoSAR inversion in forested areas. It is essential to conduct transformation to represent the continuous backscattering profile of forest with a set of sparse coefficients after projecting to a certain orthogonal basis. Modern deep learning techniques can assist in both the transformation process and the selection of the optimal basis. Networks can be trained to recognize specific forest structures and adaptively choose or even generate the best basis for projection.

Bibliography

[1] X. X. Zhu. *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring — A Sparse and Nonlinear Tour*. PhD thesis, Technische Universität München, May 2011.

[2] R. Bamler and P. Hartl. Synthetic aperture radar interferometry. *Inverse Problems*, 14(4):R1, aug 1998. URL: `https://dx.doi.org/10.1088/0266-5611/14/4/001`, `doi:10.1088/0266-5611/14/4/001`.

[3] V. Monga, Y. Li, and Y. C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. `doi:10.1109/MSP.2020.3016905`.

[4] Y. Wang and X. X. Zhu. Sar tomography via nonlinear blind scatterer separation. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5751–5763, 2021. `doi:10.1109/TGRS.2020.3022209`.

[5] J. C. Curlander and R. N. Mcdonough. Synthetic aperture radar: Systems and signal processing. 1991.

[6] I. G. Cumming and F. H. Wong. Digital processing of synthetic aperture radar data: Algorithms and implementation. *Artech House*, 2005.

[7] R. Werninghaus and S. Buckreuss. The terrasar-x mission and system design. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2):606–614, 2010. `doi:10.1109/TGRS.2009.2031062`.

[8] F. Caltagirone, A. Capuzi, A. Coletta, G. F. De Luca, E. Scorzafava, R. Leonardi, S. Rivola, S. Fagioli, G. Angino, M. L'Abbate, M. Piemontese, E. Zampolini Faustini, A. Torre, C. De Libero, and P. G. Esposito. The cosmo-skymed dual use earth observation program: Development, qualification, and results of the commissioning of the overall constellation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(7):2754–2762, 2014. `doi:10.1109/JSTARS.2014.2317287`.

[9] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. `doi:10.1109/TIT.2006.871582`.

[10] R. G. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4):118–121, 2007. `doi:10.1109/MSP.2007.4286571`.

[11] E. J. Candes and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008. `doi:10.1109/MSP.2007.914731`.

[12] X. X. Zhu and R. Bamler. Tomographic sar inversion by l_1 -norm regularization—the compressive sensing approach. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3839–3846, 2010. `doi:10.1109/TGRS.2010.2048117`.

[13] X. X. Zhu and R. Bamler. Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic sar. *IEEE Transactions on Geoscience and Remote Sensing*, 50(1):247–258, 2012. `doi:10.1109/TGRS.2011.2160183`.

[14] A. Budillon, A. Evangelista, and G. Schirinzi. Sar tomography from sparse samples. In *2009 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, pages IV–865–IV–868, 2009. `doi:10.1109/IGARSS.2009.5417514`.

[15] 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 Geoscience and Remote Sensing Letters*, 12(9):1853–1857, 2015. doi:10.1109/LGRS.2015.2431254.

[16] X. Wang, F. Xu, and Y. Jin. The iterative reweighted alternating direction method of multipliers for separating structural layovers in sar tomography. *IEEE Geoscience and Remote Sensing Letters*, 14(11):1883–1887, 2017. doi:10.1109/LGRS.2017.2694825.

[17] X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017. doi:10.1109/MGRS.2017.2762307.

[18] X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu, and R. Bamler. Deep learning meets sar: Concepts, models, pitfalls, and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, pages 0–0, 2021. doi:10.1109/MGRS.2020.3046356.

[19] Y. Wu, M. Rosca, and T. Lillicrap. Deep compressed sensing. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6850–6860. PMLR, 09–15 Jun 2019.

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

[21] D. Massonnet and J.-C. Souyris. *Imaging with Synthetic Aperture Radar*. EPFL Press, May 2008. URL: https://doi.org/10.1201/9781439808139, doi:10.1201/9781439808139.

[22] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013.

[23] R. Bamler, M. Eineder, N. A. Adam, X. X. Zhu, and S. Gernhardt. Interferometric potential of high resolution spaceborne sar. *Photogrammetrie Fernerkundung Geoinformation*, (5):403–415, 2009. URL: https://elib.dlr.de/60371/.

[24] H. A. Zebker and R. M. Goldstein. Topographic mapping from interferometric synthetic aperture radar observations. *Journal of Geophysical Research*, 91(B5):4993, 1986. URL: https://doi.org/10.1029/jb091ib05p04993, doi:10.1029/jb091ib05p04993.

[25] D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, and T. Rabaute. The displacement field of the landers earthquake mapped by radar interferometry. *Nature*, 364(6433):138–142, July 1993. URL: https://doi.org/10.1038/364138a0, doi:10.1038/364138a0.

[26] A. K. GABRIEL and R. M. GOLDSTEIN. Crossed orbit interferometry: theory and experimental results from SIR-b. *International Journal of Remote Sensing*, 9(5):857–872, May 1988. URL: https://doi.org/10.1080/01431168808954901, doi:10.1080/01431168808954901.

[27] F. Gatelli, A. Monti Guamieri, F. Parizzi, P. Pasquali, C. Prati, and F. Rocca. The wavenumber shift in sar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):855–865, 1994. doi:10.1109/36.298013.

[28] C. Ruf, C. Swift, A. Tanner, and D. Le Vine. Interferometric synthetic aperture microwave radiometry for the remote sensing of the earth. *IEEE Transactions on Geoscience and Remote Sensing*, 26(5):597–611, 1988. doi:10.1109/36.7685.

[29] M. Eineder, N. Adam, R. Bamler, N. Yague-Martinez, and H. Breit. Spaceborne spotlight sar interferometry with terrasar-x. *IEEE Transactions on Geoscience and Remote Sensing*, 47(5):1524–1535, 2009. doi:10.1109/TGRS.2008.2004714.

[30] F. Li and R. Goldstein. Studies of multibaseline spaceborne interferometric synthetic aperture radars. *IEEE Transactions on Geoscience and Remote Sensing*, 28(1):88–97, 1990. doi:10.1109/36.45749.

[31] J. Homer, I. Longstaff, and G. Callaghan. High resolution 3-d sar via multi-baseline interferometry. In *IGARSS '96. 1996 International Geoscience and Remote Sensing Symposium*, volume 1, pages 796–798 vol.1, 1996. doi:10.1109/IGARSS.1996.516478.

[32] P. Pasquali, C. Prati, F. Rocca, M. Seymour, J. Fortuny, E. Ohlmer, and A. Sieber. A 3-d sar experiment with emsl data. In *1995 International Geoscience and Remote Sensing Symposium, IGARSS '95. Quantitative Remote Sensing for Science and Applications*, volume 1, pages 784–786 vol.1, 1995. doi:10.1109/IGARSS.1995.520585.

[33] A. Reigber and A. Moreira. First demonstration of airborne sar tomography using multibaseline l-band data. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2142–2152, 2000. doi:10.1109/36.868873.

[34] G. Fornaro, F. Serafino, and F. Soldovieri. Three-dimensional focusing with multipass sar data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(3):507–517, 2003. doi:10.1109/TGRS.2003.809934.

[35] F. Lombardini, M. Montanari, and F. Gini. Reflectivity estimation for multibaseline interferometric radar imaging of layover extended sources. *IEEE Transactions on Signal Processing*, 51(6):1508–1519, 2003. doi:10.1109/TSP.2003.811239.

[36] X. X. Zhu, N. A. Adam, and R. Bamler. First demonstration of spaceborne high resolution sar tomography in urban environment using terrasar-x data. In *Proc. CEOS SAR Workshop on Calibration and Validation '08*, pages 1–8. CEOS, 2008. URL: https://elib.dlr.de/57586/.

[37] A. Ferretti, C. Prati, and F. Rocca. Analysis of permanent scatterers in sar interferometry. In *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120)*, volume 2, pages 761–763 vol.2, 2000. doi:10.1109/IGARSS.2000.861695.

[38] A. Ferretti, C. Prati, and F. Rocca. Permanent scatterers in sar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39(1):8–20, 2001. doi:10.1109/36.898661.

[39] H. Breit, T. Fritz, U. Balss, M. Lachaise, A. Niedermeier, and M. Vonavka. Terrasar-x sar processing and products. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2):727–740, 2010. doi:10.1109/TGRS.2009.2035497.

[40] W. Pitz and D. Miller. The terrasar-x satellite. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2):615–622, 2010. doi:10.1109/TGRS.2009.2037432.

[41] H. Breit, M. Fischer, U. Balss, and T. Fritz. Terrasar-x staring spotlight processing and products. In *In 10th European Conference on Synthetic Aperture Radar (EUSAR)*, pages 193–196, 06 2014.

[42] F. Lombardini. Differential tomography: a new framework for sar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1):37–44, 2005. doi:10.1109/TGRS.2004.838371.

[43] X. X. Zhu and R. Bamler. Let's do the time warp: Multicomponent nonlinear motion estimation in differential sar tomography. *IEEE Geoscience and Remote Sensing Letters*, 8(4):735–739, 2011. doi:10.1109/LGRS.2010.2103298.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[45] G. Nishida, A. Bousseau, and D. G. Aliaga. Procedural modeling of a building from a single image. *Computer Graphics Forum*, 37(2):415–429, 2018. doi:https://doi.org/10.1111/cgf.13372.

[46] N. Ibtehaz and M. S. Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020.

[47] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing, 2015.

[48] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 399–406, Madison, WI, USA, 2010. Omnipress.

[49] Z. Wang, Q. Ling, and T. Huang. Learning deep l_0 encoders. 30th AAAI Conference on Artificial Intelligence, AAAI 2016, pages 2194–2200. AAAI Press, January 2016.

[50] 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 *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2764–2772, 2016. doi:10.1109/CVPR.2016.302.

[51] Z. Wang, S. Chang, J. Zhou, M. Wang, and T. S. Huang. *Learning A Task-Specific Deep Architecture For Clustering*, pages 369–377. doi:10.1137/1.9781611974348.42.

[52] W. Zhangyang, Y. Yingzhen, C. Shiyu, L. Qing, and T. Huang. Learning a deep l_∞ encoder for hashing. *IJCAI International Joint Conference on Artificial Intelligence*, 2016-January:2174–2180, 2016.

[53] P. Thompson, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz. *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Kluwer Academic Publishers, USA, 1996.

[54] P. Stoica and R. Moses. Spectral analysis of signals. *Prentice Hall*, 01 2005.

[55] S. C. Andreas Reigber, Konstantinos Papathanassiou and A. Moreira. SAR tomography and interferometry for the remote sensing of forested terrain. *Frequenz*, 55(3-4):119–122, March 2001. URL: https://doi.org/10.1515/freq.2001.55.3-4.119, doi:10.1515/freq.2001.55.3-4.119.

[56] M. Nannini and R. Scheiber. Height dependent motion compensation and coregistration for airborne sar tomography. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 5041–5044, 2007. doi:10.1109/IGARSS.2007.4423994.

[57] O. Frey, F. Morsdorf, and E. Meier. Tomographic imaging of a forested area by airborne multi-baseline p-band sar. *Sensors*, 8(9):5884–5896, 2008. URL: https://www.mdpi.com/1424-8220/8/9/5884, doi:10.3390/s8095884.

[58] X. Zhu and R. Bamler. Very high resolution spaceborne sar tomography in urban environment. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12):4296–4308, 2010. 00125. doi:10.1109/TGRS.2010.2050487.

[59] P. Stoica and R. L. Moses. Introduction to spectral analysis. 1997.

[60] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969. doi:10.1109/PROC.1969.7278.

[61] G. Bienvenu. Influence of the spatial coherence of the background noise on high resolution passive methods. In *ICASSP '79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 306–309, 1979. doi:10.1109/ICASSP.1979.1170720.

[62] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986. doi:10.1109/TAP.1986.1143830.

[63] J. Li and P. Stoica. Efficient mixed-spectrum estimation with applications to target feature extraction. *IEEE Transactions on Signal Processing*, 44(2):281–295, 1996. doi:10.1109/78.485924.

[64] G. Fornaro, F. Lombardini, and F. Serafino. Three-dimensional multipass sar focusing: experiments with long-term spaceborne data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):702–714, 2005. doi:10.1109/TGRS.2005.843567.

[65] P. Zhao and B. Yu. On model selection consistency of lasso. *Journal of Machine Learning Research*, 7(90):2541–2563, 2006. URL: http://jmlr.org/papers/v7/zhao06a.html.

[66] K. P. Burnham and D. R. Anderson. Multimodel inference. *Sociological Methods and Research*, 33(2):261–304, November 2004. URL: https://doi.org/10.1177/0049124104268644, doi:10.1177/0049124104268644.

[67] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. URL: http://www.jstor.org/stable/2958889.

[68] G. Fornaro, S. Verde, D. Reale, and A. Pauciullo. Caesar: An approach based on covariance matrix decomposition to improve multibaseline–multitemporal interferometric sar processing. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):2050–2065, 2015. doi:10.1109/TGRS.2014.2352853.

[69] G. Fornaro, A. Pauciullo, D. Reale, and S. Verde. Sar coherence tomography: A new approach for coherent analysis of urban areas. In *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, pages 73–76, 2013. doi:10.1109/IGARSS.2013.6721095.

[70] T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171 – 1220, 2008. doi:10.1214/009053607000000677.

[71] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger. Nl-sar: A unified nonlocal framework for resolution-preserving (pol)(in)sar denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):2021–2038, 2015. doi:10.1109/TGRS.2014.2352555.

[72] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, jul 2006. doi:10.1162/neco.2006.18.7.1527.

[73] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006. doi:10.1126/science.1127647.

[74] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[75] A. Budillon, A. C. Johnsy, G. Schirinzi, and S. Vitale. Sar tomography based on deep learning. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 3625–3628, 2019. doi:10.1109/IGARSS.2019.8900616.

[76] K. Qian, Y. Wang, Y. Shi, and X. X. Zhu. Super-resolving SAR tomography using deep learning. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1–5, 2021.

[77] K. Qian, Y. Wang, Y. Shi, and X. X. Zhu. γ-net: Superresolving SAR tomographic inversion via deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. doi:10.1109/TGRS.2022.3164193.

[78] 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.

[79] S. Osher, Y. Mao, B. Dong, and W. Yin. Fast linearized bregman iteration for compressed sensing and sparse denoising. *Communications in Mathematical Sciences*, 8:93–111, 01 2010. doi:10.4310/CMS.2010.v8.n1.a6.

[80] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink. Tandem-x: A satellite formation for high-resolution sar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3317–3341, 2007. doi:10.1109/TGRS.2007.900693.

[81] F. Rodriguez Gonzalez, N. Adam, A. Parizzi, and R. Brcic. The Integrated Wide Area Processor (IWAP): A Processor for Wide Area Persistent Scatterer Interferometry. Edinburgh, UK, September 2013. 00000.

[82] N. Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151, 1999. doi:https://doi.org/10.1016/S0893-6080(98)00116-6.

[83] M. D. Zeiler. Adadelta: An adaptive learning rate method, 2012. arXiv:1212.5701.

[84] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12(null):2121–2159, jul 2011.

[85] J. T. Zhou, K. Di, J. Du, X. Peng, H. Yang, S. J. Pan, I. W. Tsang, Y. Liu, Z. Qin, and R. S. M. Goh. Sc2net: Sparse lstms for sparse coding. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, pages 4588–4595, New Orleans, Louisiana, Feb. 2018. AAAI.

[86] G.-B. Zhou, J. Wu, C.-L. Zhang, and Z.-H. Zhou. Minimal gated unit for recurrent neural networks. *Int. J. Autom. Comput.*, 13(3):226–234, jun 2016. doi:10.1007/s11633-016-1006-2.

[87] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[88] R. Jozefowicz, W. Zaremba, and I. Sutskever. An empirical exploration of recurrent network architectures. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 2342–2350, 2015.

[89] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber. Lstm: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10):2222–2232, 2017. doi:10.1109/TNNLS.2016.2582924.

[90] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning*, 2014.

[91] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*.

[92] J. Liu, X. Chen, Z. Wang, and W. Yin. ALISTA: Analytic weights are as good as learned weights in LISTA. In *International Conference on Learning Representations*, 2019. URL: https://openreview.net/forum?id=B1lnzn0ctQ.

[93] X. Chen, J. Liu, Z. Wang, and W. Yin. Hyperparameter tuning is all you need for LISTA. In *Advances in Neural Information Processing Systems*, 2021.

[94] Y. Xu and W. Yin. Block stochastic gradient iteration for convex and nonconvex optimization. *SIAM Journal on Optimization*, 25(3):1686–1716, 2015. doi:10.1137/140983938.

[95] Z. Peng, T. Wu, Y. Xu, M. Yan, and W. Yin. Coordinate-friendly structures, algorithms and applications. *Annals of Mathematical Sciences and Applications*, 1(1):57–119, 2016. doi:10.4310/amsa.2016.v1.n1.a2.

[96] K. Qian, Y. Wang, P. Jung, Y. Shi, and X. X. Zhu. Basis pursuit denoising via recurrent neural network applied to super-resolving SAR tomography. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi:10.1109/TGRS.2022.3221185.

[97] X. X. Zhu and R. Bamler. Demonstration of super-resolution for tomographic SAR imaging in urban environment. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8):3150–3157, 2012. doi:10.1109/TGRS.2011.2177843.

[98] 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 Transactions on Geoscience and Remote Sensing*, 56(10):6148–6158, 2018. doi:10.1109/TGRS.2018.2832721.

[99] Y. Shi, R. Bamler, Y. Wang, and X. X. Zhu. Sar tomography at the limit: Building height reconstruction using only 3–5 tandem-x bistatic interferograms. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11):8026–8037, 2020. doi:10.1109/TGRS.2020.2986052.

[100] Y. Chi and M. Ferreira Da Costa. Harnessing sparsity over the continuum: Atomic norm minimization for superresolution. *IEEE Signal Processing Magazine*, 37(2):39–57, 2020. doi:10.1109/MSP.2019.2962209.

[101] S. Li, D. Yang, G. Tang, and M. B. Wakin. Atomic norm minimization for modal analysis from random and compressed samples. *IEEE Transactions on Signal Processing*, 66(7):1817–1831, 2018. doi:10.1109/TSP.2018.2793907.

[102] J. He, H. Wymeersch, and M. Juntti. Channel estimation for ris-aided mmwave mimo systems via atomic norm minimization. *IEEE Transactions on Wireless Communications*, 20(9):5786–5797, 2021. doi:10.1109/TWC.2021.3070064.

[103] X. Wang and F. Xu. Tomographic sar inversion by atomic-norm minimization—the gridless compressive sensing approach. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. doi:10.1109/TGRS.2022.3223524.

[104] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65 vol. 2, 2005. doi:10.1109/CVPR.2005.38.

[105] Y. Shi, X. x. Zhu, and R. Bamler. Nonlocal compressive sensing-based sar tomography. *IEEE Transactions on Geoscience and Remote Sensing*, 57(5):3015–3024, 2019. doi:10.1109/TGRS.2018.2879382.

[106] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. doi:10.1109/TIP.2007.901238.

[107] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014. URL: http://arxiv.org/abs/1312.6114.

Appendices

A K. Qian, Y. Wang, Y. Shi and X. X. Zhu, "Gamma-Net: Superresolving SAR Tomographic Inversion via Deep Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022, Art no. 4706116, doi: 10.1109/TGRS.2022.3164193.

γ-Net: Superresolving SAR Tomographic Inversion via Deep Learning

Kun Qian, Yuanyuan Wang, *Member, IEEE*, Yilei Shi, *Member, IEEE*, and Xiao Xiang Zhu, *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 $\boldsymbol{\gamma}$ 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 SL1MMER [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 "*SL1MMER*." This approach speeds up the processing by pre-classifying the pixels and reducing the percentage of pixels that require SL1MMER 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 $\boldsymbol{\gamma}$-Net. $\boldsymbol{\gamma}$-Net could be trained using data simulated by spatial baselines of given stacked SAR acquisitions. Once well-trained, $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-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., $\boldsymbol{\gamma}$-Net. Experiments demonstrate that $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-Net in computational efficiency and evidences that $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-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 nth acquisition is the integral of the reflected signal along the elevation direction and can be expressed as follows:

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

where $\gamma(s)$ depicts the reflectivity profiles along the elevation direction. $\xi_n = -2b_n/(\lambda r)$ 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

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

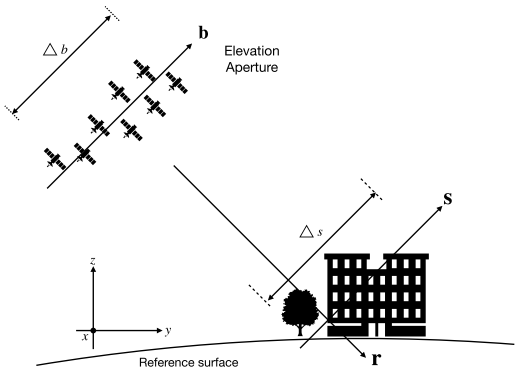


Fig. 1. SAR imaging geometry. The elevation synthetic aperture is built up by SAR data acquired from slightly different viewing angles. Flight direction is orthogonal into the plane.

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 $\boldsymbol{\varepsilon}$, the discrete TomoSAR imaging model can be expressed as follows:

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{3}$$

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 $R_{nl} = \exp(-j2\pi\xi_n s_l)$, 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:

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \left\{ \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \right\} \tag{4}$$

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.

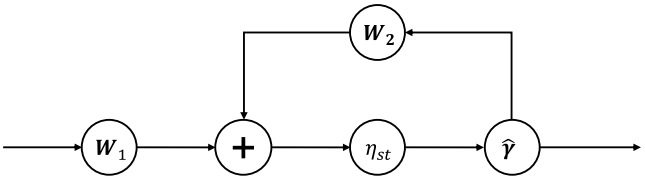


Fig. 2. RNN structure of ISTA by viewing an iteration of ISTA as a layer of the RNN.

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

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

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

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

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:

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

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

If we regard the soft-thresholding function in (7) as an activation function, we find that (7) is the basic form of the ith 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:

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

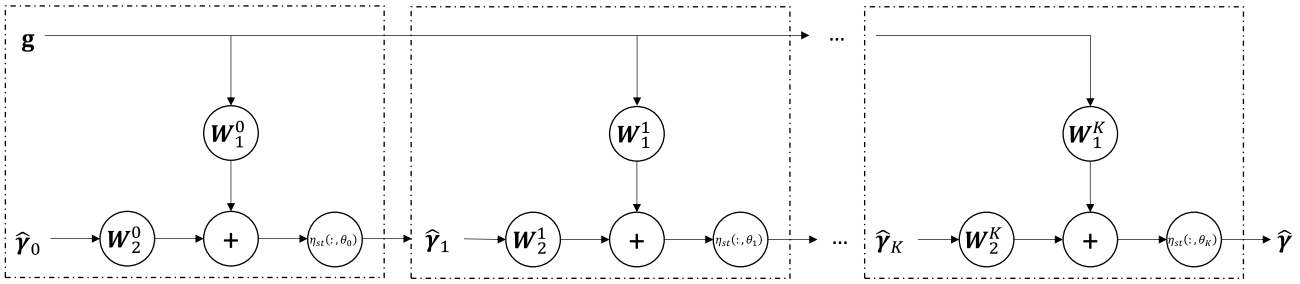


Fig. 3. Unfolded LISTA architecture. A K-layer LISTA unrolls the RNN and truncates it into K iterations, thus leading to a side-connected feedforward neural network.

where T denotes the number of samples in the training data and $\boldsymbol{\Psi} = \{\mathbf{W}_1, \mathbf{W}_2, \boldsymbol{\theta}\}$ 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 adaptions to (7):

$$\tilde{\boldsymbol{\gamma}}_i = \eta_{st}\left\{\widetilde{\mathbf{W}}_1^i \tilde{\mathbf{g}} + \widetilde{\mathbf{W}}_2^i \tilde{\boldsymbol{\gamma}}_{i-1}, \theta_i\right\} \tag{9}$$

where

$$\begin{aligned} \widetilde{\mathbf{W}}_j^i &= \begin{bmatrix} \Re\left(\mathbf{W}_j^i\right) & -\Im\left(\mathbf{W}_j^i\right) \\ \Im\left(\mathbf{W}_j^i\right) & \Re\left(\mathbf{W}_j^i\right) \end{bmatrix} \\ \tilde{\mathbf{g}} &= \begin{bmatrix} \Re(\mathbf{g}) \\ \Im(\mathbf{g}) \end{bmatrix} \\ \tilde{\boldsymbol{\gamma}} &= \begin{bmatrix} \Re(\hat{\boldsymbol{\gamma}}) \\ \Im(\hat{\boldsymbol{\gamma}}) \end{bmatrix} \end{aligned} \tag{10}$$

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

C. $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-Net. Second, we tried to use the acceleration technique, support selection, which was originally developed for LASSO to boost the convergence in $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-Net as pure "black-box" networks, we simplify the CV-LISTA and propose $\boldsymbol{\gamma}$-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:

$$\mathbf{W_2^i} = \mathbf{I} - \mathbf{W_1^i}\mathbf{R}. \tag{11}$$

By employing the partial weight coupling structure, we can simplify the ith layer of $\boldsymbol{\gamma}$-Net to

$$\tilde{\boldsymbol{\gamma}}_i = \eta_{\mathrm{st}}\left\{\tilde{\boldsymbol{\gamma}}_{i-1} + \widetilde{\mathbf{W}}^i\left(\tilde{\mathbf{g}} - \widetilde{\mathbf{R}}\tilde{\boldsymbol{\gamma}}_{i-1}\right), \theta_i\right\} \tag{12}$$

with

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

where $(\Re(\mathbf{W^i}), \Im(\mathbf{W^i}), \theta^i)$ are the parameters to be learned in the ith 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 $\boldsymbol{\gamma}$-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 $\boldsymbol{\gamma}$-Net before the shrinkage step. Hereafter, the selected part will be trusted as "true suppor"' 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 ith layer, the support selection can be formally defined as

$$\eta_{\mathrm{ss}\theta_i}^{\rho^i}\left(\tilde{\boldsymbol{\gamma}}_i\right) = \begin{cases} \tilde{\boldsymbol{\gamma}}_i, & i \in \mathcal{S}^{\rho^i}(\tilde{\boldsymbol{\gamma}}) \\ \eta_{\mathrm{st}}\left(\tilde{\boldsymbol{\gamma}}_i, \theta_i\right), & i \notin \mathcal{S}^{\rho^i}(\tilde{\boldsymbol{\gamma}}) \end{cases} \tag{13}$$

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

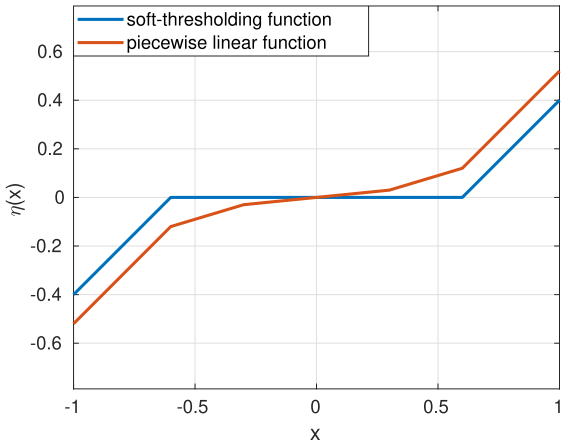


Fig. 4. Comparison between the piecewise linear function and soft-thresholding function. Instead of pruning the elements with small magnitude, the piecewise linear function just further minifies them, thus possibly avoiding the information loss.

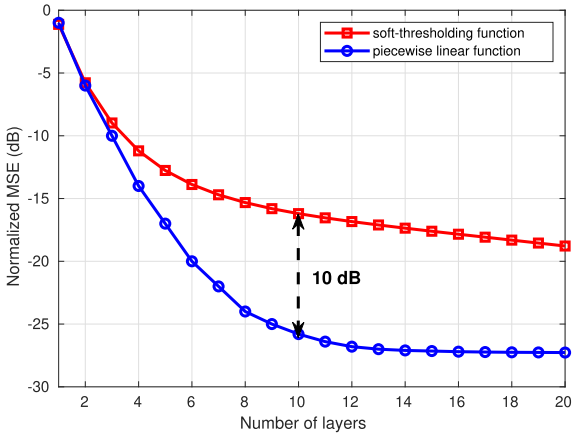


Fig. 5. Performance of γ-Net using different shrinkage function. The piecewise linear function conduces to faster convergence and improves the estimation accuracy.

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

$$\tilde{\gamma}_i = \eta_{\mathrm{ss}_{\theta_i}^{\rho^i}}\left\{\tilde{\gamma}_{i-1} + \widetilde{\mathbf{W}}^i\left(\tilde{\mathbf{g}} - \widetilde{\mathbf{R}}\tilde{\gamma}_{i-1}\right), \theta_i\right\}. \tag{14}$$

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 $\eta_{pwl}(\tilde{\gamma}, \boldsymbol{\theta})$ in γ-Net, which is defined as

$$\eta_{pwl}(\hat{\gamma}) = \begin{cases} \theta_3\hat{\gamma}, & |\hat{\gamma}| \le \theta_1 \\ e^{j\cdot\angle\hat{\gamma}}\left[\theta_4(|\hat{\gamma}| - \theta_1) + \theta_3\theta_1\right], & \theta_1 < |\hat{\gamma}| \le \theta_2 \\ e^{j\cdot\angle\hat{\gamma}}\left[\theta_5(|\hat{\gamma}| - \theta_2) + \theta_4(\theta_2 - \theta_1) + \theta_3\theta_1\right], & |\hat{\gamma}| > \theta_2. \end{cases} \tag{15}$$

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. 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

$$\mathrm{NMSE} = \frac{1}{T}\sum \frac{\|\hat{\gamma} - \gamma\|_2^2}{\|\gamma\|_2^2}. \tag{16}$$

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 $\hat{\gamma}$ [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
  $R_{nl} = \exp(-j2\pi\xi_n s_l)$, 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 $\{(\mathbf{g_i}, \boldsymbol{\gamma_i})\}_{i=1}^{T}$
  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:
    $\boldsymbol{\gamma} = \boldsymbol{\gamma}$-Net$(\mathbf{g})$
    Model order selection to remove outliers:
    $\hat{P} = \underset{P}{\text{argmin}}\{\sigma_\varepsilon^{-2}||\mathbf{g} - \mathbf{R}\hat{\gamma}\,||_2^2 + 1.5\ P \ln N\}$
    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 scatterering 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.

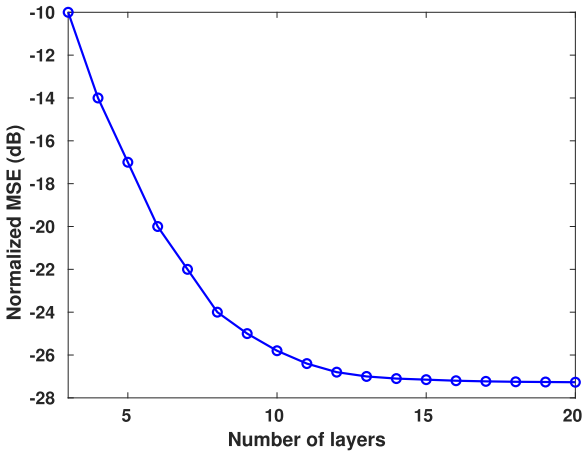


Fig. 6. $\boldsymbol{\gamma}$-Net performance w.r.t. the number of layers. After 12 layers, the performance improvement of $\boldsymbol{\gamma}$-Net is marginal with the increase number of layers. Instead, the increase of layers leads to heavier computational burden.

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

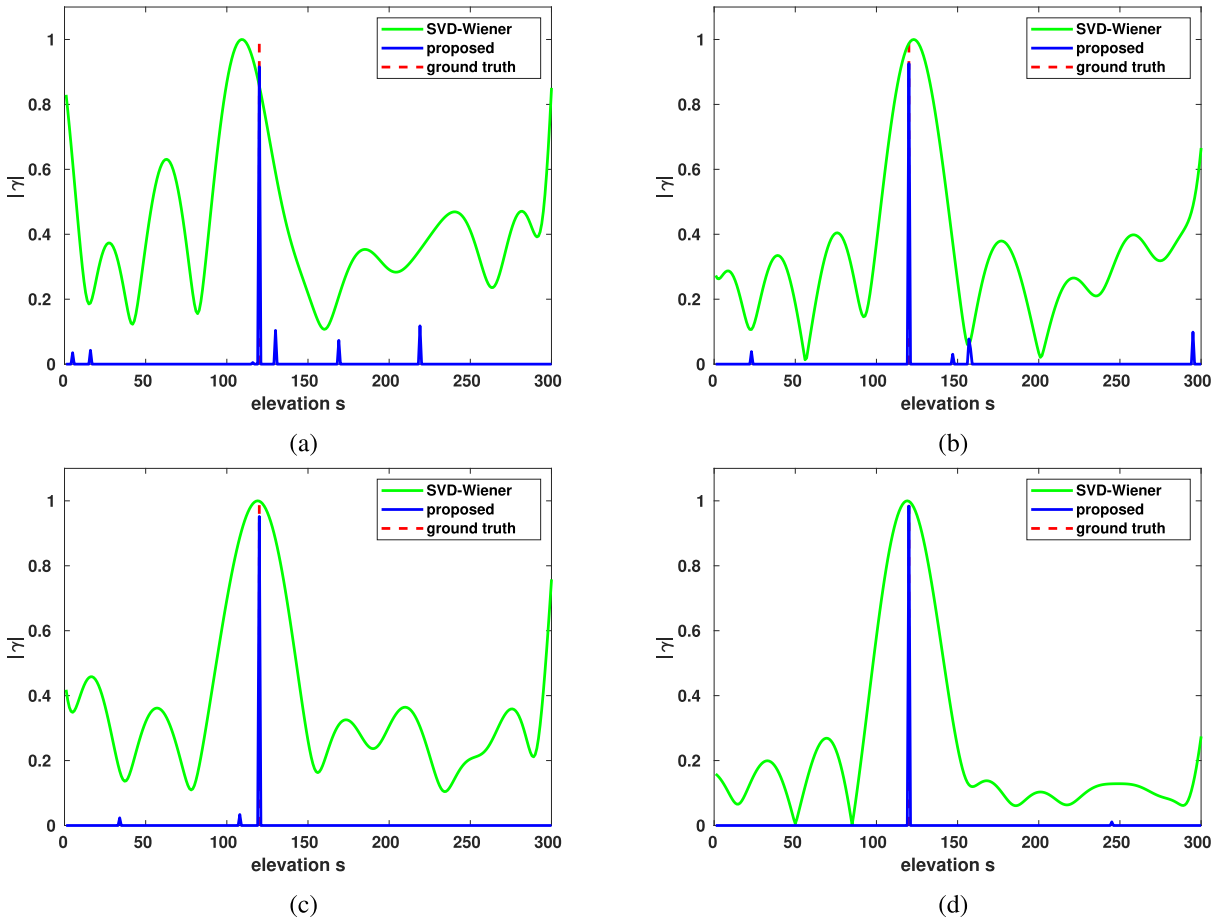


Fig. 7. Estimated reflectivity profile of simulated data with single scatterer at different SNR. (a) SNR = 0 dB. (b) SNR = 3 dB. (c) SNR = 6 dB. (d) SNR = 10 dB.

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

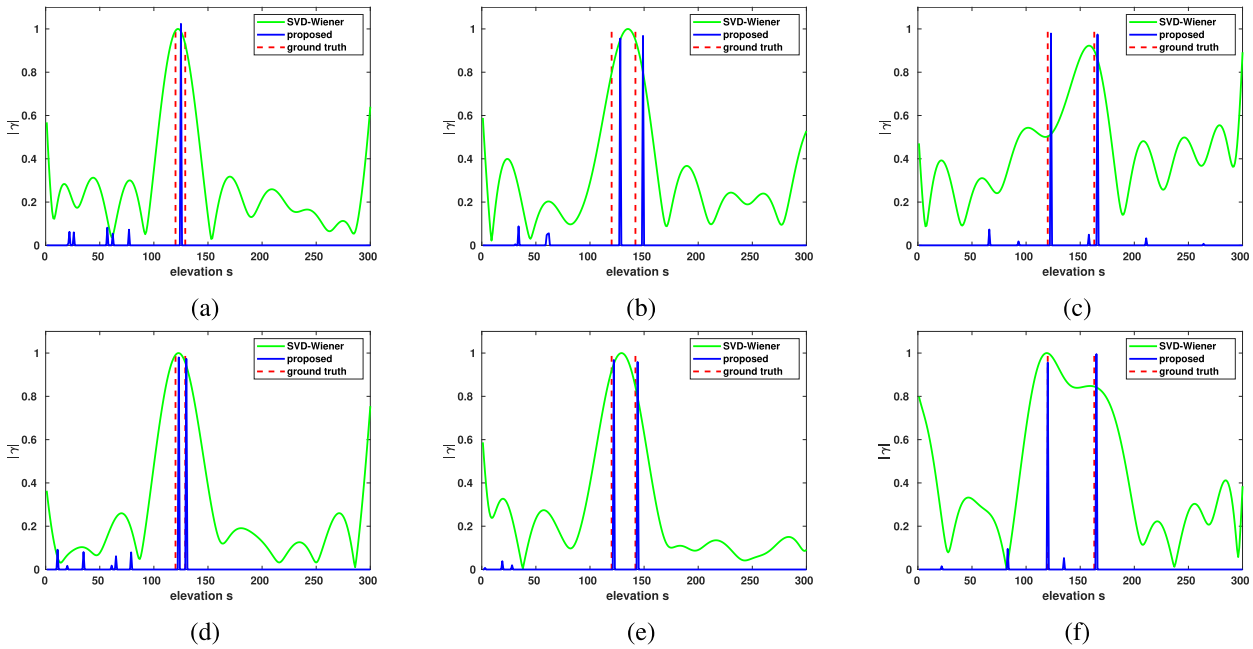


Fig. 8. Estimated reflectivity profile of simulated data with overlaid double scatterers. (a) SNR $= 0$ dB, $\alpha = 0.2$. (b) SNR $= 0$ dB, $\alpha = 0.5$. (c) SNR $= 0$ dB, $\alpha = 1$. (d) SNR $= 6$ dB, $\alpha = 0.2$. (e) SNR $= 6$ dB, $\alpha = 0.5$. (f) SNR $= 6$ dB, $\alpha = 1$.

were taken into consideration with 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

$$\alpha = \frac{d_s}{\rho_s}. \tag{17}$$

From Fig. 8, one can see that both the trained $\boldsymbol{\gamma}$-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, $\boldsymbol{\gamma}$-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, $\boldsymbol{\gamma}$-Net is still capable of detecting the double scatterers in most cases, which exhibits its super-resolution power.

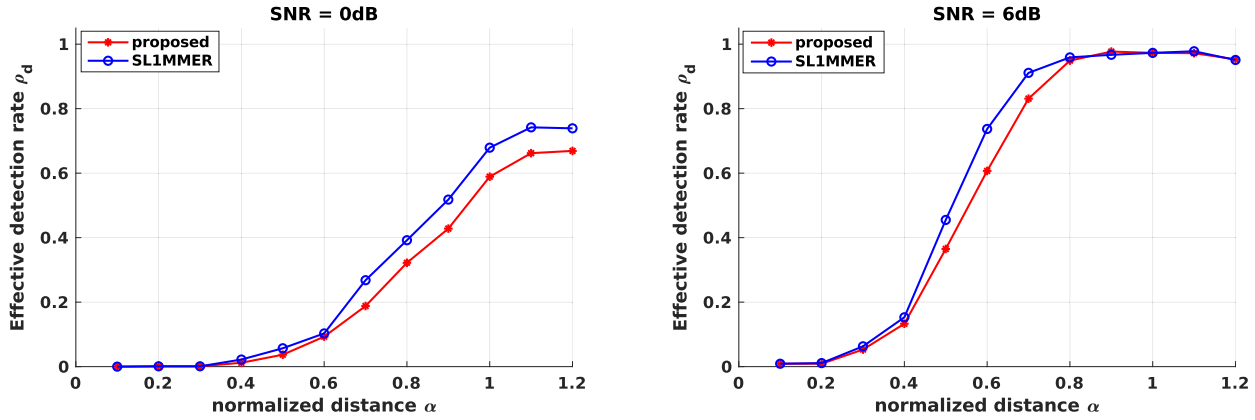
Hereafter, we compare the proposed algorithm with the state-of-the-art SL1MMER algorithm [8] focusing on the detection rate and the estimation accuracy. Similar to the single scatterer case, we use the ***effective detection rate*** to fairly evaluate the detection rate. An effective detection of double 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 ± 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. ± 0.5 d_s is a much stricter constraint in such cases, which will reflect the true performance of the algorithm.

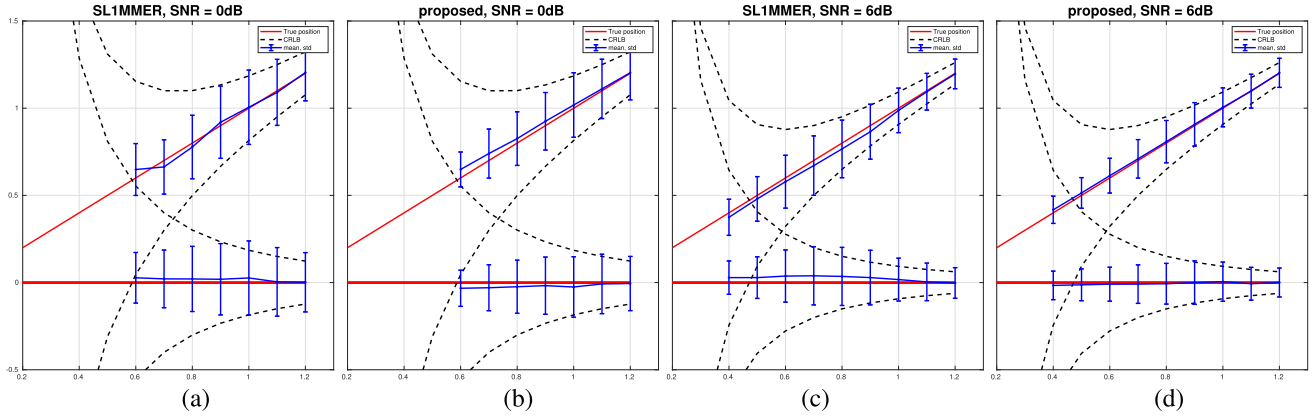
Fig. 9 compares the effective detection rate P_d of SL1MMER and the proposed algorithm for the case $N = 25$. For each pair of (SNR, α), 0.2 million Monte Carlo trials for the worst case in TomoSAR inversion, i.e., the double scatterers have the same amplitude and phase, were simulated. The effective 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 SL1MMER, respectively. As we can see from Fig. 9, the proposed algorithm has comparable super-resolution power as SL1MMER.

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 facade and ground, respectively, while the dashed lines show the true elevation $\pm 1 \times$ CRLB (normalized). Exhaustive details of the derivation of the CRLB can be found in [8]. The elevation estimates of simulated facade 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 SL1MMER. To be specific, at 0 dB SNR, although both the proposed algorithm and SL1MMER have similar estimate bias, the proposed algorithm leads to much smaller variance. In high SNR case, the proposed algorithm outperforms SL1MMER in super-resolving



(a) (b)

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 SL1MMER (dashed circle) with SNR = 0 and 6 dB, $N = 25$ and phase difference $\triangle\phi = 0$ (worst case) under 0.2 million Monte Carlo trials. Comparison in (a) 0 dB and (b) 6 dB.



(a) (b) (c) (d)

Fig. 10. Estimated elevation of simulated facade and ground, (a) SNR = 0 dB with SL1MMER, (b) SNR = 0 dB with the proposed algorithm, (c) SNR = 6 dB with SL1MMER, 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 $\pm 1 \times$ CRLB normalized w.r.t. the Rayleigh resolution.

cases w.r.t. the elevation estimation accuracy. As can be seen that, SL1MMER 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.

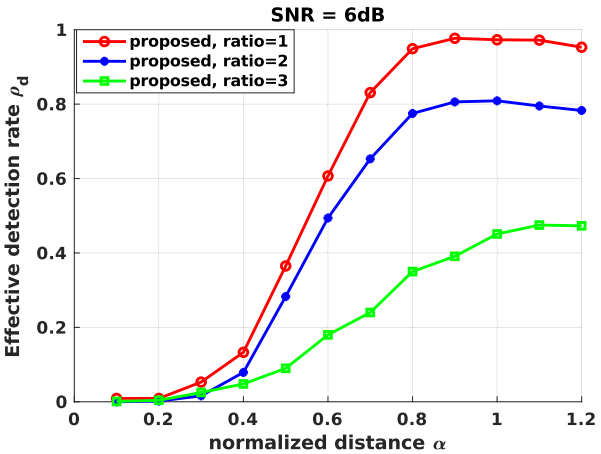


Fig. 11. Effective detection rate ρ_d as a function of amplitude ratio at 6 dB SNR.

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

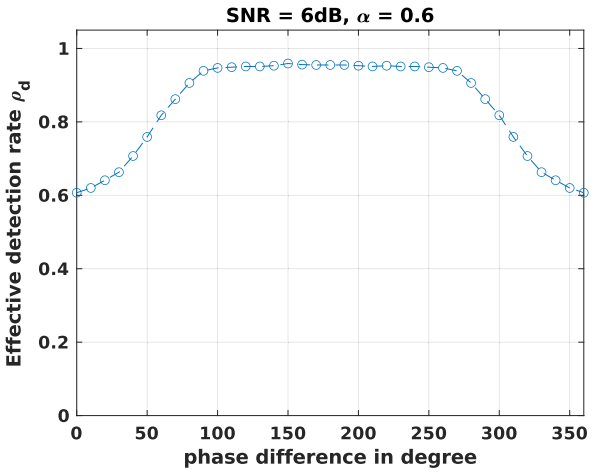


Fig. 12. Effective detection rate ρ_d as a function of phase difference $\triangle\phi$ under the case: $N = 25$, SNR $= 6$ dB and $\alpha = 0.6$.

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 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 $\triangle\phi$ affects the performance, the proposed algorithm is still capable of providing satisfactory super-resolution power even in the worst case, where the phase difference $\triangle\phi = 0$.

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., SNR $= 0$ dB, the two algorithms have similar performance. However, with the increase of the SNR level, the proposed algorithm outperforms SL1MMER by a fair margin. To be specific, the proposed algorithm outperforms SL1MMER 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 ρ_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.

$\boldsymbol{\gamma}$-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, $\boldsymbol{\gamma}$-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 SL1MMER. 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 SL1MMER 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

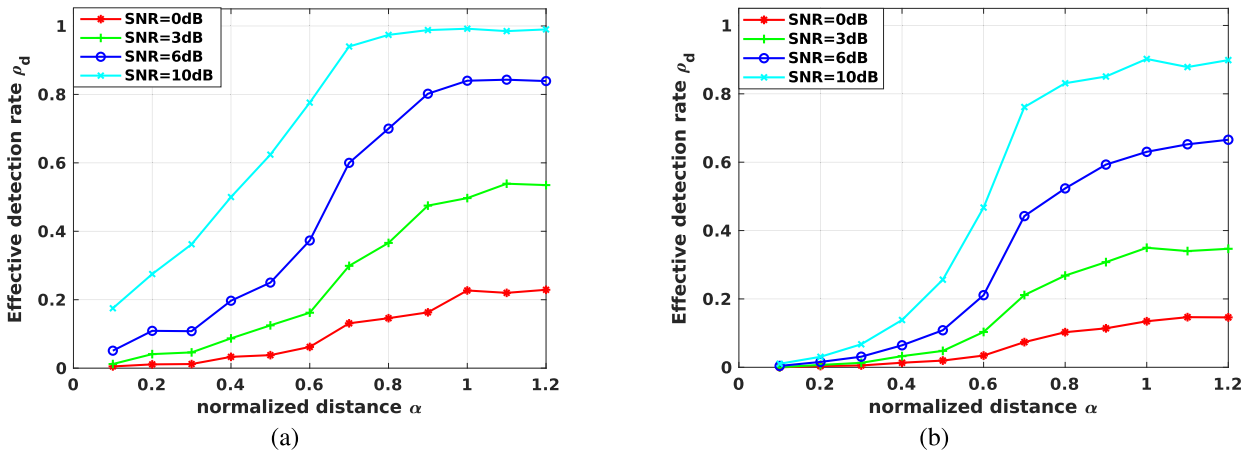


Fig. 13. Effective detection rate P_d as a function of the normalized elevation distance between double scatterers simulated with six real baselines. The simulated double scatterers are set to have identical phase and amplitude, i.e., the worst case. For each pair of (SNR, α), 0.2 million Monte Carlo trials were simulated. (a) Proposed algorithm. (b) SL1MMER.

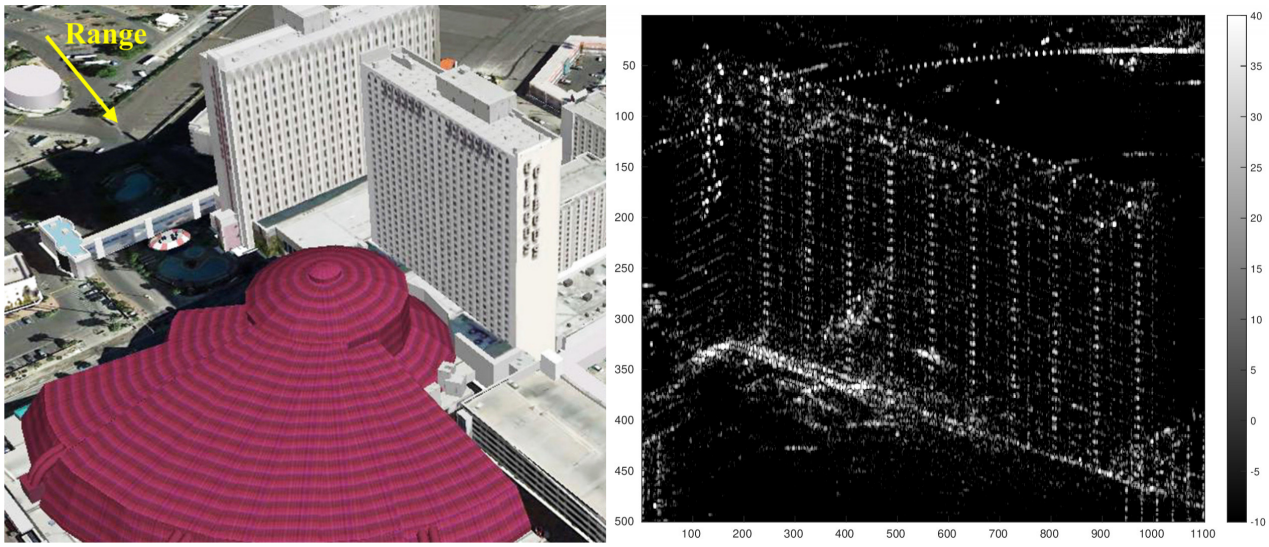


Fig. 14. Test site. (Left) Optical image from Google Earth. (Right) SAR mean intensity image.

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

SL1MMER, the proposed algorithm captures more reflection from the facade. This confirms the finding in Fig. 13 that the proposed algorithm outperforms SL1MMER at low number of measurements.

In terms of detection rate, the proposed algorithm is comparable to that of SL1MMER. 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 SL1MMER, respectively. 36.56% of the pixels were detected as double scatterers by both algorithms. Only 2.4% were only detected double scatterers by the proposed algorithm, and only 1.2% were only detected by SL1MMER.

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 %
SL1MMER	30.41 %	31.83 %	37.76 %

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

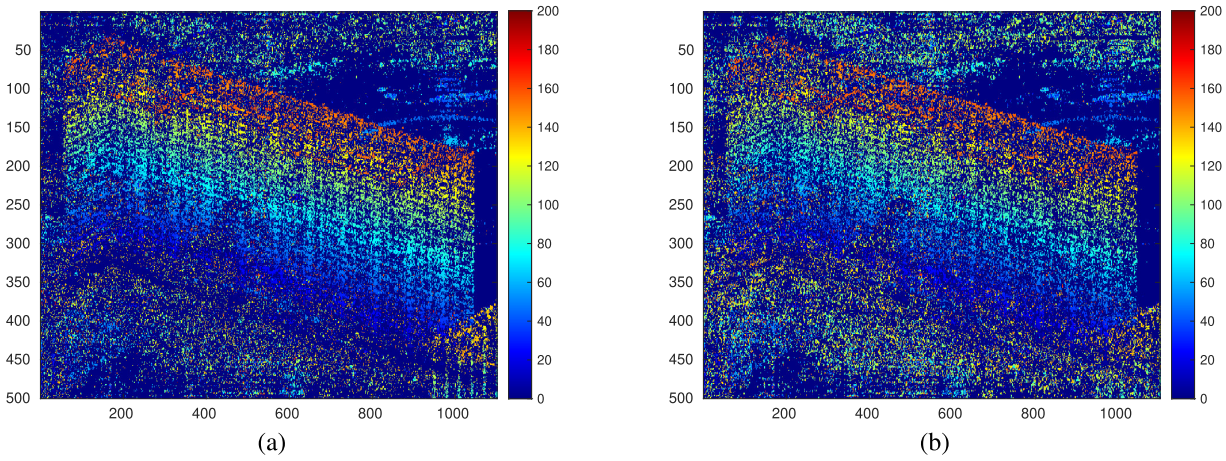


Fig. 15. Reconstructed and color-coded elevation of detected single scatterer. (a) Proposed algorithm. (b) SL1MMER algorithm.

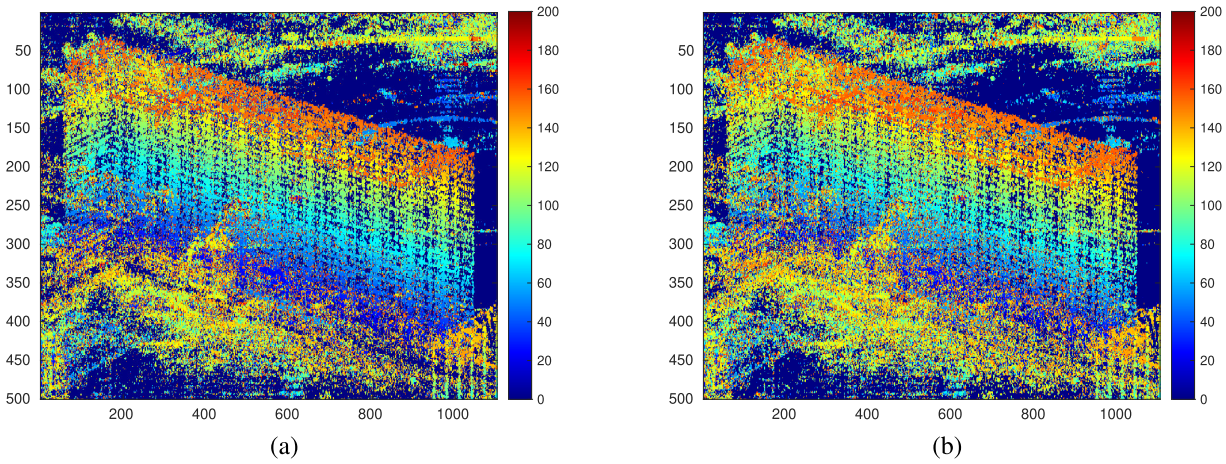


Fig. 16. Reconstructed and color-coded elevation of detected single scatterer + top layer of the detected double scatterers. (a) Proposed algorithm. (b) SL1MMER algorithm.

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, SL1MMER requires about ten CPU hours for the TomoSAR processing. On the contrary, it takes only a few CPU minutes when a trained $\boldsymbol{\gamma}$-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 SL1MMER escalates with the increasing amount of data, especially when limited measurements are available.

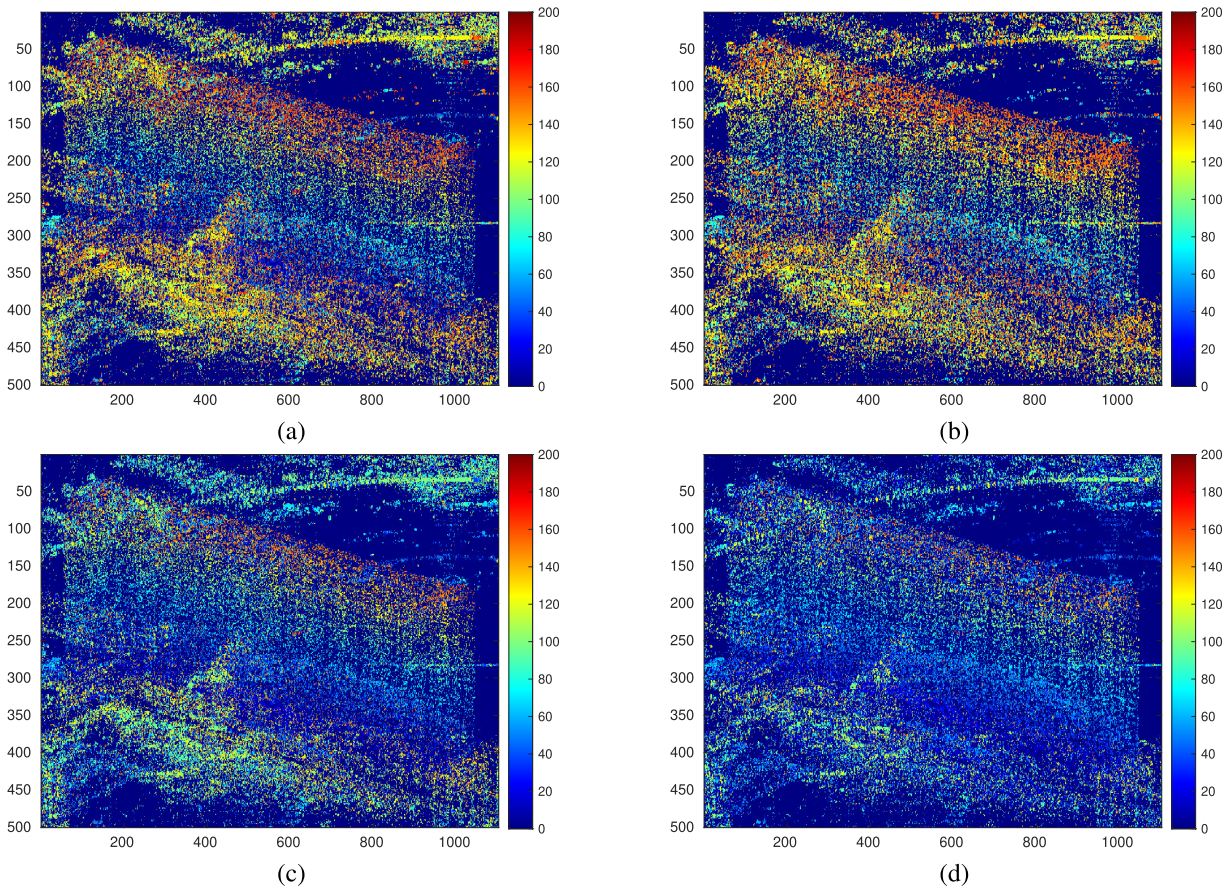


Fig. 17. Reconstructed and color-coded elevation of detected double scatterers. Top layer detected by the (a) proposed algorithm and (b) SL1MMER. Bottom layer detected by the (c) proposed algorithm and (d) SL1MMER.

In real-world TomoSAR processing, the number of pixels is usually tens or even hundreds of million, thus blocking the application of SL1MMER 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 $\mathbf{R^H R}$. 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 ith layer of γ-Net for the support selection

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

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

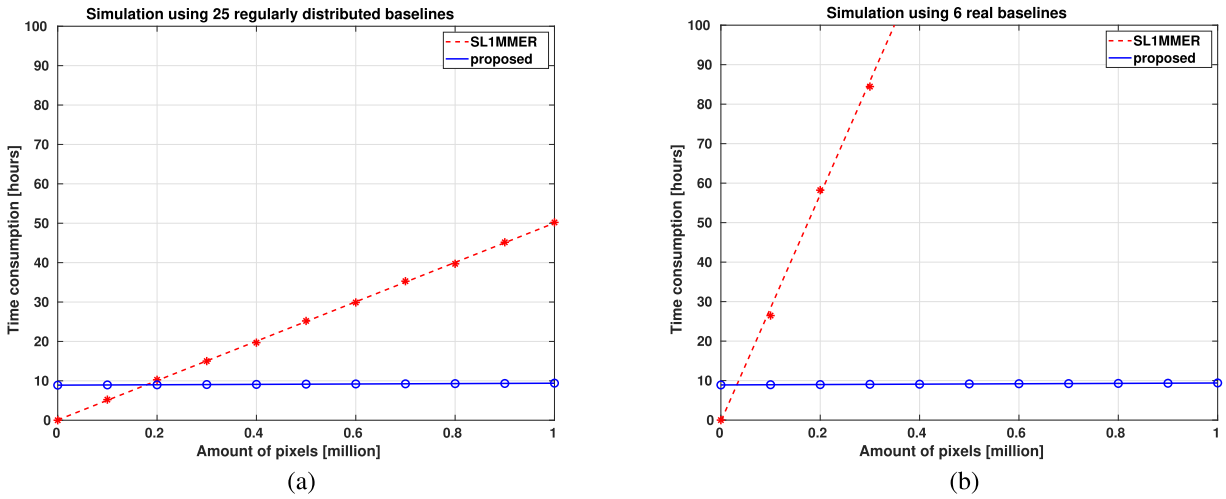


Fig. 18. Comparison of time consumption between the proposed algorithm and SL1MMER. (a) On dataset simulated using 25 regularly distributed baselines at 6 dB. (b) On dataset simulated using six real baselines at 6 dB. The training time will be affected only by the size of training data and the number 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 $\boldsymbol{\gamma}$-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 minifies 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 $\boldsymbol{\gamma}$-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 -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. Schirinzi, "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. Pauciullo, "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 multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 69, pp. 3459–3472, 2021.

[20] A. Budillon, A. C. Johnsy, G. Schirinzi, 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: 10.1561/2200000016.

[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

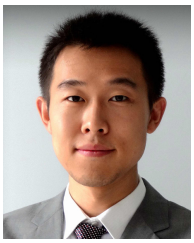
Kun Qian received the B.Sc. degree in remote sensing and information engineering from Wuhan University, Wuhan, China, and the B.Sc. degree in aerospace engineering and geodesy from the University of Stuttgart, Stuttgart, Germany, respectively, in 2016, and the M.Sc. degree in aerospace engineering and 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*.

B K. Qian, Y. Wang, P. Jung, Y. Shi and X. X. Zhu, "Basis Pursuit Denoising via Recurrent Neural Network Applied to Super-Resolving SAR Tomography," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-15, 2022, Art no. 4710015, doi: 10.1109/TGRS.2022.3221185.

Basis Pursuit Denoising via Recurrent Neural Network Applied to Super-Resolving SAR Tomography

Kun Qian, Yuanyuan Wang, *Member, IEEE*, Peter Jung, *Member, IEEE*, Yilei Shi, *Member, IEEE*, and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*— Finding sparse solutions of underdetermined linear systems commonly requires the solving of L_1 regularized least-squares minimization problem, which is also known as the basis pursuit denoising (BPDN). They are computationally expensive since they cannot be solved analytically. An emerging technique known as *deep unrolling* provided a good combination of the descriptive ability of neural networks, explainable, and computational efficiency for BPDN. Many unrolled neural networks for BPDN, e.g., learned iterative shrinkage thresholding algorithm and its variants, employ shrinkage functions to prune elements with small magnitude. Through experiments on synthetic aperture radar tomography (TomoSAR), we discover the shrinkage step leads to unavoidable information loss in the dynamics of networks and degrades the performance of the model. We propose a recurrent neural network (RNN) with novel sparse minimal gated units (SMGUs) to solve the information loss issue. The proposed RNN architecture with SMGUs benefits from incorporating historical information into optimization and, thus, effectively preserves full information in the final output. Taking TomoSAR inversion as an example, extensive simulations demonstrated that the proposed RNN outperforms the state-of-the-art deep learning-based algorithm in terms of super-resolution power and generalization ability. It achieved 10%–20% higher double-scatterer detection rate and is less sensitive to phase and amplitude ratio difference between scatterers. Test on real TerraSAR-X spotlight images also shows the high-quality 3-D reconstruction of the test site.**

***Index Terms*— Basis pursuit denoising (BPDN), recurrent neural network (RNN), sparse reconstruction, synthetic aperture radar tomography (TomoSAR).**

I. Introduction

A. Motivation

SPARSE solutions are ordinarily desired in a multitude of fields, such as radar imaging, medical imaging, and acoustics signal processing. Compressive sensing (CS) theory tells that the exact solution in the absence of noise is the signal with the minimum L_0-norm while still fulfilling the forward model. As the L_0-norm minimization is NP-hard, this is often solved by L_1-norm minimization. The unconstrained form of a linear system can be formulated as follows:

$$\min_x ||\mathbf{Ax} - \mathbf{b}||_2^2 + \lambda||x||_1 \tag{1}$$

where $\mathbf{A}$, $\mathbf{x}$, and $\mathbf{b}$ are the sensing matrix, the signal to be retrieved, and the measurements. Solving (1) is an unconstrained convex optimization problem, whose objective function is nondifferentiable. It is also known as basis pursuit denoising (BPDN) [1]. In the field of remote sensing, sparse signals are widely expected. Therefore, BPDN is broadly employed to exploit sparsity prior in various remote sensing applications, including, but not limited to, pan-sharpening [2], spectral unmixing [3], microwave imaging [4], and synthetic aperture radar tomography (TomoSAR) [5]. In this work, we focus on addressing BPDN in TomoSAR inversion, but our findings are applicable to general sparse reconstruction problems in other fields as well.

Generic solvers for BPDN are either first- or second-order CS-based methods [6], [7], [8]. First-order methods are typically based on a linear approximation of gradient, e.g., the iterative shrinkage thresholding algorithm (ISTA) [9], coordinate descent (CD) [10], and alternating direction method of multipliers (ADMM) [11]. Second-order methods usually have much better performance than first-order methods. An example of the second-order method is the prime dual inferior point method (PDIPM) [12]. It was demonstrated in [5] and [13] that CS-based methods are able to achieve unprecedented super-resolution ability and location accuracy compared to conventional linear algorithm [14], [15]. In spite of the good performance of CS-based methods, they often suffer from heavy computational burdens due to their iterative properties and are hard to extend to practical use.

In the past years, the advent of deep neural networks has attracted the interest of many researchers and triggered extensive studies due to their excellent learning and expression power. Deep neural networks have demonstrated their availability and advanced the state-of-the-art for many problems. More recently, an emerging deep learning algorithm coined ***deep unfolding*** [16] was proposed to provide a concrete and systematic connection between iterative physical model-based algorithms and deep neural networks. Inspired by this concept,

Manuscript received 31 May 2022; revised 9 September 2022; accepted 22 October 2022. Date of publication 2 December 2022; date of current version 20 December 2022. This work was supported in part by the European Research Council under Grant Agreement ERC-2016-StG-714087 (So2Sat), in part by the Helmholtz Association through the Framework of the Helmholtz Excellent Professorship "Data Science in Earth Observation—Big Data Fusion for Urban Research" under Grant W2-W3-100, and in part by the German Federal Ministry of Education and Research in the framework of the International Future AI Lab "AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" under Grant 01DD20001. *(Corresponding author: Xiao Xiang Zhu.)*

Kun Qian, Yuanyuan Wang, Peter Jung, and Xiao Xiang Zhu are with the Chair of Data Science in Earth Observation, Technical University of Munich, 80333 Munich, Germany (e-mail: kun.qian@tum.de; y.wang@tum.de; peter.jung@tum.de; xiaoxiang.zhu@tum.de).

Yilei Shi is with the Chair of Remote Sensing Technology, Technical University of Munich, 80333 Munich, Germany (e-mail: yilei.shi@tum.de).

Digital Object Identifier 10.1109/TGRS.2022.3221185

various neural networks were proposed to solve BPDN in CS problems by unrolling iterative CS solvers. The first work of deep unfolding dates back to learned ISTA (LISTA) [17], which was designed for solving sparse recovery. LISTA unrolls ISTA, one of the most popular algorithms, and substantially improves computational efficiency and parameter tuning. Yang et al. [18] proposed ADMM-CSnet by unrolling the ADMM algorithm to deep hierarchical network architecture and applied ADMM-CSnet to magnetic resonance imaging (MRI) and natural image CS. Results in [18] indicate the favorable performance of ADMM-CSnet in high computational speed. For remote sensing applications, CSR-net [19] was proposed by combining deep unfolding structures and convolutional neural network modules, and achieved fast and accurate 3-D microwave imaging. In addition, Wei et al. [20] proposed AF-AMPNet by unrolling approximate message passing with phase error estimation (AF-AMP) to a deep neural network. AF-AMPNet was employed in sparse aperture (SA) inverse SAR (ISAR) imaging and accelerated the imaging process. Inspired by the encouraging achievements made by deep unfolding, the TomoSAR community started to design deep neural networks by unrolling iterative optimization solvers for solving BPDN in TomoSAR inversion. Gao et al. [21] unrolled and mapped vector AMP (VAMP) [22] into a neural network for line spectral estimation and applied it to tackle TomoSAR inversion. Results in [21] show that L-VAMP is able to separate overlaid scatterers. $\boldsymbol{\gamma}$-Net was proposed in [23] by tailoring the complex-valued (CV) LISTA network. $\boldsymbol{\gamma}$-Net introduced weight coupling structure [24] and support selection scheme [24] to each iteration block in LISTA and improved the conventional soft-thresholding function by the piecewise linear function. It was demonstrated in [23] that $\boldsymbol{\gamma}$-Net improves the computational efficiency by two to three orders of magnitude compared to the state-of-the-art second-order TomoSAR solver SL1MMER [13] while showing no degradation in super-resolution ability and location accuracy.

However, unrolled neural networks do not consider historical information in the updating rules. To be exact, the output is generated exclusively based on the output of its previous layer. This kind of learning architecture leads to an error propagation phenomenon, where the error in the first few layers will be propagated and even amplified in the upcoming layers. Moreover, when the unrolled neural networks are designed for sparse reconstruction, shrinkage steps are usually required to promote sparsity. The shrinkage step utilizes thresholding functions to prune elements with a small magnitude to zero, and such pruning causes information loss in the dynamics of the neural network. Once useful information is discarded in the previous layers, the upcoming layers have no longer a chance to utilize the discarded information, thus degrading the performance of the neural network and sometimes leading to a large error in the final output.

B. Contribution of This Article

In this article, we aim to address the problem of information loss caused by shrinkage steps in unrolled neural networks designed for sparse reconstruction. To this end, we propose a novel architecture, termed the sparse minimal gated unit (SMGU), to incorporate historical information into optimization so that we can promote sparsity using thresholding functions and preserve full information simultaneously. In addition, we extend SMGU to the CV domain as CV-SMGU and use it to build a gated recurrent neural network (RNN) for solving TomoSAR inversion. The main contribution of this article is listed in the following.

1) We addressed the problem of information loss in unrolled neural networks for sparse reconstruction by a novel gated RNN. The gated RNN is built using SMGUs, which incorporate historical information into optimization. The proposed gated RNN is able to promote sparsity by employing shrinkage thresholding functions. Simultaneously, the pruned information will be reserved in the cell state of SMGUs; thus, full information can be preserved in the dynamics of the network.
2) We extend the SMGU to the CV domain, called CV-SMGU, and apply the gated RNN built with CV-SMGUs to solve TomoSAR inversion. To the best of our knowledge, it is the first attempt to bridge the gated RNN and TomoSAR inversion. We may provide novel insights and open a new prospect for future deep learning-based TomoSAR inversion.
3) We carry out a systematic evaluation to demonstrate that the proposed gated RNN outperforms the state-of-the-art deep learning-based TomoSAR algorithm $\boldsymbol{\gamma}$-Net in terms of super-resolution power and generalization ability for TomoSAR inversion.

The remainder of this article is outlined as follows. The TomoSAR imaging model and $\boldsymbol{\gamma}$-Net is briefly reviewed in Section II. Section III provides an overview of the formulation of SMGUs and CV-SMGUs with application to TomoSAR inversion. Results of systematic evaluation, using simulated and real data, are presented in Section IV. Section V discussed the generalization ability w.r.t. baseline discrepancy and analyzed the model convergence. Finally, the conclusion of this article is drawn in Section VI.

II. Background

A. TomoSAR Imaging Model

In this section, we briefly introduce the TomoSAR imaging model. Fig. 1 demonstrates the SAR imaging model at a fixed azimuth position. A stack of CV SAR acquisitions over the illuminated area is obtained at slightly different orbit positions (the elevation aperture). The CV measurement g_n of the nth acquisition is the integral of the reflectivity profiles $\gamma(s)$ along the elevation direction s. The discrete TomoSAR imaging model can be written as

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{2}$$

where $\mathbf{g} \in \mathbb{C}^{N\times 1}$ is the CV SAR measurement vector and $\boldsymbol{\gamma} \in \mathbb{C}^{L\times 1}$ denotes the discrete reflectivity profile uniformly sampled at elevation position $s_l (l = 1, 2, \ldots, L)$ along the elevation direction. N is the number of measurements, and L is the number of discrete elevation indices. $\mathbf{R} \in \mathbb{C}^{N\times L}$ is the irregularly sampled discrete Fourier transformation mapping matrix with $R_{nl} = \exp(-j2\pi\xi_n s_l)$, where ξ_n is the frequency

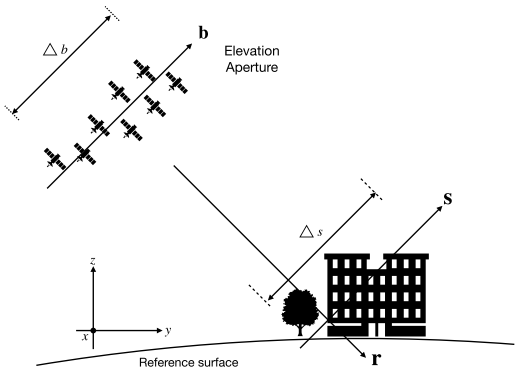


Fig. 1. SAR imaging geometry at a fixed azimuth position. The elevation synthetic aperture is built up by acquisition from slightly different incidence angles. The flight direction is orthogonal to the plane.

proportional to the perpendicular baseline of the nth acquisition. The readers can refer to [14] for more details of the SAR imaging model.

Since the reflectivity profile γ is sufficiently sparse in urban areas [5], retrieving γ is a sparse reconstruction problem. Accordingly, γ in the presence of measurement noise $\boldsymbol{\varepsilon}$ can be estimated by BPDN optimization, which is formulated as follows:

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \left\{ \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \right\} \tag{3}$$

where λ is a regularization parameter balancing the sparsity and data-fitting terms. It should be adjusted according to the noise level and the desired sparsity level. The choice of a proper λ is described in great detail in [1].

B. Review of $\boldsymbol{\gamma}$-Net

As shortly mentioned previously, conventional CS-based BPDN solvers for (3) are extremely computationally expensive. To overcome the heavy computational burden and make super-resolving TomoSAR inversion for large-scale processing feasible, the authors proposed $\boldsymbol{\gamma}$-Net in [23], which tailors the first unrolling ISTA network, to mimic a CS-based BPDN solver. To be specific, $\boldsymbol{\gamma}$-Net introduces the weight coupling structure and support selection scheme, and improves the conventional soft-thresholding function by the piecewise linear function. Fig. 2 illustrates us the architecture of the ith layer in $\boldsymbol{\gamma}$-Net. $\boldsymbol{SS}$ in $\boldsymbol{\gamma}$-Net indicates a special thresholding scheme called support selection, which will select ρ^i percentage of entries with the largest magnitude and trust them as "true support." The "true support" will be directly fed to the next layer, bypassing the shrinkage step. η_{pwl} is a novel thresholding function, called piecewise linear function, to execute shrinkage in the $\boldsymbol{\gamma}$-Net. It contributes to improving the convergence rate and reducing reconstruction error. More details about $\boldsymbol{\gamma}$-Net formulation and the full model structure can be found in the Appendix.

However, as one can see in Fig. 2, $\boldsymbol{\gamma}$-Net inherits the learning architecture of LISTA despite modifications made by the authors to improve the performance. Therefore, it can be imagined that $\boldsymbol{\gamma}$-Net will suffer from the same problem as LISTA. Specifically speaking, in the learning architecture of $\boldsymbol{\gamma}$-Net, the output at the current layer is generated exclusively from the previous output. As a natural consequence, the final output can only utilize the information from the second last layer. When useful or important information is pruned by shrinkage steps in the intermediate layers, the discarded information is no longer possible to contribute to the final output. Consequently, a large reconstruction error in the final output can be expected. Fig. 3 demonstrates an unsuccessful detection of double scatterers in our experiments. In this experiment, the double scatterers were assumed to have identical phase and amplitude and were spaced by 0.6 Rayleigh resolution, i.e., in a super-resolution regime, and the SNR level was set as 6 dB. In general, if we cannot resolve the overlaid double scatterers, the reflectivity profile should have a dominant amplitude peak between the true elevation position of the double scatterers, as it is shown by the estimate of a nonsuper-resolving algorithm SVD-Wiener [14] in Fig. 3. However, $\boldsymbol{\gamma}$-Net was able to detect one of the double scatterers with very high localization accuracy but failed to find the other one. From our perspective, it was abnormal, and we supposed that this unsuccessful double-scatterer separation should attribute to the information loss caused by shrinkage steps in $\boldsymbol{\gamma}$-Net. Inspecting the intermediate layers in $\boldsymbol{\gamma}$-Net, we discovered that the information of the second scatterer gradually diminished after each shrinkage step in the intermediate layers. Until the second last layer, the information of the second scatterer fell out completely. As a result, the final output of $\boldsymbol{\gamma}$-Net, i.e., the estimate of γ, did not contain the information of the second scatterer. Hence, we cannot detect the second scatterer.

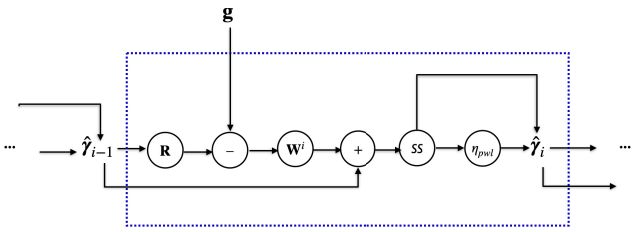


Fig. 2. Illustration of the ith layer in $\boldsymbol{\gamma}$-Net.

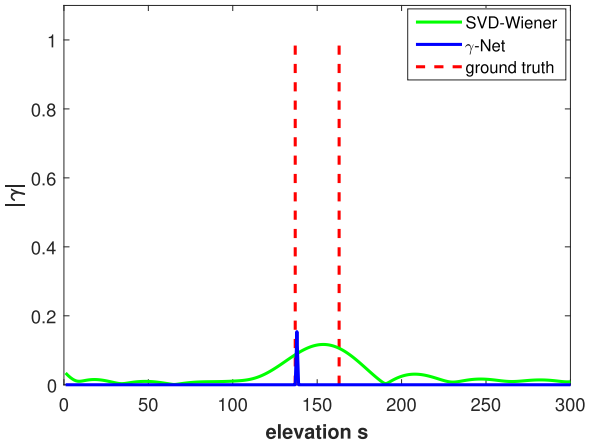


Fig. 3. Example of unsuccessful detection of double scatterers caused by information loss. $\boldsymbol{\gamma}$-Net detects one of the double scatterers with very high localization accuracy but fails to find the other one.

III. Methodology

A. Adaptive ISTA and sc2net

In the optimization community, it has been extensively studied and proved [25], [26], [27] that incorporating historical

TABLE I

FORMAL DEFINITION OF THE tTH LAYER IN DIFFERENT MODELS AND COMPARISON OF THEIR DIFFERENCE. γ-NET HAS NO GATED EXPRESSION. SLSTM UNIT INTRODUCES FORGET AND INPUT GATES TO INCORPORATE HISTORICAL INFORMATION. SMGU HAS THE MINIMAL NUMBER OF GATES WHILE MAINTAINING THE PERFORMANCE COMPARED TO THE SLSTM UNIT. CV-SMGU EXTENDS SMGU TO THE CV DOMAIN. THE FORGET GATE IS ACTIVATED ON THE MAGNITUDE USING ***tanh*** FUNCTION INSTEAD OF THE SIGMOID FUNCTION TO GUARANTEE THE ACTIVATION VALUE RANGING FROM 0 TO 1

	γ-Net layer	SLSTM unit	SMGU	CV-SMGU
Input gate	-	$\mathbf{i}^{(t)} = \sigma\{\mathbf{W}_{i2}^{(t)}\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_{i1}^{(t)}\mathbf{g}\}$	-	-
Forget gate	-	$\mathbf{f}^{(t)} = \sigma\{\mathbf{W}_{f2}^{(t)}\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_{f1}^{(t)}\mathbf{g}\}$	$\mathbf{f}^{(t)} = \sigma\{\mathbf{W}_{f2}^{(t)}\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_{f1}^{(t)}\mathbf{g}\}$	$\mathbf{f}^{(t)} = \tanh\{\lvert\tilde{\mathbf{W}}_{f2}^{(t)}\tilde{\boldsymbol{\gamma}}^{(t-1)} + \tilde{\mathbf{W}}_{f1}^{(t)}\tilde{\mathbf{g}}\rvert\}$
Cell state	-	$\bar{\mathbf{c}}^{(t)} = \mathbf{W}_2\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_1\mathbf{g}$ $\mathbf{c}^{(t)} = \mathbf{f}^{(t)} \odot \mathbf{c}^{(t-1)} + \mathbf{i}^{t} \odot \bar{\mathbf{c}}^{t}$	$\bar{\mathbf{c}}^{(t)} = \mathbf{W}_2(\mathbf{f}^{(t)} \odot \hat{\boldsymbol{\gamma}}^{(t-1)}) + \mathbf{W}_1\mathbf{g}$ $\mathbf{c}^{(t)} = (1 - \mathbf{f}^{(t)}) \odot \hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{f}^{(t)} \odot \bar{\mathbf{c}}^{(t)}$	$\bar{\mathbf{c}}^{(t)} = \tilde{\mathbf{W}}_2(\mathbf{f}^{(t)} \odot \tilde{\boldsymbol{\gamma}}^{(t-1)}) + \tilde{\mathbf{W}}_1\tilde{\mathbf{g}}$ $\mathbf{c}^{(t)} = (1 - \mathbf{f}^{(t)}) \odot \tilde{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{f}^{(t)} \odot \bar{\mathbf{c}}^{(t)}$
Output	$\tilde{\boldsymbol{\gamma}}^{(t)} = \eta_{ss}\rho_{\theta t}^{t}\{\tilde{\boldsymbol{\gamma}}^{(t-1)} + \tilde{\mathbf{W}}^{t}(\tilde{\mathbf{g}} - \tilde{\mathbf{R}}\tilde{\boldsymbol{\gamma}}^{(t-1)}), \boldsymbol{\theta}_t\}$	$\hat{\boldsymbol{\gamma}}^{(t)} = \eta_{dt}\left(\mathbf{c}^{(t)}\right)$	$\hat{\boldsymbol{\gamma}}^{(t)} = \eta_{dt}\left(\mathbf{c}^{(t)}\right)$	$\tilde{\boldsymbol{\gamma}}^{(t)} = \eta_{cv-dt}\{\mathbf{c}^{(t)}\}$

information contributes to improving the algorithm performance. Inspired by the high-level ideas from the previous research, researchers proposed adaptive ISTA in [28] to integrate and make use of historical information by introducing two adaptive momentum vectors $\mathbf{f}$ and $\mathbf{i}$ into ISTA in each iteration, which is formulated as follows:

$$\begin{aligned}\bar{\mathbf{c}}^{(t)} &= \mathbf{W}_2\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_1\mathbf{g}\\ \mathbf{c}^{(t)} &= \mathbf{f}^{(t)} \odot \mathbf{c}^{(t-1)} + \mathbf{i}^{(t)} \odot \bar{\mathbf{c}}^{(t)}\\ \hat{\boldsymbol{\gamma}}^{(t)} &= \eta_{\mathrm{st}}\left(\mathbf{c}^{(t)}\right)\end{aligned} \tag{4}$$

where η_{st} indicates the conventional soft-thresholding function and its CV version reads

$$\eta_{\mathrm{st}}\left(\hat{\boldsymbol{\gamma}}_i, \theta_i\right) = \begin{cases} \dfrac{\hat{\boldsymbol{\gamma}}_i}{|\hat{\boldsymbol{\gamma}}_i|}\max\left(|\hat{\boldsymbol{\gamma}}_i| - \theta_i, 0\right), & |\hat{\boldsymbol{\gamma}}_i| \neq 0 \\ 0, & \text{else.} \end{cases} \tag{5}$$

Compared to ISTA, whose update rule can be equivalently expressed as $\hat{\boldsymbol{\gamma}}^{(t)} = \eta_{\mathrm{st}}(\bar{\mathbf{c}}^{(t)})$ using the same notation, the adaptive ISTA takes not only the current information but also the previous information into consideration. To be exact, at the tth iteration of the adaptive ISTA, the estimate is generated by linear combining the historical information $\mathbf{c}^{(t-1)}$ at the previous iteration and the current information $\bar{\mathbf{c}}^{(t)}$ at the current iteration. The historical information $\mathbf{c}^{(t-1)}$ and the current information $\bar{\mathbf{c}}^{(t)}$ are weighted by the adaptive momentum vectors $\mathbf{f}^{(t)}$ and $\mathbf{i}^{(t)}$, respectively. By this means, the final estimate of the adaptive ISTA will accumulate historical information weighted by different $\mathbf{f}^{(t)}$ and $\mathbf{i}^{(t)}$ for different iterations.

However, one problem of the adaptive ISTA is that the two momentum vectors in each adaptive ISTA iteration are difficult to determine. So far, there has been no analytical way to determine the values of the adaptive momentum vectors $\mathbf{f}^{(t)}$ and $\mathbf{i}^{(t)}$. Usually, they are selected by tediously handcraft tuning, which takes a lot of time and cannot guarantee optimal performance. To address this issue, the authors proposed ***sc2net*** in [28] by recasting the adaptive ISTA as an RNN to parameterize the two momentum vectors and learn them from data. The ***sc2net*** is built by sparse long short-term memory (SLSTM) [28] units, as it is demonstrated in Fig. 4. Each SLSTM unit represents an individual layer of ***sc2net***. At the tth layer of ***sc2net***, the input gate and the forget gate correspond to the momentum vectors $\mathbf{i}^{(t)}$ and $\mathbf{f}^{(t)}$ in each adaptive ISTA iteration, respectively. Hence, we use the same

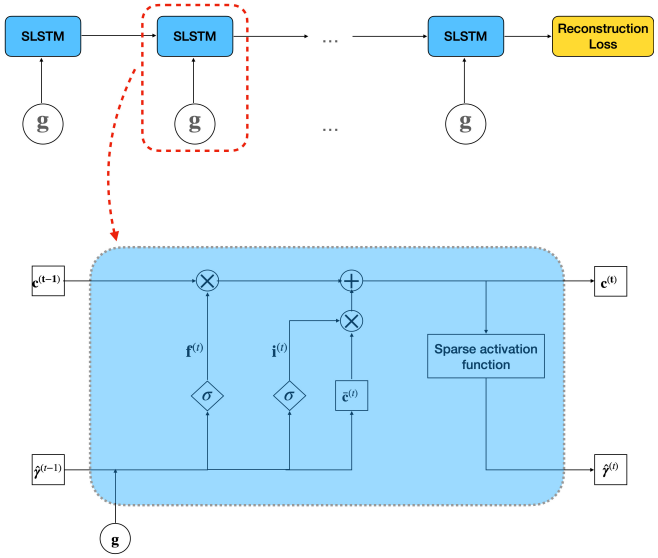


Fig. 4. Sc2net and detailed learning architecture of SLSTM unit. Each SLSTM unit builds an individual layer of sc2net.

notation in SLSTM units to describe the input and forget gates. The two gates in each SLSTM unit are parameterized with the input data $\mathbf{g}$ and the output $\hat{\boldsymbol{\gamma}}^{(t-1)}$ at the previous layer as follows:

$$\begin{aligned}\mathbf{i}^{(t)} &= \sigma\left(\mathbf{W}_{i2}^{(t)}\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_{i1}^{(t)}\mathbf{g}\right)\\ \mathbf{f}^{(t)} &= \sigma\left(\mathbf{W}_{f2}^{(t)}\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_{f1}^{(t)}\mathbf{g}\right).\end{aligned} \tag{6}$$

To clarify, the SLSTM unit does not have an output gate like conventional LSTM units. By substituting (6) into (4), we have the formal definition of the SLSTM unit, as it is listed in Table I. $\mathbf{W}_{i1}, \mathbf{W}_{i2}, \mathbf{W}_{f1}$, and $\mathbf{W}_{f2}$ denote four trainable weight matrices to determine the input and forget gates in each SLSTM unit. It is worth mentioning that the weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are also learned from data, while they are shared for all SLSTM units in an individual ***sc2net***. $\sigma(\cdot)$ indicates the conventional sigmoid function, which is expressed as

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{7}$$

The sparse activation function employed in the SLSTM to promote sparse codes is the double hyperbolic tangent function, which is abbreviated as $\eta_{dt}(\cdot)$ and defined as follows:

$$\eta_{dt}(\hat{\boldsymbol{\gamma}}, s, \theta) = s \cdot \left[\tanh(\hat{\boldsymbol{\gamma}} + \theta) + \tanh(\hat{\boldsymbol{\gamma}} - \theta)\right] \tag{8}$$

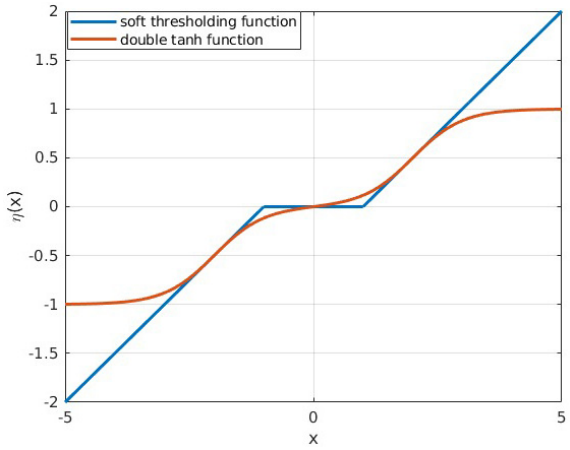


Fig. 5. Comparison of the double hyperbolic tangent function $\eta_{dt}(\cdot)$ and the soft-thresholding function. $\eta_{dt}(\cdot)$ effectively imitates the soft-thresholding function within the interval of $[-\theta, \theta]$.

where s and θ denote two trainable parameters. It is worth noting that the double hyperbolic tangent function can be viewed as a smooth and continuously differentiable alternative to the conventional soft-thresholding function. Its advantages are mainly twofold. On the one hand, its second derivative sustains for a long span, thus contributing to addressing the gradient vanishing problem caused by the cell recurrent connection [29]. On the other hand, it is able to effectively imitate the soft-thresholding function within the interval of $[-\theta, \theta]$. Fig. 5 demonstrates an example of the double hyperbolic tangent function and compares it to the soft-thresholding function.

To sum up, ***sc2net*** inherits the advantage of the adaptive ISTA, which incorporates historical information into optimization. The cell state $\mathbf{c}^{(t)}$ in each SLSTM unit of ***sc2net*** acts as an "eye" to supervise the optimization from two aspects. First, the long-term dependence on the previous outputs can be captured and maintained. Second, important information will be automatically accumulated, whereas useless or redundant information will be forgotten, in the dynamics of ***sc2net***.

However, when we tried to apply ***sc2net*** in TomoSAR inversion, we discovered that a drawback of ***sc2net*** impedes its application. As it is known, a complicated RNN model, on the one hand, hinders theoretical analysis and empirical understanding. On the other hand, it also implies that we have to learn more parameters and tune more components. As a natural result, more training sequences, which mean more training time, and (perhaps) larger training datasets are required. When ***sc2net*** is applied to solve TomoSAR inversion, we need to learn four weight matrices $\mathbf{W}_{f1}^{(t)}$, $\mathbf{W}_{f2}^{(t)}$, $\mathbf{W}_{i1}^{(t)}$, and $\mathbf{W}_{i2}^{(t)}$, which have the dimension $L \times L$, $L \times N$, $L \times L$, and $L \times N$, respectively, to determine the forget gate $\mathbf{f}^{(t)}$ and the input gate $\mathbf{i}^{(t)}$ in each individual SLSTM unit. Moreover, SAR data are CV. Hence, there weight matrices to be learned should be CV as well, thus duplicating the number of trainable components and parameters since two weight matrices need to be learned simultaneously as the real and imaginary parts of a CV weight matrix. Through our research and experiments, we found that such a large amount of high-dimensional weight matrices to be learned makes the training procedure time-consuming. More seriously, it is difficult for the model to converge in the training process.

B. Complex-Valued Sparse Minimal Gated Unit

To address the aforementioned issue and better leverage the power of incorporating historical information for solving TomoSAR inversion, it is necessary to reduce the components and simplify the model architecture. Recently, studies and evaluations in [30], [31], and [32] demonstrated that the gated unit contributes to significantly improving the performance of an RNN compared to that without any gated unit. However, it does not signify that the more the gates, the better the performance of an RNN. Based on this fact, the author proposed an RNN model with only one gate, termed the minimal gated unit (MGU), and revealed that fewer gated units reduce the complexity but not necessarily the performance.

Inspired by the valuable works in [33] and [34], we proposed SMGU, as illustrated in Fig. 6, by coupling the input gate to the forget gate, thus further the simplifying SLSTM unit. The detailed equations for defining the SMGU are listed in Table I.

In the tth layer of an RNN with SMGUs, we will first compute the forget gate $\mathbf{f}^{(t)}$. In addition, the short-term response $\bar{\mathbf{c}}^{(t)}$ is generated by combining the input data $\mathbf{g}$ and the "forgotten" portion $(\mathbf{f}^{(t)} \odot \hat{\boldsymbol{y}}^{(t-1)})$ of the output from the previous layer. Hereafter, the new hidden state $\mathbf{c}^t$ of the current layer can be formulated by combining part of $\hat{\boldsymbol{y}}^{(t-1)}$ and the short-term response $\bar{\mathbf{c}}^{(t)}$, which are determined by $(1-\mathbf{f}^{(t)})$ and $\mathbf{f}^{(t)}$, respectively. Eventually, the sparse activation function, i.e., the double hyperbolic function, will be applied to the current hidden state $\mathbf{c}^t$ for shrinkage and thresholding to promote sparsity of the output.

In this formulation, we can see that the SMGU is able to simultaneously execute a twofold task with only one forget gate. On the one hand, SMGU allows a compact representation by enabling the hidden state $\mathbf{c}^{(t)}$ to discard irrelevant or redundant information. On the other hand, SMGU is capable of controlling how much information from the previous layer takes over. In addition, comparing the formulation of SMGU to SLSTM in Table I, we can see that the parameter size of SMGU is only about half of that of SLSTM since the weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are shared for different layers in a network. The main advantage brought by the significant elimination of trainable parameters is that we can reduce the requirement for training data, training time, and architecture tuning.

In addition to the improvements using SMGU, an extension of SMGU to the complex domain is required. CV-SMGU has essentially the same structure as SMGU despite the two differences. First, each neuron in CV-SMGU has two channels indicating the real and imaginary parts of a complex number, respectively. Often, the real and imaginary parts are not directly activated. Instead, the activation is performed on the magnitude of the complex number. Hence, it is no longer appropriate to use the sigmoid function for activation to generate the forget gate since the magnitude is always greater than

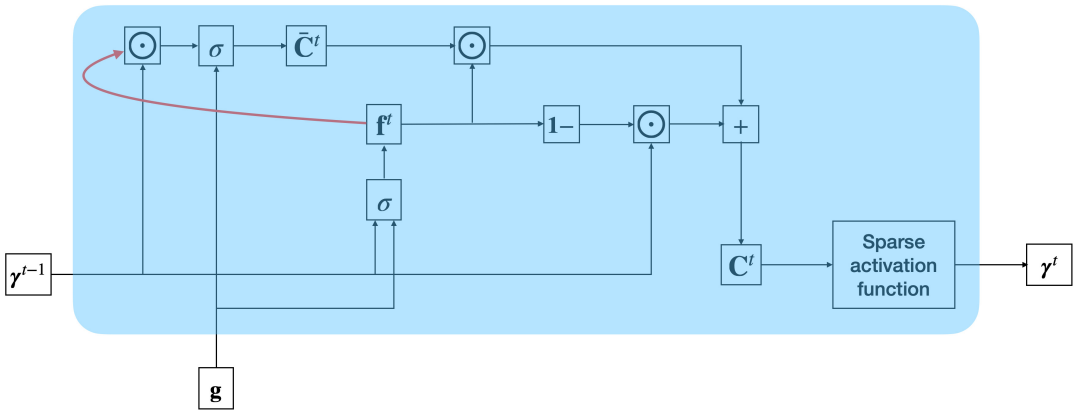


Fig. 6. Structure of the proposed SMGU. **f** indicated the only gate in each SMGU.

zero leading to the undesired result being always greater than 0.5 after activation. To tackle this problem, we employed the "tanh" function instead of sigmoid to guarantee that the value of the forget gate vector varies from 0 to 1 after activation, as it is originally designed. By applying the aforementioned adaptions, we have the formulations of CV-SMGU, as listed in Table I as well. The symbols $\tilde{\mathbf{W}}_*$, $\tilde{\mathbf{g}}$, and $\tilde{\boldsymbol{\gamma}}^*$ represent

$$\tilde{\mathbf{W}}_* = \begin{bmatrix} \mathrm{Re}(\mathbf{W}_*) & -\mathrm{Im}(\mathbf{W}_*) \\ \mathrm{Im}(\mathbf{W}_*) & \mathrm{Re}(\mathbf{W}_*) \end{bmatrix}$$
$$\tilde{\mathbf{g}} = \begin{bmatrix} \mathrm{Re}(\mathbf{g}) \\ \mathrm{Im}(\mathbf{g}) \end{bmatrix}$$
$$\tilde{\boldsymbol{\gamma}}^* = \begin{bmatrix} \mathrm{Re}(\hat{\boldsymbol{\gamma}}^*) \\ \mathrm{Im}(\hat{\boldsymbol{\gamma}}^*) \end{bmatrix}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary operators, respectively. $\eta_{cv-dt}(\cdot)$ is the CV version of the double hyperbolic function applied componentwise and expressed as follows:

$$\eta_{cv-dt}(\hat{\boldsymbol{\gamma}}, s, \theta) = \begin{cases} \dfrac{\hat{\boldsymbol{\gamma}}_i}{|\hat{\boldsymbol{\gamma}}_i|} s \cdot e^{j\cdot\angle(\hat{\boldsymbol{\gamma}})} \big[\tanh(|\hat{\boldsymbol{\gamma}}|+\theta) + \tanh(|\hat{\boldsymbol{\gamma}}|-\theta)\big], & |\hat{\boldsymbol{\gamma}}_i| \neq 0 \\ 0, & \text{else.} \end{cases} \tag{9}$$

Table II summarizes and compares the features of different unrolled RNNs. Through experiments, we found that gated unrolled RNNs require significantly fewer layers to achieve comparable or even better performance. Moreover, the SMGU simplifies the model structure by coupling the two gates, thus significantly eliminating the number of free trainable parameters. Even if the CV-SMGU duplicates the number of parameters for determining the gate, it will not induce a serious memory burden or computational expense.

IV. Performance Evaluation

A. Simulation Setup and Model Training

In the simulation, we applied the same settings as [23], i.e., 25 regularly distributed spatial baselines in the range of −135 to 135 m were simulated. The corresponding inherent elevation resolution, i.e., Rayleigh resolution, amounts to about 42 m.

In the experiment, about four million training samples, half of which are single scatterer and the others are two-scatterer mixtures, were simulated to generate the training dataset. To make the training dataset adequate and the simulation more realistic, we randomized many parameters, i.e., SNR level, amplitude, phase, and elevation position of scatterers, when we simulated the training samples. The simulation details of single scatterer and double scatterers are listed as follows.

1) *Single Scatterer:* For single scatterer, the scattering phase ϕ is set to follow a uniform distribution, i.e., $\phi \sim U(-\pi, \pi)$. In addition, the amplitude A of the scatterer is simulated to be uniformly distributed in the range of $(1, 4)$. Hereafter, the CV scattering coefficient γ can be generated by $\gamma = A \cdot \exp(j\phi)$. The elevations of the simulated scatterers are regularly distributed on a 1-m grid between −20 and 300 m. Once the elevation is determined, the echo signal $\mathbf{g} \in \mathbb{C}^{25}$ is generated with different levels of SNR, which is regularly distributed between [0 dB, 10 dB] with 11 samples.
2) *Double Scatterers:* We simulated two single scatterers inside each resolution unit. The simulation of the two single scatterers is identical to the previous step. As a consequence, different amplitude ratios, different scattering phase offsets, and different elevation distances between the two scatterers are considered.

The model was implemented and trained under the framework of Pytorch [35]. The employed optimizer was Adam [36]. The learning rate was set to be adaptive according to the number of training epochs with the initial value being 0.0001. The loss function over the training data $\{(\mathbf{g}_i, \boldsymbol{\gamma}_i)\}_{i=1}^T$ is mean square error (mse) loss, which is defined as follows:

$$\underset{\boldsymbol{\Psi}}{\text{minimize}}\ \mathcal{L}(\boldsymbol{\Psi}) = \frac{1}{T}\sum_{i=1}^{T} ||\hat{\boldsymbol{\gamma}}(\boldsymbol{\Psi}, \mathbf{g}_i) - \boldsymbol{\gamma}_i||_2^2 \tag{10}$$

where $\boldsymbol{\Psi}$ denotes the set of all parameters to be learned from data. To determine the optimal structure of the network, we validated the performance of the network with different numbers of CV-SMGUs in terms of normalized mse (NMSE) on a validation dataset. The validation dataset was composed of 50 000 noise-free samples simulated using the same settings

TABLE II
COMPARISON OF DIFFERENT UNROLLED RNNS FOR SPARSE RECONSTRUCTION

Features	**γ-net**	**sc2net**	**SMGU**	**CV-SMGU**
complex-value	Yes	No	No	Yes
gates expression	No	Yes	Yes	Yes
number of gates	0	2	1	1
number of parameters for gates	0	$2 \cdot (L^2 + NL)$	$L^2 + NL$	$2 \cdot (L^2 + NL)$
required number of layers	≈ 15	≈ 5	≈ 5	≈ 5

introduced in Section III, and the NMSE is defined as follows:

$$\text{NMSE} = \frac{1}{T}\sum_{i=1}^{T}\frac{\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|_2^2}{\|\boldsymbol{\gamma}_i\|_2^2}. \tag{11}$$

As we can see from Table III, the NMSE gradually converges with increasing the number of SMGUs. Moreover, after six CV-SMGUs, a further increase in the number of CV-SMGUs leads to marginal performance improvement. Instead, a heavier computational burden will be brought about. Therefore, the network we designed is composed of six CV-SMGUs.

B. Performance Assessment and Comparison to γ-Net

In this section, we carry out experiments to systematically evaluate the performance of the proposed algorithm in terms of super-resolution power, estimation accuracy, and generalization ability against different amplitude ratios and phase differences of scatterers.

C. Super-Resolution Power and Estimation Accuracy

The first experiment sets out to study the super-resolution power and estimation accuracy of the proposed algorithm via a TomoSAR benchmark test [5], [14]. In the experiment, we mimicked a facade-ground interaction by simulating two-scatterer mixtures with increasing elevation distance between them. The double scatterers were simulated to have identical phase and amplitude, i.e., the worst case for TomoSAR processing [13]. The proposed algorithm and $\boldsymbol{\gamma}$-Net were employed to resolve overlaid double scatterers at two SNR levels, i.e., SNR$\in \{0, 6\}$ dB, which represents typical SNR levels of a high-resolution spaceborne SAR image. We use the effective detection rate defined in [23] to fairly evaluate the super-resolution power. An effective detection should satisfy the following three criteria.

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 ± 0.5 d_s w.r.t. their true elevation.

Here, d_s indicates the distance between the double scatterers. Fig. 7 compares the effective detection rate P_d of the proposed algorithm and γ-Net. It is presented as a function of the normalized distance α, which is the ratio of the scatterers' distance and the Rayleigh resolution $\alpha = d_s/\rho_s$. For each combination of SNR and α, we simulated 0.2 million Monte Carlo trials. From Fig. 7, one can see that the proposed algorithm and

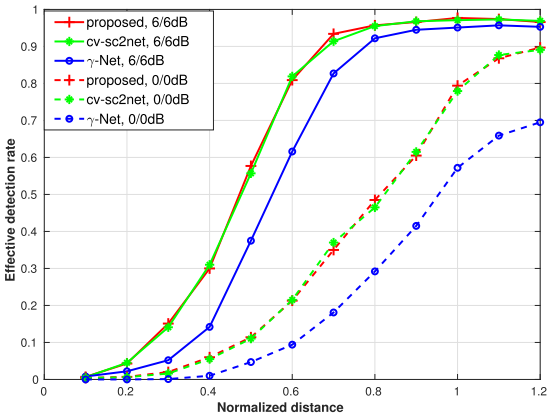


Fig. 7. Effective detection rate of the proposed algorithm, CV-sc2net and $\boldsymbol{\gamma}$-Net as a function of the normalized elevation distance between the simulated facade and ground with SNR $= 0$ and 6 dB under 0.2 million Monte Carlo trials.

sc2net with CV-SLSTMs (CV-sc2net) have quite similar performance in terms of effective detection rate. This is the same as we expected since the CV-SMGU is constructed by simplifying the CV-SLSTM. The purpose of CV-SMGU is to reduce network components while maintaining performance. The advantages of the proposed algorithm compared to CV-sc2net are analyzed and discussed in Section V. When we compare the proposed algorithm and CV-sc2net to γ-Net, we can see that both the proposed algorithm and CV-sc2net outperform γ-Net by a fair margin at both SNR levels. Specifically, they are able to deliver 10%–20% higher effective detection rate in moderate super-resolving cases at 6-dB SNR. In the noisy case at 0-dB SNR, the proposed algorithm and CV-sc2net gradually approach about 90% effective detection rate with the increase in the normalized distance, whereas γ-Net reaches only about 70% effective detection rate. The superior performance of the proposed algorithm and CV-sc2net attributes to that they overcome the information loss in the dynamics of the network by incorporating historic data and preserving full information. As we have mentioned in Section II, the detection of double scatterers is affected by information loss. We cannot detect the scatterers whose information is discarded.

To better manifest how the incorporation of historic information improves the performance, we simulated 2000 samples containing double scatterers with increasing scatterers distance at 6-dB SNR. We made a scatter plot of their elevation estimates and color-coded the points by the detector decision in Fig. 8. The x-axis refers to the true normalized elevation distance of the scatterers. The y-axis shows their normalized elevation estimates. The ideal reconstruction would be a horizontal and a diagonal straight line, which represents the ground

TABLE III

PERFORMANCE OF THE NETWORK WITH DIFFERENT NUMBER OF SMGUS. AFTER 6 SMGUS, THE PERFORMANCE IMPROVEMENT IS MARGINAL WITH INCREASING THE NUMBER OF SMGUS. INSTEAD, THE INCREASE IN SMGUS LEADS TO HEAVIER COMPUTATIONAL BURDEN

Number of CV-SMGUs	2	3	4	5	6	7	8	9
NMSE [dB]	-12.4	-20.7	-26.1	-29.6	-30.2	-30.6	-30.8	-30.9

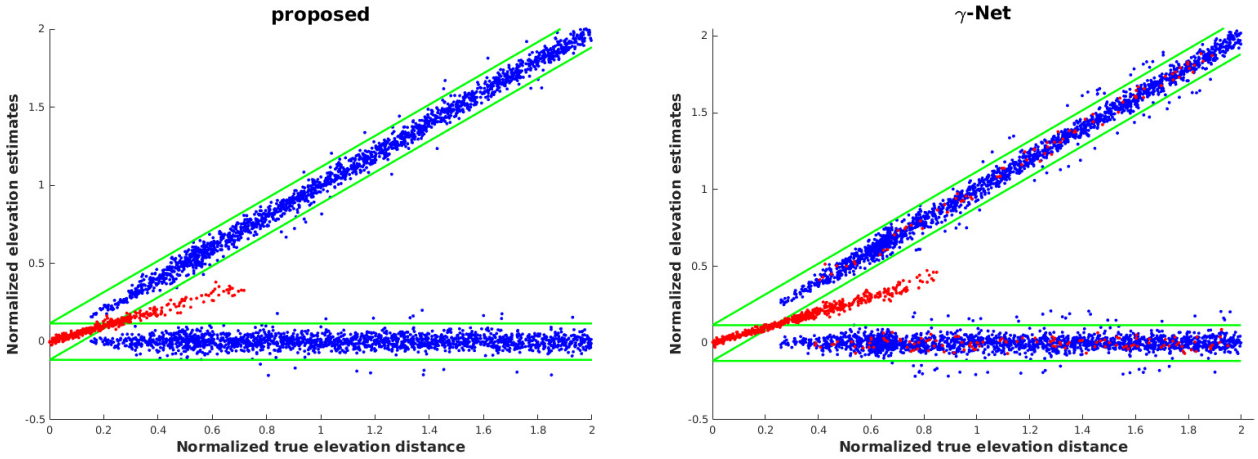


Fig. 8. Normalized estimated elevation of facade and ground of increasing elevation distance with SNR = 6 dB and N = 25. The double scatterers were simulated to have identical phase and amplitude. The true positions are a horizontal line referring to the ground and a diagonal line referring to the scatterers at variable elevations. The green lines depict true positions ± 3 times CRLB of elevation estimates for single scatterers. Red dots represent samples detected as single scatterers. Blue dots indicate detected overlaid double scatterers.

truth of the simulated ground and facade. The green lines refer to ground truth ±3 times CRLB of single scatterer elevation estimate. The blue dots indicate the detected double scatterers, whereas the red dots represent that the samples were detected as single scatterers, meaning that the second scatterer was lost in the network output. Fig. 8 clear shows the following.

1) γ-Net experiences many more red dots locate within ±3 times CRLB w.r.t. the ground truth, meaning that it occasionally can only detect one of the double scatterers but is able to estimate its elevation with high precision. We ascribe this problem to the information loss caused by the learning structure of γ-Net. On the contrary, the proposed algorithm utilizes CV-SMGUs to preserve full information, thus avoiding discarding any significant information.
2) The proposed algorithm is able to resolve double scatterers at much smaller scatterers' distances. Specifically, the proposed algorithm starts to separate double scatterers from about 0.15 Rayleigh resolution, whereas γ-Net can only detect double scatterers only after about 0.3 Rayleigh resolution.

The elevation estimates of the simulated facade and ground are plotted in Fig. 9 w.r.t. the normalized true elevation distance. The red horizontal and slant lines indicate the ground truth of the ground and façade, respectively. The black dashed curves represent the ground truth ±1× CRLB. The error bars indicate the standard deviation of the elevation estimates with the midpoint depicting the mean value of the elevation estimates at the given normalized true elevation distance. We discarded the points below an effective detection rate of 5% in the figures. Due to the strict criteria of the effective detection, both the proposed algorithm and γ-Net provide high elevation estimation accuracy, especially at 6-dB SNR, where the bias of the elevation estimates derived by both methods approaches 0. However, in the extremely noisy case, we can see that the proposed algorithm is able to estimate the elevation with a slightly lower bias compared to γ-Net.

D. Performance w.r.t. Amplitude Ratio

In this experiment, we propose to study how the proposed algorithm performs at different amplitude ratios of double scatterers. The double scatterers were set to have identical phases. The SNR level was set as 6 dB. Fig. 10 compares the effective detection rate of the proposed algorithm and γ-Net at different amplitude ratios. As can be seen, the effective detection rate of both algorithms degrades with the increase in the amplitude ratio. The reason for the degradation of the effective detection rate is twofold. First, dark scatterers suffer from larger and larger bias with the increase in the amplitude ratio since their elevation estimates tend to approach the other more prominent scatterer. Second, at a high amplitude ratio, the energy of the second scatterer is closer to the noise level. In real-world applications, we usually see dark scatterers at a high amplitude ratio (≥4) as noise. However, by comparing the two algorithms, we can see from Fig. 10 that the proposed algorithm performs much better with the increase in the amplitude ratio than $\boldsymbol{\gamma}$-Net despite the fact that the effective detection rate is seriously affected. From our perspective, the better performance of the proposed algorithm attributes to that the estimates derived by the proposed algorithm preserve the full information; thus, we have a higher chance to retrieve weak signals of dark scatterers.

E. Performance w.r.t. Phase Difference

As it was investigated in [13], the super-resolution power depends strongly on the phase difference when double

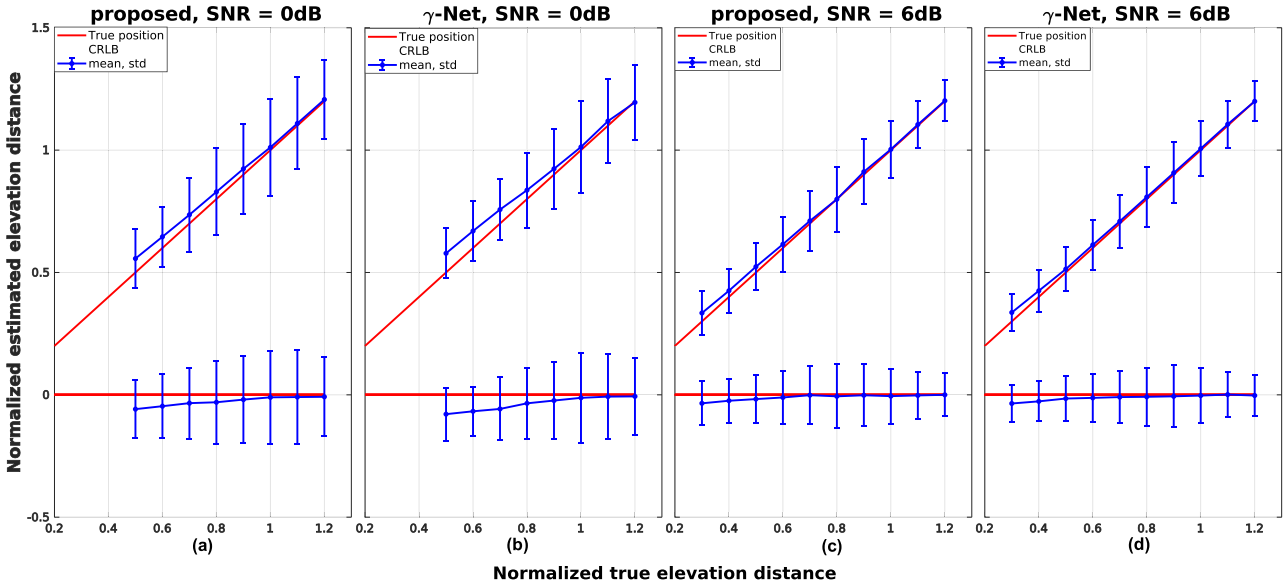


Fig. 9. Estimated elevation of simulated facade and ground: (a) SNR = 0 dB with the proposed algorithm, (b) SNR = 0 dB with γ-Net, (c) SNR = 6 dB with the proposed algorithm, and (d) SNR = 6 dB with γ-Net. Each dot has the sample mean of all estimates as its y value and the correspond standard deviation as the error bar. The red line segments represent the true elevation of the simulated facade and ground. The dashed curves denote the true elevation $\pm 1\times$ CRLB normalized w.r.t. the Rayleigh resolution.

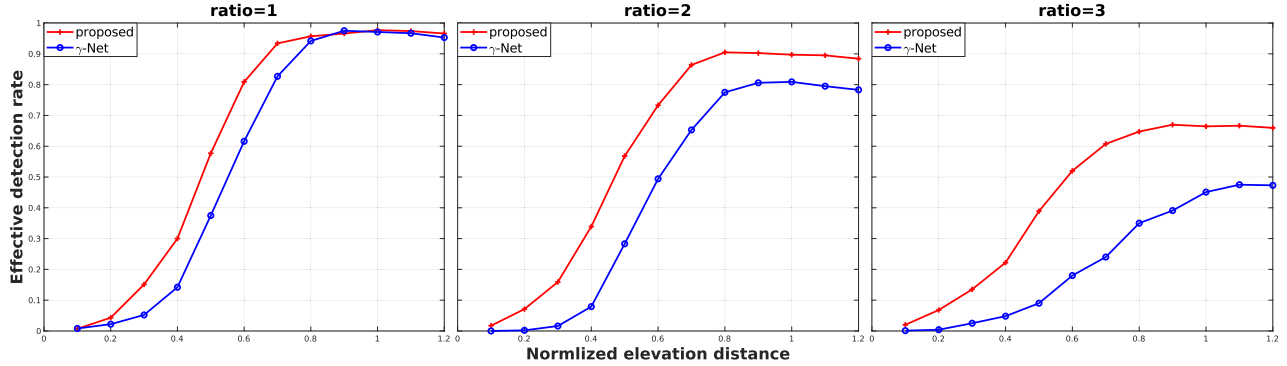


Fig. 10. Effective detection rate of the two algorithms w.r.t. the normalized elevation distance at different amplitude ratios.

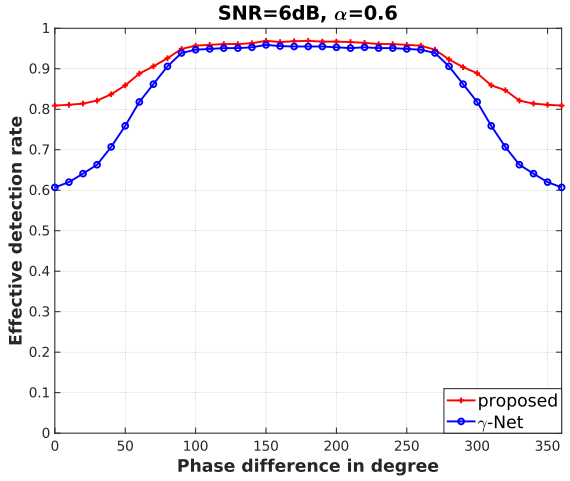


Fig. 11. Effective detection rate ρ_d of the two algorithms as a function of phase difference $\Delta\phi$ under the case: $N = 25$, SNR = 6 dB, and $\alpha = 0.6$.

scatterers were spaced within the Rayleigh resolution. To evaluate how the proposed algorithm performs w.r.t. phase difference of double scatterers in super-resolving cases, we vary the phase difference of simulated double scatterers in this experiment and test the effective detection rate. The double scatterers are simulated with identical amplitude. Fig. 11 illustrates the effective detection rate of the proposed algorithm and γ-Net for the case when $N = 25$ and SNR = 6 dB with $\alpha = 0.6$. As can be seen, both algorithms have the worst performance at $\Delta\phi = 0$ and perform better when $\Delta\phi$ approaches 180°. Compared to γ-Net, the proposed algorithm is less sensitive to the phase difference. When $\Delta\phi = 0$, the proposed algorithm delivers about 20% higher effective detection rate than γ-Net.

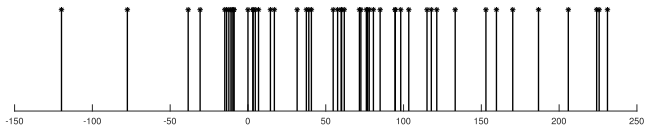
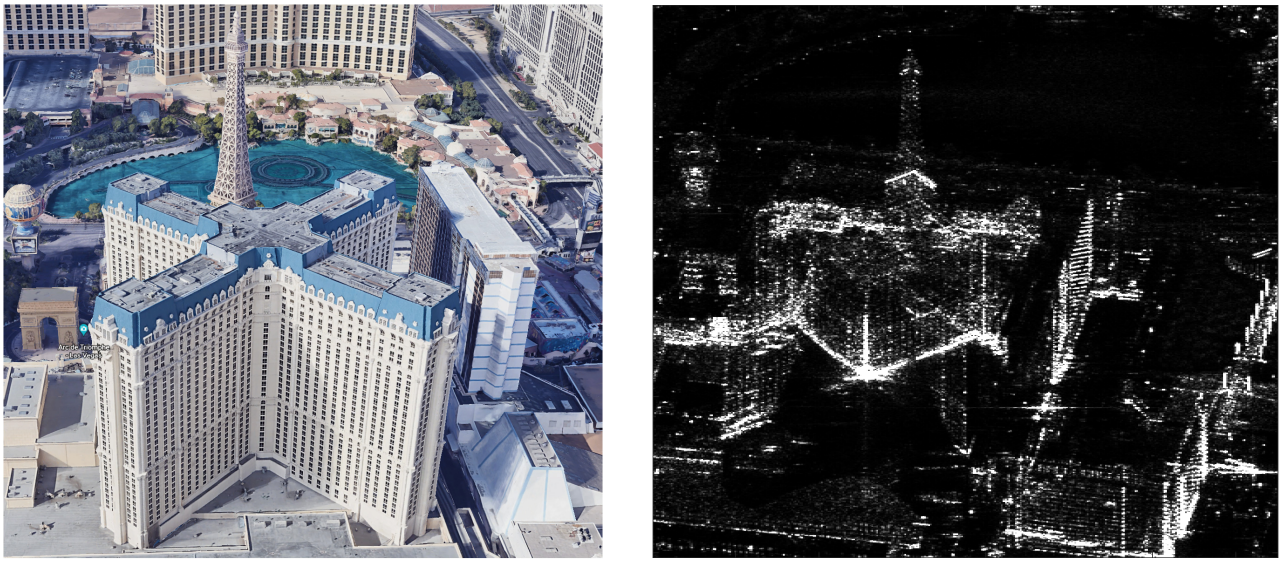


Fig. 12. Effective baselines of the 50 acquisitions.

F. Practical Demonstration

For the real data experiment, we used the test data stack over the city of Las Vegas covering Paris Hotel. The stack is composed of 50 TerraSAR-X high-resolution spotlight



(a) (b)

Fig. 13. Test site. (a) Optical image from Google Earth. (b) SAR mean intensity image.

TABLE IV
SYSTEM PARAMETERS OF THE TERRASAR-X HIGH-RESOLUTION SPOTLIGHT IMAGE STACK

parameter	value
slant-range resolution	0.6m
azimuth resolution	1.1m
acquisition time	2008-2010
range distance	704km
incidence angle	31.8°

images with a slant-range resolution of 0.6 m and an azimuth resolution of 1.1 m, whose spatial baseline distribution is demonstrated in Fig. 12. In Fig. 13, an optical image from Google Earth and the SAR mean intensity image of the test site are demonstrated. The images were acquired between 2008 and 2010. More details of the data stack that we use are listed in Table IV.

We employed the DLR's integrated wide area processor (IWAP) [37] to carry out preprocessing like multiple SAR images' coregistration and phase calibration. In addition, a coherence point on the ground was chosen as a reference.

We used the baselines of the test data stack to simulate training data. The simulation was conducted in the same way as introduced in the simulation setup in Section IV-A, and four million training samples were generated. When the network was well-trained, the proposed algorithm was directly applied to reconstruct the elevation of the test site.

The reconstruction results of the test site are demonstrated in Fig. 14 and compared to the results derived by γ-Net. Fig. 14(a) and (b) illustrates color-coded elevation of single scatterers detected by both algorithms. Fig. 14(c)–(f) depicts the reconstruction of detected double scatterers of both algorithms. The double scatterers are separated into the top and bottom layers according to their elevation estimates, and the top and bottom layers are demonstrated separately. By comparing the reconstruction results of both algorithms, we can see that the proposed algorithm detects the double scatterers

TABLE V
PERCENTAGE OF SCATTERERS' DETECTION FOR THE TWO ALGORITHMS

Algorithm	Percentage of detection as		
	0 scatterer	1 scatterer	2 scatterers
proposed	62.01 %	33.30 %	4.69 %
γ-Net	61.06 %	35.83 %	3.11 %

with a higher density, indicating that the proposed algorithm has stronger super-resolution power. A closer inspection of the reconstruction of double scatterers shows that a serious layover exists on the top of the cross-building. Moreover, the elevation estimates of detected double scatterers indicate that the top layer is mainly caused by reflections from the building roof and building facade, whereas the bottom layer is composed of scatterers on the ground or lower infrastructures.

To provide a more intuitive comparison of the super-resolution power of both algorithms, we summarized the scatterers' detection of both algorithms in Table V. As it is shown in Table V, most pixels are detected as zero scatterers by the two algorithms because the fountain and many low infrastructures in the test site exhibit no strong scattering, which can be seen in Fig. 13(b). Compared to γ-Net, the proposed algorithm detected fewer single scatterers (33.30%) but more double scatterers. Comparison between the double scatterers detected by both algorithms shows that the proposed algorithm is able to detect 95.2% of the double scatterers detected by γ-Net. Moreover, it detects 50% more double scatterers than γ-Net.

Further investigation was conducted to inspect the improvement of double-scatterer detection. The histogram of detected double scatterers' elevation difference from the proposed algorithm and γ-Net is shown in Fig. 15. In the nonsuper-resolution region, especially when the distance between double scatterers is larger than twice Raleigh resolution, the two algorithms have a comparable performance of double scatterers' detection. However, in the super-resolution region,

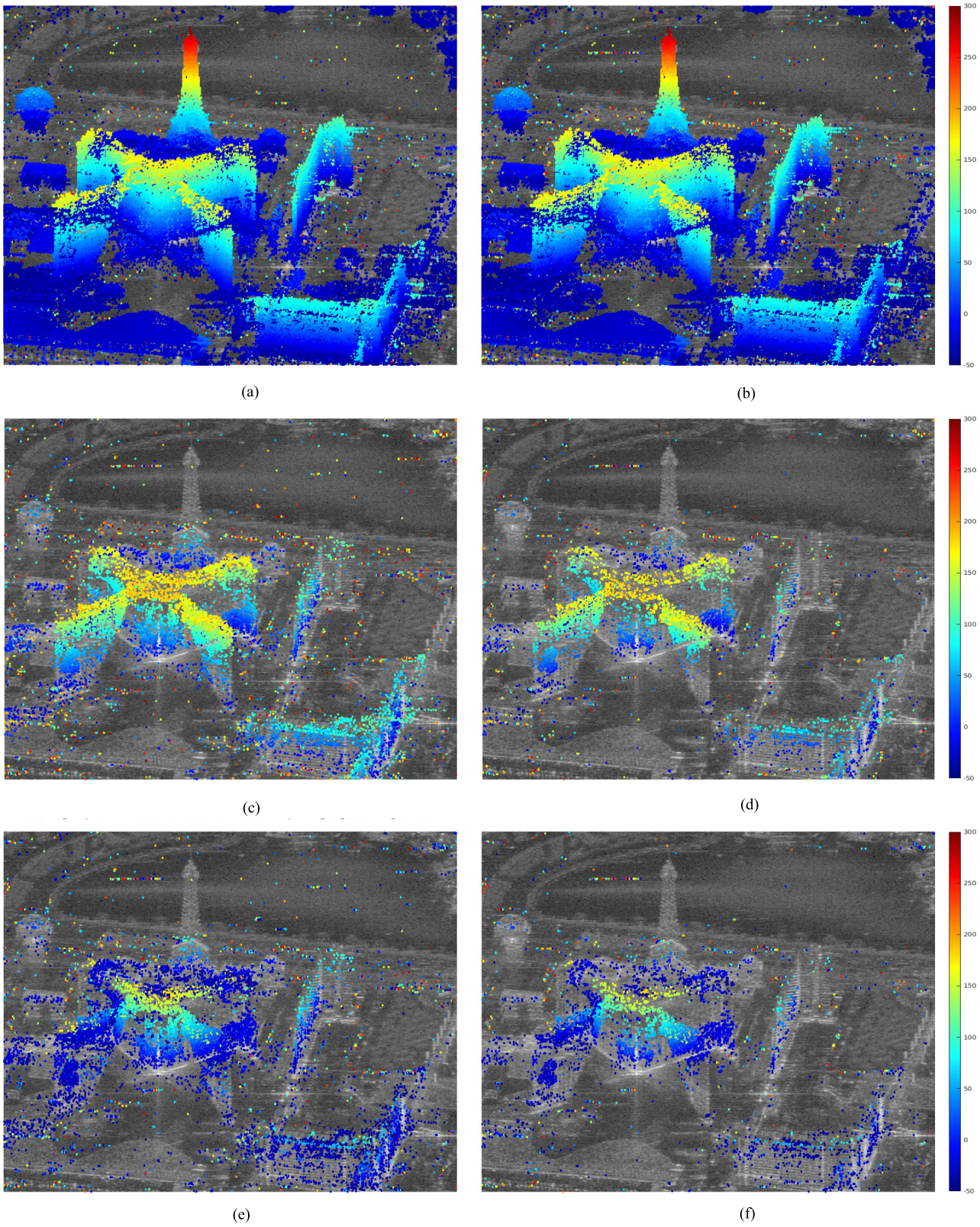


Fig. 14. Reconstructed and color-coded elevation of detected scatterers. From (Left) to (Right): elevation estimates derived by the proposed algorithm and γ-Net, respectively. From (Top) to (Bottom): Color-coded elevation of detected single scatterers, the top layer of detected double scatterers, and the bottom layer of detected double scatterers, respectively. (a) Single scatterer detected by the proposed algorithm. (b) Single scatterer detected by γ-Net. (c) Top layer of double scatterers detected by the proposed algorithm. (d) Top layer of double scatterers detected by γ-Net. (e) Bottom layer of double scatterers detected by the proposed algorithm. (f) Bottom layer of double scatterers detected by γ-Net.

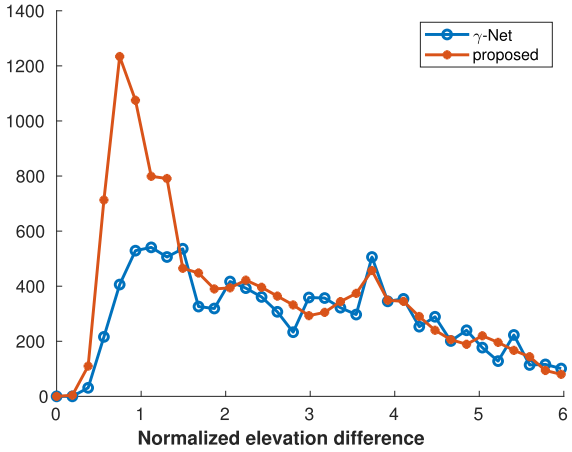


Fig. 15. Histogram of the elevation distance between the detected double scatterers from the proposed algorithm and γ-Net. The proposed algorithm shows significantly more detection in the super-resolution region.

the proposed algorithm delivers obviously stronger resolution ability.

V. Discussion

A. Generalization Ability Against Baselines' Discrepancy

The effective baseline in an SAR image varies according to the range and azimuth location. A deep learning model trained with a fixed set of baselines may have undesired performance when being applied to the whole image stack, as baseline discrepancies between training and testing data may cause data domain shifts. In this experiment, we verify the generalization ability against baseline discrepancies. The network with six CV-SMGUs is trained using 25 regularly distributed baselines as introduced in the simulation setup. Then, we add random perturbation uniformly distributed in the range [5 m, 10 m], i.e., about [7%, 14%] of the standard deviation of the 25 regularly distributed baselines, to the 25 regularly distributed baselines. 100 different baselines' distributions were generated. For each baseline distribution, we carry out a Monte Carlo simulation at 6-dB SNR for each baselines' distribution with 0.2 million Monte Carlo trials at each discrete normalized distance. Fig. 16 demonstrates the effective detection rate of the proposed algorithm when we apply the pretrained network to the data generated with baseline perturbations. The red line represents the reference, i.e., the pretrained network is applied to data simulated with the same baselines' distribution. The green line indicates the average effective detection rate of the 100 Monte Carlo simulations with the blue error bars depicting the standard deviation. As one can see, the proposed algorithm shows a good generalization ability against baselines' discrepancy with the effective detection rate decreasing only 5%–8% compared to the reference. Therefore, we see the proposed algorithm as a promising tool for large-scale TomoSAR processing since the biggest baselines' difference of a typical spaceborne SAR image will not exceed the perturbation that we simulated.

However, for baselines with large perturbations or even completely different distributions, the proposed algorithm is not an estimation efficient method. We carried out an additional experiment to test the boundary of the generalization

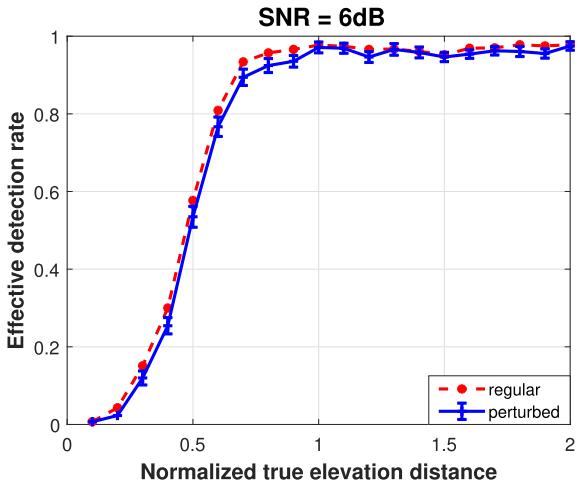


Fig. 16. Effective detection rate as a function of α at different baselines' distributions. The proposed algorithm shows a good generalization ability against baselines' discrepancy with the effective detection rate decreasing only 5%–8%.

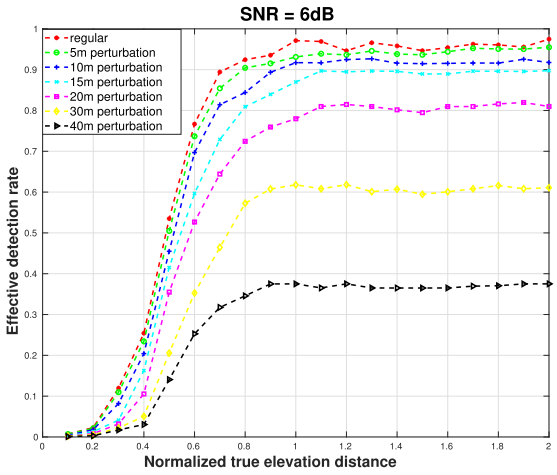


Fig. 17. Effective detection rate as a function of α at baselines with increasing perturbation. First, the effective detection rate decreases slowly with the increase in the baseline perturbation. When the perturbation is larger than 15 m, the performance of the proposed algorithm degrades dramatically.

ability by further increasing baseline perturbation. As we can see in Fig. 17, with the increase in the baseline discrepancy, the effective detection rate decreases slowly at first. When the perturbation is larger than 15 m, the performance of the proposed algorithm degrades dramatically. According to the test result, it indicates that 15 m might be the boundary for the proposed algorithm to have reasonable performance for the baseline setting in this simulation.

When we set out sights on global urban mapping using TomoSAR, the huge discrepancy between baselines of different data stacks will be a severe challenge. We still need to explore a more general and also computationally efficient algorithm.

B. Convergence Analysis

In this section, we propose to investigate the influence of CV-SGMUs on convergence performance in comparison with CV-SLSTMs. We use an RNN with six CV-SLSTMs

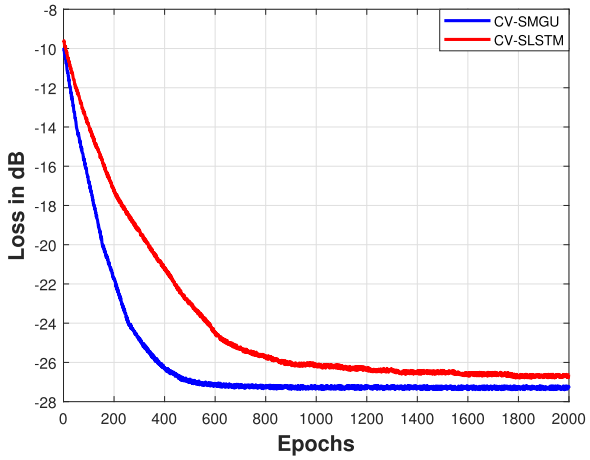


Fig. 18. Training loss [dB] versus epochs on simulated data. CV-SMGUs have faster convergence and lower overall loss.

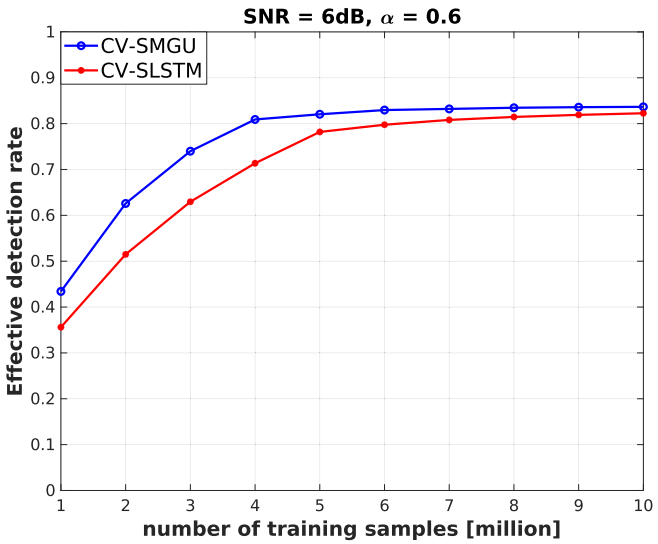


Fig. 19. Effective detection rate versus the number of training samples. The RNN with CV-SMGUs requires less training samples to achieve optimal performance.

as a baseline. Fig. 18 compares the objective loss [see (10)] with increasing training epochs. From Fig. 18, we can observe that CV-SMGUs contribute to faster convergence. To be specific, the RNN with CV-SMGUs needs only about 500 epochs to achieve convergence, while the RNN with CV-SLSTMs requires more than 1000 epochs to converge. Furthermore, CV-SMGUs lead to slightly lower overall costs than CV-SLSTMs.

C. Requirement of Training Data

As we have clarified in Section III-B, the CV-SMGU has only one gate, i.e., the minimum number of gates; thus, it has fewer trainable parameters and a simpler structure. In this experiment, we study how this simpler model contributes to reducing the requirement for training data. We compare two RNNs with six CV-SMGUs and six CV-SLSTMs, respectively, in terms of effective detection rate at 6-dB SNR. The distance between double scatterers was fixed at 0.6 Rayleigh resolution, and the double scatterers were set to have identical phase and amplitude. The result is demonstrated in Fig. 19. As can be seen, the RNN with CV-SMGUs has a better performance when the two RNNs are trained with the same amount of training samples. In addition, the RNN with CV-SMGUs requires obviously fewer training samples to achieve optimal performance.

VI. Conclusion

In this article, we proposed a novel gated RNN-based BPDN solver for sparse reconstruction. The proposed gated RNN adopted a novel architecture, termed SMGU, to avoid information loss caused by shrinkage by incorporating historical information into optimization. With the assistance of SMGUs, we are able to capture and maintain long-term dependence on information in previous layers. To be specific, important information will be automatically accumulated, while useless or redundant information will be forgotten in the dynamic of the network. Moreover, we extended the SMGU to the CV domain as CV-SMGU and applied it to solve TomoSAR inversion. Laboratory and real data experiments demonstrated that the proposed gated RNN built with CV-SMGUs outperforms the state-of-the-art deep learning-based TomoSAR method $\boldsymbol{\gamma}$-Net. The encouraging results open up a new prospect for SAR tomography using deep learning and motivate us to further investigate the potential of RNNs with gated units in practical TomoSAR processing.

Appendix
$\boldsymbol{\gamma}$-Net Formulation

Fig. 20 illustrates a K-layer $\boldsymbol{\gamma}$-Net. Each block in Fig. 20 indicates one layer of $\boldsymbol{\gamma}$-Net and is formally defined as

$$\tilde{\boldsymbol{\gamma}}_i = \eta_{ss_{\boldsymbol{\theta}_i}^{\rho^i}}\left\{\tilde{\boldsymbol{\gamma}}_{i-1} + \tilde{\mathbf{W}}^i\left(\tilde{\mathbf{g}} - \tilde{\mathbf{R}}\tilde{\boldsymbol{\gamma}}_{i-1}\right), \boldsymbol{\theta}_i\right\} \tag{12}$$

where

$$\tilde{\mathbf{W}}^i = \begin{bmatrix} \mathrm{Re}(\mathbf{W}^i) & -\mathrm{Im}(\mathbf{W}^i) \\ \mathrm{Im}(\mathbf{W}^i) & \mathrm{Re}(\mathbf{W}^i) \end{bmatrix}, \quad \tilde{\mathbf{R}} = \begin{bmatrix} \mathrm{Re}(\mathbf{R}) & -\mathrm{Im}(\mathbf{R}) \\ \mathrm{Im}(\mathbf{R}) & \mathrm{Re}(\mathbf{R}) \end{bmatrix}$$
$$\tilde{\mathbf{g}} = \begin{bmatrix} \mathrm{Re}(\mathbf{g}) \\ \mathrm{Im}(\mathbf{g}) \end{bmatrix}, \quad \tilde{\boldsymbol{\gamma}}_i = \begin{bmatrix} \mathrm{Re}(\hat{\boldsymbol{\gamma}}_i) \\ \mathrm{Im}(\hat{\boldsymbol{\gamma}}_i) \end{bmatrix}.$$

$\boldsymbol{\theta}_i = [\theta_i^1, \theta_i^2, \ldots, \theta_i^5]$ denotes the set of parameters to be learned for the piecewise linear function in the ith layer. $\mathbf{W}^i$ indicates the trainable weight matrix in the ith layer, and it is initialized using the system steering matrix $\mathbf{R}$ with $\mathbf{W^i} = \beta\mathbf{R}^H$. β is the step size. Usually, a proper step size can be taken as $(1/L_s)$, with L_s being the largest eigenvalue of $\mathbf{R^H R}$. $\hat{\boldsymbol{\gamma}}_i$ is the output of the ith layer. Re$(\cdot)$ and Im$(\cdot)$ denote the real and imaginary operators, respectively.

$\boldsymbol{SS}$ in $\boldsymbol{\gamma}$-Net indicates a special thresholding scheme called support selection, which is formally defined as follows:

$$\eta_{ss_{\boldsymbol{\theta}_i}^{\rho^i}}(\tilde{\boldsymbol{\gamma}}_i) = \begin{cases} \tilde{\boldsymbol{\gamma}}_i & i \in \mathcal{S}^{\rho^i}(\tilde{\boldsymbol{\gamma}}) \\ \eta_{pwl}(\tilde{\boldsymbol{\gamma}}_i, \boldsymbol{\theta}_i) & i \notin \mathcal{S}^{\rho^i}(\tilde{\boldsymbol{\gamma}}). \end{cases} \tag{13}$$

In the ith layer, the support selection will select ρ^i percentage of entries with the largest magnitude and trust them as "true support," which will be directly fed to the next layer, bypassing the shrinkage step. The remaining part will go through the shrinkage step as usual. The shrinkage is executed using the

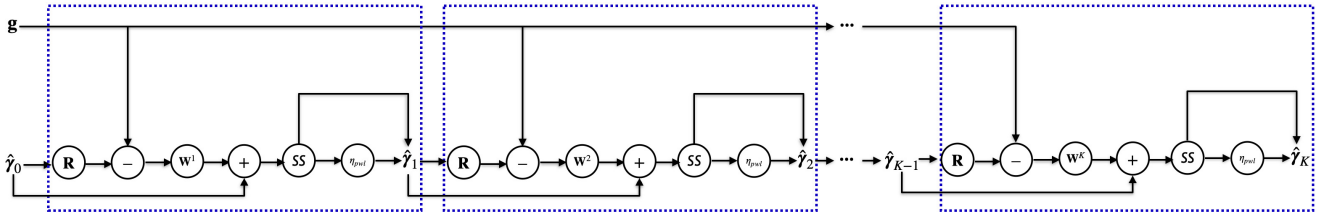


Fig. 20. Illustration the learning architecture of a K-layer $\boldsymbol{\gamma}$-Net.

piecewise linear function η_{pwl}, which is a novel shrinkage thresholding function to promote sparsity while improving convergence rate and reducing reconstruction error in the meanwhile and expressed as

$$\eta_{pwl}(\hat{\boldsymbol{\gamma}}, \boldsymbol{\theta}_i) = \begin{cases} \theta_i^3 \hat{\boldsymbol{\gamma}}, & |\hat{\boldsymbol{\gamma}}| \le \theta_i^1 \\ e^{j\cdot\angle\hat{\gamma}}\left[\theta_i^4\left(|\hat{\boldsymbol{\gamma}}| - \theta_i^1\right) + \theta_i^3\theta_i^1\right], & \theta_i^1 < |\hat{\boldsymbol{\gamma}}| \le \theta_i^2 \\ e^{j\cdot\angle\hat{\gamma}}\left[\theta_i^5\left(|\hat{\boldsymbol{\gamma}}| - \theta_i^2\right) + \theta_i^4\left(\theta_i^2 - \theta_i^1\right) + \theta_i^3\theta_i^1\right], & |\hat{\boldsymbol{\gamma}}| > \theta_i^2. \end{cases} \tag{14}$$

REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[2] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013.

[3] J. Bieniarz, E. Aguilera, X. X. Zhu, R. Müller, and P. Reinartz, "Joint sparsity model for multilook hyperspectral image unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 696–700, Apr. 2015.

[4] B. Zhang, W. Hong, and Y. Wu, "Sparse microwave imaging: Principles and applications," *Sci. China Inf. Sci.*, vol. 55, no. 8, p. 33, 2012.

[5] 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, Jun. 2010.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 118–121, Jan. 2007.

[8] 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.

[9] 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.

[10] Y. Li and S. Osher, "Coordinate descent optimization for L_1 minimization with application to compressed sensing; a greedy algorithm," *Inverse Problems Imag.*, vol. 3, no. 3, pp. 487–503, 2009.

[11] 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, Nov. 2010, doi: 10.1561/2200000016.

[12] S. J. Wright, *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics, 1997.

[13] 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.

[14] 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.

[15] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.

[16] J. R. Hershey, J. L. Roux, and W. Felix, "Deep unfolding: Model-based inspiration of novel deep architectures," 2014, *arXiv:1409.2574*, doi: 10.48550/arXiv.1409.2574.

[17] 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.

[18] 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.

[19] 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.

[20] S. Wei, J. Liang, M. Wang, J. Shi, X. Zhang, and J. Ran, "AF-AMPNet: A deep learning approach for sparse aperture ISAR imaging and autofocusing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[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)*, 2019, pp. 1–5, doi: 10.1109/ICSIDP47821.2019.9173392.

[22] S. Rangan, P. Schniter, and A. Fletcher, "Vector approximate message passing," *IRE Prof. Group Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[23] K. Qian, Y. Wang, Y. Shi, and X. X. Zhu, "γ-Net: Superresolving SAR tomographic inversion via deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[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] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.

[26] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*, doi: 10.48550/arXiv.1212.5701.

[27] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Feb. 2011.

[28] J. T. Zhou et al., "SC2Net: Sparse LSTMs for sparse coding," in *Proc. 32th AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 4588–4595.

[29] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS Workshop Deep Learn.*, Dec. 2014, pp. 1–9.

[30] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 2342–2350.

[31] K. Greff, R. K. Srivastava, J. Koutnìk, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[32] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *NIPS Workshop Deep Learn.*, 2014, pp. 1–9.

[33] G.-B. Zhou, J. Wu, C.-L. Zhang, and Z.-H. Zhou, "Minimal gated unit for recurrent neural networks," *Int. J. Automat. Comput.*, vol. 13, no. 3, pp. 226–234, Jun. 2016.

[34] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1–15.

[35] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 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.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*, doi: 10.48550/arXiv.1412.6980.

[37] F. R. 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

Kun Qian received the double B.Sc. degree in remote sensing and information engineering from Wuhan University, Wuhan, China, and in aerospace engineering and geodesy from the University of Stuttgart, Stuttgart, Germany, in 2016, and the M.Sc. degree in aerospace engineering and geodesy from the University of Stuttgart in 2018. He is currently pursuing the Ph.D. degree in data science in Earth observation with the 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, China, in 2008, and the M.Sc. and Dr.Ing. degrees from the Technical University of Munich, Munich, Germany, in 2010 and 2015, respectively.

In June and July 2014, he was a Guest Scientist with the Institute of Visual Computing, ETH Zürich, Zürich, Switzerland. He is currently a Guest Professor with the German International AI Future Laboratory (AI4EO), Technical University of Munich. He is also with the Department of EO Data Science, Remote Sensing Technology Institute, German Aerospace Center, Wessling, Germany, where he leads the working group Big SAR Data. His research interests include optimal and robust parameters estimation in multibaseline synthetic aperture radar interferometry (InSAR) techniques, multisensor fusion algorithms of synthetic aperture radar (SAR) and optical data, nonlinear optimization with complex numbers, machine learning in SAR, uncertainty quantification and mitigation in machine learning, 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.

Peter Jung (Member, IEEE) received the Dipl.Phys. degree in high-energy physics from Humboldt University, Berlin, Germany, in 2000, in cooperation with DESY Hamburg, Hamburg, Germany, and the Dr.rer.nat (Ph.D.) degree in Weyl–Heisenberg representations in communication theory from the Technical University of Berlin (TUB), Berlin, Germany, in 2007.

Since 2001, he has been with the Department of Broadband Mobile Communication Networks, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut (HHI), Berlin. Since 2004, he has been with the Fraunhofer German-Sino Laboratory for Mobile Communications, Berlin. He is currently working under Deutsche Forschungsgemeinschaft (DFG) grants at TUB in the field of signal processing, information and communication theory, and data science. He is also a Visiting Professor with the Technical University of Munich (TU Munich), Munich, Germany, and associated with the Munich AI Future Laboratory (AI4EO), Munich. He is giving lectures in compressed sensing, estimation theory, and inverse problems. His research interests include the area compressed sensing, machine learning, time–frequency analysis, dimension reduction, and randomized algorithms.

Dr. Jung is also a member of Verband Deutscher Elektrotechniker/Informationstechnische Gesellschaft (VDE/ITG).

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 currently 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 SAR interferometry (InSAR) processing, machine learning and deep learning for a 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. degree, the Dr.Ing. degree, and the Habilitation degree in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She was the Founding Head of the Department of EO Data Science, Remote Sensing Technology Institute, German Aerospace Center (DLR), Germany. Since 2019, she has been a Co-Coordinator of the Munich Data Science Research School, Munich. Since 2019, she has been heading the Helmholtz Artificial Intelligence, Munich—research fields: aeronautics, space, and transport. Since May 2020, she has been the Director of the International Future AI Lab "AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond," Munich. Since October 2020, she has been the 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 the University of California at Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is currently the Chair Professor for data science in Earth observation (former: signal processing in Earth observation) with TUM. She is currently a Visiting AI Professor with the Phi-Lab, European Space Agency (ESA), Paris. Her main research interests are remote sensing and Earth observation, signal processing, machine learning, and data science, with a special application focus on global urban mapping.

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

C K. Qian, Y. Wang, P. Jung, Y. Shi and X. X. Zhu, "HyperLISTA-ABT: An Ultra-light Unfolded Network for Accurate Multi-component Differential Tomographic SAR Inversion," submitted to IEEE TGRS

HyperLISTA-ABT: An Ultra-light Unfolded Network for Accurate Multi-component Differential Tomographic SAR Inversion

Kun Qian, Yuanyuan Wang, *Member, IEEE;* Peter Jung, *Member, IEEE;* Yilei Shi, *Member, IEEE;* and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*—Deep neural networks based on unrolled iterative algorithms have achieved remarkable success in sparse reconstruction applications, such as synthetic aperture radar (SAR) tomographic inversion (TomoSAR). However, the currently available deep learning-based TomoSAR algorithms are limited to three-dimensional (3D) reconstruction. The extension of deep learning-based algorithms to four-dimensional (4D) imaging, i.e., differential TomoSAR (D-TomoSAR) applications, is impeded mainly due to the high-dimensional weight matrices required by the network designed for D-TomoSAR inversion, which typically contain millions of freely trainable parameters. Learning such huge number of weights requires an enormous number of training samples, resulting in a large memory burden and excessive time consumption. To tackle this issue, we propose an efficient and accurate algorithm called HyperLISTA-ABT. The weights in HyperLISTA-ABT are determined in an analytical way according to a minimum coherence criterion, trimming the model down to an ultra-light one with only three hyperparameters. Additionally, HyperLISTA-ABT improves the global thresholding by utilizing an adaptive blockwise thresholding scheme, which applies block-coordinate techniques and conducts thresholding in local blocks, so that weak expressions and local features can be retained in the shrinkage step layer by layer. Simulations were performed and demonstrated the effectiveness of our approach, showing that HyperLISTA-ABT achieves superior computational efficiency and with no significant performance degradation compared to state-of-the-art methods. Real data experiments showed that a high-quality 4D point cloud could be reconstructed over a large area by the proposed HyperLISTA-ABT with affordable computational resources and in a fast time.**

***Index Terms*—Differential SAR tomography (D-TomoSAR), HyperLISTA, sparse recovery, unrolling algorithms.**

I. Introduction

Synthetic aperture radar tomography (TomoSAR) has attracted significant interest due to its capability in 3-D reconstruction, particularly for urban areas [1] [2] [3] [4]. Compressive sensing [5] [6] (CS)-based algorithms are usually preferred for solving TomoSAR inversion [7] [8] [9]. However, the heavy computational cost of CS-based methods makes them less applicable for large-scale processing. Among the different methods aiming to tackle this issue, deep neural networks have been employed in speeding up TomoSAR inversion. In the work presented in [10], TomoSAR inversion was approached as a classification problem, and a conventional convolutional neural network (CNN) was employed to solve the problem. However, this approach was limited to the detection of single scatterers, and it did not fully address the challenges of TomoSAR inversion for complex scenes with multiple scatterers and variations in the elevation direction. More recently, thanks to an emerging deep learning technique called deep unfolding [11], the authors proposed γ-Net in [12] for improving the unrolled iterative shrinkage thresholding algorithm (ISTA)-network. It was shown that γ-Net could succeed in accelerating the processing speed by 2-3 order of magnitude while maintaining a comparable super-resolution power and location accuracy compared to second-order CS-solvers. In addition, a gated recurrent structure, dubbed as complex-valued sparse minimal gated units (CV-SMGUs), was proposed in [13] that incorporates historical information into the dynamics of network, thus preserving the full information. As discussed in [13], CV-SMGUs could outperform γ-Net by a fair margin.

However, to the best of our knowledge, deep learning-based TomoSAR algorithms are to date still confined to 3-D reconstruction cases. Considering that spaceborne datasets are usually acquired in the repeat-pass mode at different time stamps, often over several years, it is necessary to additionally account for a potential deformation of objects in the estimation, such as seasonal motion caused by thermal dilation or linear motions like subsidence. The 4-D imaging technique taking into account additional deformation parameters is known as differential TomoSAR (D-TomoSAR) [1] [14] [15] [16].

The limitation of deep learning-based algorithms in solving D-TomoSAR inversion is mainly attributed to the high-dimensional weight matrices to be learned in the network. For modern deep learning-based algorithms, like γ-Net and CV-SMGUs, the size of the weight matrices is usually related to the discretization level. In D-TomoSAR cases, especially when multi-component motion terms are considered, weight matrices can easily contain over one million free trainable parameters. As a consequence, it would be extremely compu-

This work is supported by the European Research Council with the grant agreement No. [ERC-2016-StG-714087], Acronym: So2Sat, by the Helmholtz Association through the Framework of the Helmholtz Excellent Professorship "Data Science in Earth Observation - Big Data Fusion for Urban Research"(grant number: W2-W3-100), and by the German Federal Ministry of Education and Research in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (grant number: 01DD20001).

Corresponding author: Xiao Xiang Zhu.

K. Qian, Y. Wang, P. Jung and X. X. Zhu are with the Chair Data Science in Earth Observation, Technical University of Munich, Munich, Germany. (e-mails: {kun9361.qian,peter.jung,y.wang,xiaoxiang.zhu}@tum.de).

Y. Shi is with the Chair of Remote Sensing Technology, Technical University of Munich, Munich, Germany. (e-mail: yilei.shi@tum.de)

tationally inefficient to learn those weights without mentioning the enormous number of training samples required. A detailed analysis of this issue can be found in Section III.A of the present paper.

To tackle the computational challenges posed by learning huge weights, a pioneering solution was introduced in the seminal work by Liu et al. [17]. In their research, they proposed an analytical weight determination method. This method was further refined by them and extended into HyperLISTA [18]. By employing analytical weight determination, a novel perspective emerges for leveraging deep learning-based algorithms in D-TomoSAR inversion. Specifically, the optimization of weights using a data-free approach becomes possible, circumventing the need for an extensive number of training samples. This innovative approach effectively overcomes the limitations that were previously encountered in deep learning-based D-TomoSAR inversion.

However, it is important to note that, similar to LISTA, HyperLISTA employs a global thresholding scheme where a unified threshold is used to prune all entries. The choice of an appropriate threshold is thus crucial. A high threshold value may result in the loss of significant information [13], such as local features generated by echoed signals from dark scatterers. On the other hand, a low threshold value can delay convergence and yield a solution that lacks sufficient sparsity.

A. Contribution of the present study to the field

To overcome the aforementioned issue in to D-TomoSAR inversion, we proposed an ultra-light model, named HyperLISTA-ABT, that improves HyperLISTA [18] through incorporating an adaptive blockwise thresholding (ABT) scheme. Same as HyperLISTA, the proposed HyperLISTA-ABT can be viewed as an unrolled ISTA network, whereas the weight matrices therein can be determined with analytical optimization according to the minimum coherence criterion. A system matrix with low mutual coherence implies a recovery of high probability, which is the fundamental concept of compressive sensing. The adaptive blockwise thresholding scheme in HyperLISTA-ABT enables updating the block coordinates and conducting a shrinkage in local regions. Moreover, the blocksize is adjusted layer by layer for a better fine-focusing ability. The main contribution of this paper is listed as follows:

1) We propose the efficient and accurate algorithm HyperLISTA-ABT and, to the best of our knowledge, are the first to apply deep neural networks to solve D-TomoSAR and multi-component D-TomoSAR inversion.
2) We apply a block-coordinate technique and propose an adaptive blockwise thresholding scheme to replace global thresholding in most shrinkage thresholding methods. Therefore, the local features from a weakly echoed signal can be possibly retained.
3) We carry out a systematic performance evaluation using both simulated and real data. The results demonstrate that the proposed HyperLISTA-ABT provides competitive estimation accuracy and superior computational efficiency. Large-scale D-TomoSAR processing was conducted, demonstrated by a 4-D point cloud reconstruction over Las Vegas.

II. Background

A. High-dimensional SAR imaging model for D-TomoSAR

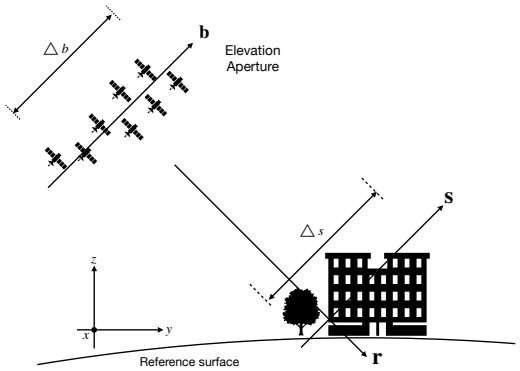


Fig. 1: SAR imaging geometry at a fixed azimuth position. The elevation synthetic aperture is built up by acquisition from slightly different incidence angles. Flight direction is orthogonal into the plane.

D-TomoSAR, employs multi-baseline and multi-temporal SAR acquisitions to estimate scatterering profiles. Based on the estimated scatterering profiles, we can reconstruct the 3-D distribution of scatterers along the elevation direction and the motion history assigned to each elevation position [1][15][19]. The following describes the D-TomoSAR imaging model:

$$g_n = \int_{\Delta s} \boldsymbol{\gamma}(s) \exp\left(-j2\pi\left(\xi_n s + 2d\left(s, t_n\right)/\lambda\right)\right) \mathrm{d}s. \quad (1)$$

where g_n is the complex-valued SAR acquisition at a certain azimuth-range pixel at time t_n $(n = 1, 2, \cdots, N)$; $\boldsymbol{\gamma}(s)$ denotes the scatterering profile along the elevation direction with an extent of $\triangle s$; $\xi = 2b_n/\lambda r$ is the elevation frequency proportional to the respective aperture position b_n; $d(s, t_n)$ depicts the line-of-sight (LOS) motion, which is a function of elevation and time. The LOS motion relative to the master acquisition can be modeled with a linear combination of M base functions $\tau_M(t_n)$

$$d\left(s, t_n\right) = \sum_{m=1}^{M} p_m(s)\tau_m\left(t_n\right) \quad (2)$$

where $p_m(s)$ is the corresponding motion coefficient to be estimated. The choice of the base functions $\tau_m(t_n)$ depends on the underlying physical motion processes. Great details about how to choose proper base functions can be found in [16]. Taking multi-component motion into consideration, we generalize Eq. (1) as:

$$g_n = \int \cdots \iint \gamma(s)\delta\left(p_1 - p_1(s), \ldots, p_M - p_M(s)\right) \quad (3)$$
$$\exp\left(j2\pi\left(\xi_n s + \eta_{1,n}p_1 + \ldots + \eta_{M,n}p_M\right)\right) dsdp_1 \ldots dp_M$$

The inversion of the system model with multi-component motion retrieves the elevation information as well as the

the motion history assigned to each elevation position, even if multiple scatterers are overlaid inside an resolution unit. Therefore, we can acquire a high-dimensional map of the scatterers. In the presence of noise $\boldsymbol{\varepsilon}$, the discrete high-dimensional D-TomoSAR model can be expressed as:

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{4}$$

where $\mathbf{g} \in \mathbb{C}^{N\times 1}$ is the complex-valued SAR measurement vector and $\mathbf{R} \in \mathbb{C}^{N\times L}$ is the irregular sampled Fourier transformation steering matrix, where N is the number of SAR acquisitions and L is the amount of the discretization in the signal to be reconstructed.

As investigated in [7], usually only a few (less than 4) scatterers are overlaid inside an individual pixel in urban areas, such that $\boldsymbol{\gamma}$ is sufficiently sparse so that retrieving $\boldsymbol{\gamma}$ can be formulated as a sparse reconstruction problem. Accordingly, solving $\boldsymbol{\gamma}$ in the presence of noise can be formulated as a basis pursuit denoising (BPDN) optimization problem, which can be expressed as follows:

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \left\{ \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \right\}, \tag{5}$$

where λ is the regularization parameter controlling the data-fit terms and the signal sparsity. Great details about how to choose a proper value of λ according to the noise level can be found in [20].

B. Review of the deep learning-based TomoSAR algorithms and their limitation in solving D-TomoSAR inversion

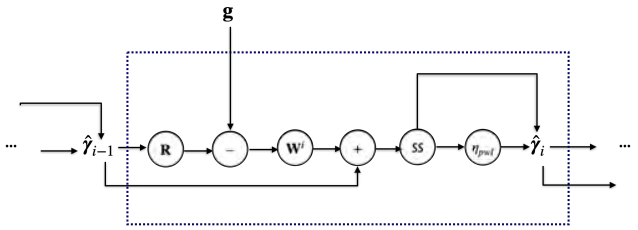


Fig. 2: Illustration of an intermediate layer in γ-Net.

In [12] [13], the respective authors proposed two advanced deep learning-based algorithms by improving unrolled neural networks. Experimental results on both laboratory and real data demonstrated their strong super-resolution power and high location accuracy. However, their application to date is still limited to 3-D reconstruction. Taking γ-Net as an example, we will explain the difficulty of applying deep learning-based algorithms for solving D-TomoSAR inversion. To start with, we briefly go through the basics of the γ-Net architecture. Fig. 2 illustrates us the structure of an intermediate layer of γ-Net, which can be formally defined as follows:

$$\hat{\boldsymbol{\gamma}}_k = \eta_{\theta_k}^{\rho^k}\{\hat{\boldsymbol{\gamma}}_{k-1} + \mathbf{W}_k^H(\mathbf{g} - \mathbf{R}\hat{\boldsymbol{\gamma}}_{k-1}), \boldsymbol{\theta_i}\} \tag{6}$$

More details about γ-Net can be found in [12].

As we can see, in each γ-Net layer, a weight matrix $\mathbf{W}_k$ of the size $N \times L$ needs to be learned. For 3-D reconstruction cases, the value of L is only determined by the grids number after the discretization of the elevation extent, thus it is typically in the range of hundreds for spaceborne sensors and $N \times L$ will be in thousands then. However, this number increases exponentially in D-TomoSAR cases when multi-component motion terms, usually linear and periodic motions, are taken into consideration. The training of the network then becomes conversely a challenge due to the tremendous amount of free trainable parameters. For instance, when we consider two motion terms, i.e., linear and seasonal motion, the value of L will be determined by the product of the discretized grid numbers along each direction $L = L_s \times L_v \times L_a$, where L_s, L_v, and L_a indicate the discretization levels for elevation, linear motion, and seasonal motion, respectively. A very conservative level of discretization in elevation, linear motion, and periodic motion for TerraSAR-X image stacks L_s, L_v, and L_a would be 200, 50, and 50. When multiplied, the value of L will then be 0.5 million, meaning that, there will be millions of parameters to be learned in each weight matrix. Such large weight matrices result in two unavoidable downsides. First, the model tends to converge at the ground truth instead of the LASSO minimizer, because the update direction $\mathbf{W}^H(\mathbf{g} - \mathbf{R}\gamma)$ does not align with the gradient of the l_2 term in the LASSO objective $\mathbf{R}^H(\mathbf{g} - \mathbf{R}\gamma)$. Therefore, we always need to train the model in a supervised way. Consequently, a massive number of training samples are required to train the model with huge weight matrices, thus making the training procedure extremely inefficient. Second, the training of the huge model requires a significant amount of GPU memory, which is usually not feasible with consumer-level GPUs.

III. METHODOLOGY

A. HyperLISTA with adaptive blockwise thresholding (HyperLISTA-ABT)

To circumvent the tedious and troublesome model training caused by needing to learn huge weight matrices, an analytical weight optimization method, which is based on coherence minimization, was proposed in ALISTA [17] to determine the weights in an unrolled neural network designed for sparse recovery, such as LISTA. ALISTA combines the superior empirical performance of fully learned methods and significantly reduces the number of parameters, leaving only thresholds and stepsize parameters to be learned. In addition, an ultra-light model, called HyperLISTA, was proposed in [18], which further trimmed down the training complexity. In HyperLISTA, weight matrices can be computed in a similar way to [17] and the training is reduced to tuning only three hyperparameters from the data. The following shows us the formal update rules of HyperLISTA:

$$\begin{aligned}\boldsymbol{\gamma}_{k+1} = \eta_{\theta^k}^{p^k}(\boldsymbol{\gamma}_k + \boldsymbol{W}^H(\boldsymbol{g} - \boldsymbol{R}\boldsymbol{\gamma}_k) \\ +\beta^k(\boldsymbol{\gamma}_k - \boldsymbol{\gamma}_{k-1}))\end{aligned} \tag{7}$$

where

$$\theta^k = c_1 \left\| \boldsymbol{R}^+ (\boldsymbol{R}\boldsymbol{\gamma}_k - \boldsymbol{g}) \right\|_1 \tag{8}$$

$$\beta^k = c_2 \left\| \boldsymbol{\gamma}_k \right\|_0 \tag{9}$$

$$p^k = c_3 \min\left(\log\left(\frac{\|\boldsymbol{R}^+\boldsymbol{g}\|_1}{\|\boldsymbol{R}^+(\boldsymbol{R}\boldsymbol{\gamma}_k - \boldsymbol{g})\|_1} \right), L \right) \tag{10}$$

where c_1, c_2, and c_3 indicate the three hyperparameters to be tuned. It is possible to learn the hyperparameters via backpropagation, albeit this method may be an overkill as it involves passing gradients through deep neural network layers to learn just three parameters. Less computationally expensive methods, such as grid search, could be employed to obtain a set of proper hyperparameters. Despite providing a bit less accurate estimate, the empirical findings in [18] showed that HyperLISTA is robust to perturbations in the values of c1, c2, and c3. In grid search, a coarse grid is first applied to find an interest region, and then this is zoom-in with a fine-grained grid. The hyperparameters are determined by minimizing the normalized mean square error (NMSE) over the simulated ground truth. The NMSE is defined as:

$$\mathrm{NMSE} = \frac{1}{T} \sum \frac{\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_2^2}{\|\boldsymbol{\gamma}\|_2^2} \tag{11}$$

where T denotes the number of samples, and $\eta_{\theta^{(k)}}^{p^{(k)}}$ is the soft-thresholding function combined with the support selection scheme,

$$\eta_{\theta_{(k)}}^{\rho^{(k)}}(\boldsymbol{\gamma}^{(k)}) = \begin{cases} \boldsymbol{\gamma}^{(k)} & i \in \mathcal{S}^{\rho^{(k)}}(\boldsymbol{\gamma}) \\ \eta_{st}(\boldsymbol{\gamma}^{(k)}, \theta_{(k)}) & i \notin \mathcal{S}^{\rho^{(k)}}(\boldsymbol{\gamma}) \end{cases} . \tag{12}$$

$\mathcal{S}^{\rho^{(k)}}(\boldsymbol{\gamma})$ contains the entries with the $\rho^{(k)}$ largest magnitudes. $\mathbf{W}$ denotes the optimized weight matrix determined with the minimum coherence criterion, which is defined as follows:

$$\begin{aligned} \hat{\mathbf{W}} &= \arg\min_{\mathbf{W}} \mu(\mathbf{W}, \mathbf{R}) \\ &= \arg\min_{\mathbf{W}} \inf_{\mathbf{W} \in \mathbb{C}_{N \times L}} \max_{i \neq j} \mathbf{W}_{:,i}^T \mathbf{R}_{:,j} \\ &\text{s.t. } \forall i \in \{1, \ldots, L\} : \mathbf{W}_{:,i}^T \mathbf{R}_{:,i} = 1 \end{aligned} \tag{13}$$

Rigorous proof of the convergence and recovery upper and lower bound of HyperLISTA can be found in [18]. An efficient numerical algorithm to calculate the optimized weights is discussed in the Appendix.

Inspired by the outstanding efficiency and performance demonstrated in [18], we consider HyperLISTA should have great potential in our high-dimensional D-TomoSAR inversion. However, through experiments, we discovered a drawback of HyperLISTA when applied to TomoSAR. Similar to most thresholding algorithms, HyperLISTA suffers from an inherent limitation caused by the global thresholding scheme. Precisely, in the signal projection process for identifying the presence of a dictionary atom within the signal, the selection of an appropriate threshold is of utmost importance. The threshold should be chosen carefully to account for both strong and weak spikes in the reflectivity profile. By selecting a well-suited threshold, the signal projection can distinguish between significant spikes and noise, enabling an accurate identification of dictionary atoms within the signal. However, when utilizing HyperLISTA and other methods that employ global thresholding, the task of selecting an optimal threshold becomes exceedingly challenging. The use of a global threshold implies that the same threshold value is applied uniformly across all entries in the signal. This approach may lead to suboptimal results, as a threshold that effectively captures strong spikes might inadvertently suppress weaker but still meaningful spikes in the reflectivity profile. Consequently, we usually need to choose a relatively small c_1 to have a small threshold so that we can maintain some small spikes caused by reflection from dark scatterers. Otherwise, the information of dark scatterers would be discarded in the thresholding step layer by layer. However, the use of a small threshold brings about two main problems. First, the convergence would be considerably slow. Second, small thresholds yield solutions that are not sparse enough.

To cope with the aforementioned issue and better leverage the power of HyperLISTA in our application, we propose HyperLISTA-ABT, which is an improvement of the original HyperLISTA by incorporation of an adaptive blockwise thresholding (ABT) scheme that explores a local thresholding strategy. The advantages of HyperLISTA-ABT is three-fold. First, it conducts the thresholding in each local block, thus allowing for a more refined thresholding process and possibly retaining weak expressions of reflections from dark scatterers. Then, it becomes possible to better capture the diverse range of spike magnitudes encountered in the signal, enhancing the accuracy and reliability of the reflectivity profile characterization. Second, HyperLISA-ABT has been shown to be more efficient since it updates only one block of variables at each time instead of updating all the variables together. Therefore, HyperLISTA-ABT has been found to be more appropriate for our large-scale and high-dimensional application. Last but not the least, HyperLISTA-ABT reduces the blocksize layerwise and contributes to a better fine-focusing ability.

According to [21] [22], the update rules of HyperLISTA-ABT after applying block coordinate techniques can be written as:

$$\begin{aligned} \boldsymbol{\gamma}_{i_p}^{k+1} = \eta_{\theta_{i_p}^k}(\boldsymbol{\gamma}_{i_p}^k + \boldsymbol{W}_{i_p}^T(\boldsymbol{y} - \boldsymbol{R}_{i_p}\boldsymbol{\gamma}_{i_p}^k) \\ + \beta_{i_p}^k(\boldsymbol{\gamma}_{i_p}^k - \boldsymbol{\gamma}_{i_p}^{k-1})) \end{aligned} \tag{14}$$

where i_p is the index of the updated block. To clarify, in HyperLISTA-ABT, we remove the support selection scheme and just use the conventional soft-thresholding function. The threshold $\theta^{(k)}$ and the factor $\beta^{(k)}$ are determined for each block as well:

$$\theta_{i_p}^k = c_1 \left\| \boldsymbol{R}_{i_p}^+ \left(\boldsymbol{R}_{i_p} \boldsymbol{\gamma}_{i_p}^k - \boldsymbol{g} \right) \right\|_1 \tag{15}$$

$$\beta_{i_p}^k = c_2 \left\| \boldsymbol{\gamma}_{i_p}^k \right\|_0 \tag{16}$$

where $c_1 > 0$, $c_2 > 0$, and $c_3 \in (0, 1)$ are the three hyperparameters. Notably, c_3 is a latent hyperparameter and plays a crucial role in controlling the blocksize despite it not explicitly appearing in the formula. In our application, we usually initialize the blocksize according to the grid number within half of the Rayleigh resolution. The block is chosen with a random variants scheme where i_p follows the probability distribution given by:

$$P_{i_p} = \frac{L_{i_p}}{\sum_{j=1}^{J} L_{i_p}}, \quad i_p = 1, \ldots, J \tag{17}$$

where J is the number of blocks and $L_{i_p} = ||R_{i_p}^T R_{i_p}||$. All the hyperparameters c_1, c_2, and c_3 can be selected using the same grid search method as in HyperLISTA.

With the blockwise threshlding scheme, local features and weak expressions can be possibly retained. This is due to the fact that many elements of the entries are not strictly driven to zero but to some extremely small value, thus making the output not strictly sparse. Therefore, a post-processing is usually required to clean the output and make it sparse. The framework of the proposed HyperLISTA-ABT is summarized in the following table.

Algorithm 1 Summary of the proposed algorithm

Generate steering matrix R for given baselines
  Analytic weight optimization $\mathbf{W}$ according to Eq. (13)
Tuning of hyperparameters
  Simulate ground truth of reflectivity profile $\boldsymbol{\gamma}$ [12]
  Simulate noise-free SAR acquisitions $\mathbf{g} = \mathbf{R}\boldsymbol{\gamma}$
  Grid search to determine the hyperparameters by minimizing NMSE over simulated data
Inference
  Init: $\boldsymbol{\gamma}^{(0)} = \mathbf{R}^H\mathbf{g}$ and blocksize B_1
  for $k = 1, 2, \cdots, K$ **do**
    Determine the number of blocks J_k based on the blocksize B_k
    for $i_p = 1, 2, \cdots, J_k$ **do**

$$\begin{aligned}\boldsymbol{\gamma}_{i_p}^{(k+1)} &= \eta_{\theta_{i_p}^{(k)}}(\boldsymbol{\gamma}_{i_p}^{(k)} + \boldsymbol{W}_{i_p}^T(\boldsymbol{y} - \boldsymbol{R}_{i_p}\boldsymbol{\gamma}_{i_p}^{(k)}) \\ &\quad + \beta_{i_p}^{(k)}(\boldsymbol{\gamma}_{i_p}^{(k)} - \boldsymbol{\gamma}_{i_p}^{(k-1)})) \\ \theta_{i_p}^{(k)} &= c_1 \left\| \boldsymbol{R}_{i_p}^{+} \left(\boldsymbol{R}_{i_p}\boldsymbol{\gamma}_{i_p}^{(k)} - \boldsymbol{g} \right) \right\|_1 \\ \beta_{i_p}^{(k)} &= c_2 \left\| \boldsymbol{\gamma}_{i_p}^{(k)} \right\|_0\end{aligned}$$

    end for
    Update blocksize with $B_{k+1} = c_3 \cdot B_k$
  end for
Output clean-up
Model order selection and final estimation

IV. Simulations

To demonstrate the improvement of the proposed HyperLISTA-ABT to the original HyperLISTA and compare it to the state-of-the-art CS-based and deep learning-based methods, we first conducted experiments based on TomoSAR inversion using simulated data. Since existing deep learning-based algorithms are not feasible to use with D-TomoSAR cases as explained in Section I, we only focused on TomoSAR inversion for 3-D reconstruction in the simulation.

A. Simulation setup

In the simulation, we conducted a well-known TomoSAR benchmark test [1] [7] using the same simulation settings as used in [12] [13]. specifically, we simulated an interferometric stack containing 25 baselines that are regularly distributed in the range of -135m to 135m and a two-scatterer mixture in each resolution cell. We used the ***effective detection rate***, which is able to simultaneously reflect the super resolution power and elevation estimation accuracy, to fairly evaluate the performance. Detailed definition of the ***effective detection rate*** can be found in [12] [13].

B. Performance improvement compared to the original HyperLISTA

The first experiment set out to study the performance improvement of HyperLISTA-ABT compared to the original HyperLISTA. In this experiment, the overlaid double scatterers were simulated to have identical scattering phase varying amplitude ratios. The two algorithms were set to have 15 layers, which is a typical number for a LISTA network and its variants. Fig. 3 shows the effective detection rate of HyperLISTA-ABT and the original HyperLISTA as a function of the normalized elevation distance between the simulated double scatterers at 6dB SNR at different amplitude ratios. The results demonstrate that HyperLISTA-ABT achieved a significantly higher effective detection rate than the original HyperLISTA. Both algorithms (in fact, all other methods) experience performance degradation with respect to an increase in amplitude ratio. This is attributed to two main factors. Firstly, dark scatterers experience a large bias in their elevation estimates at high amplitude ratios due to their elevation estimates approaching the more prominent ones. Consequently, many detections of double scatterers will not be recognized as effective due to the large elevation estimate bias. Secondly, the energy of dark scatterers is close to the noise level at high amplitude ratios. This makes it particularly challenging for HyperLISTA, which employs a global thresholding scheme, to detect the local features of dark scatterers. Further elaborating, when high-intensity scatterers are present in the signal, their strong energy can overshadow the low-energy regions where dark scatterers are located. This overshadowing effect can lead to the suppression or even annihilation of the weaker expressions associated with the dark scatterers. Consequently, the presence of these strong intensity scatterers can mask or obscure the signals originating from the dark scatterers, making their detection and characterization challenging. In contrast, HyperLISTA-ABT conducts thresholding in each local block, which can allow retaining local information and, thus, it can detect dark scatterers. This results in a higher effective detection rate at high amplitude ratios.

C. Comparison with the state-of-the-art algorithms

In this section, we compared HyperLISTA-ABT to other state-of-the-art algorithms for further evaluation, which are deep learning-based algorithms γ-Net [12] and CV-SMGUs [13], as well as the traditional CS-based method SL1MMER [23] with second-order optimization. To highlight the super-resolution ability of these methods, we also involved a conventional spectral estimator SVD-Wiener [1] as a baseline in the comparison.

The comparison was first based on the effective detection rate. Two different scenarios were taken into consideration:

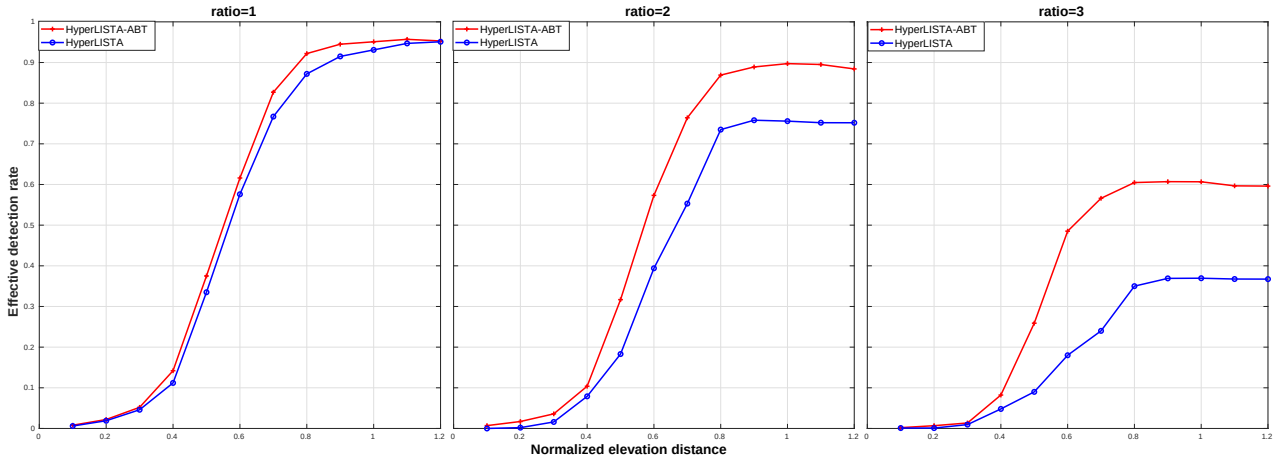


Fig. 3: Effective detection rate of HyperLISTA-ABT and the original HyperLISTA with respect to the normalized elevation distance at different amplitude ratios. The overlaid double scatterers were set to have an identical phase and the SNR level was 6dB. HyperLISTA-ABT significantly outperformed HyperLISTA at high amplitude ratios between the scatterers.

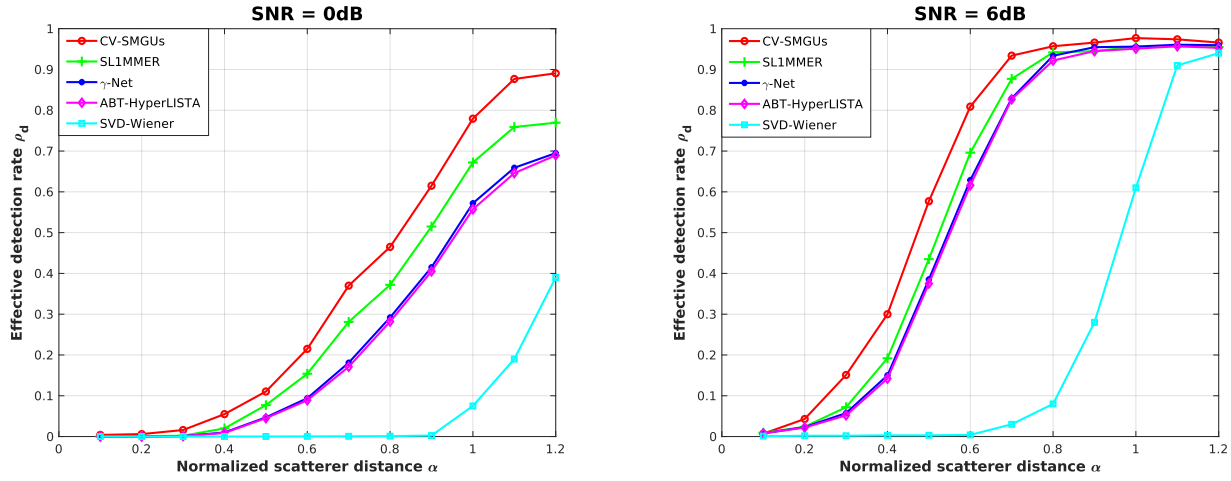
SNR $\in \{0, 6\}$dB, which represents a noisy case, and a regular case with a typical SNR level in a high-resolution spaceborne SAR image. The comparison results are demonstrated in Fig. 4. At each discrete normalized elevation distance, 0.2 million Monte Carlo trials with an identical phase and amplitude, which represents the worst case [23] in TomoSAR inversion, were simulated. The deep learning-based algorithms γ-Net and CV-SMGUs were built with 12 and 6 hidden layers, respectively. The training followed the same training strategy introduced in [12] [13] and was carried out using a single NVIDIA RTX2080 GPU. For HyperLISTA-ABT, the training involved analytical weight optimization and determining the hyperparameters via the grid search method. The number of iterations in HyperLISTA-ABT was set as 15.

From the comparison results, we can see that all the methods except the conventional spectral estimator SVD-Wiener showed a great super-resolution power. The proposed HyperLISTA-ABT delivered almost the same super-resolution ability as γ-Net and approached the performance of SL1MMER in both scenarios. When focusing solely on the effective detection rate, it was challenging to proclaim a clear advantage of the proposed HyperLISTA-ABT method over the existing state-of-the-art approaches. In fact, when comparing it to CV-SMGUs, we could observe a slight underperformance. However, all the state-of-the-art methods come with a relatively high computational cost. Both γ-Net and CV-SMGU require tailored training according to the baseline distribution of the stack. SL1MMER is a model-based algorithm, thus needs no training, yet requires significantly computational time for solving the L1-norm minimization.

We tested and recorded the time consumption of different algorithms for processing the 0.2 million Monte Carlo trials as well as the requirements for training data. The results are summarized in Table I. To clarify, all inference was conducted using a local CPU for a fair comparison. As can be seen in Table I, it took about ten hours for the deep learning-based algorithms to process 0.2 million Monte Carlo trials, which was predominantly the training time. In addition, a large amount of training samples was essential as well. For SL1MMER, it took about 20 hours for the processing since the iterative second-order optimization is computationally expensive. Further inspecting the table, we can see that HyperLISTA-ABT showed similar efficiency in the inference as the other deep learning-based algorithms. However, HyperLISTA-ABT required no training data and it took much less time for the training. In total, HyperLISTA-ABT speeded up the processing by about one order of magnitude compared to the other algorithms tested in the experiment.

Upon evaluating the performance and efficiency, it was observed that HyperLISTA-ABT achieved comparable performance to existing state-of-the-art methods while significantly improving the computational efficiency by approximately one order of magnitude. This is especially advantageous in the multi-component D-TomoSAR case. The application of the aforementioned deep learning-based algorithms and SL1MMER are very limited in the D-TomoSAR case due to the need of time-consuming model training and the heavy computational expense. On the contrary, the application of HyperLISTA-ABT can be easily extended to computationally efficient D-TomoSAR processing. Therefore, HyperLISTA-ABT is a more applicable approach for the large-scale processing of real data.

Furthermore, HyperLISTA-ABT demonstrates superior transferability compared to deep learning-based algorithms. Deep learning models are typically trained to fit specific baseline configurations, such as a fixed number of SAR acquisitions and a specific baseline distribution. While they may exhibit satisfactory generalizability to small baseline discrepancies [12] [13], directly applying a trained deep learning model to a new data stack with a different number of acquisitions or a completely different baseline distribution is not feasible. In such cases, time-consuming retraining of the



(a) (b)

Fig. 4: Detection rate P_d as a function of the normalized elevation distance between the simulated facade and ground with SNR = 0 dB and 6 dB, N = 25, and phase difference $\triangle\phi = 0$ (worst case) under 0.2 million Monte Carlo trials.

TABLE I: Comparison of the number of required training samples and time consumption for processing 0.2 million Monte Carlo trials with each algorithm. The training time of HyperLISTA-ABT indicates the combined duration of both the analytic weight optimization process and the tuning of hyperparameters. It provides a measure of the overall time required for these essential steps.

Algorithm	number of training samples	training time	inference time	total time consumption	transferability
CV-SMGUs	4 million	$\approx$ 10 hours	$\approx$ 0.25 h	$\approx$ 10 h	low
γ-Net	3 million	$\approx$ 8 hours	$\approx$ 0.17 h	$\approx$ 8 h	low
SL1MMER	-	-	$\approx$ 20 h	$\approx$ 20 h	high
HyperLISTA-ABT	-	$<$ 1 hour	$\approx$ 0.25 h	$\approx$ 1 h	medium

model becomes necessary, resulting in low transferability.

In contrast, HyperLISTA-ABT offers better transferability. Although it requires analytical optimization of the weight matrix for each new data stack, the efficiency of the analytical optimization process allows for scalability and improved transferability. This finding highlights the potential of HyperLISTA-ABT in enabling global urban mapping using TomoSAR, as it can be effectively applied to diverse data stacks with varying acquisition configurations and baseline distributions.

V. Real Data Experiment

A. Bellagio hotel

In this real data experiment, due to the fact that there was no available ground truth, we purposely used the same data as in [24] so that we can compare our results to the results obtained with SL1MMER. The datastack was composed of 29 TerraSAR-X high-resolution spotlight images covering the Bellagio Hotel in Las Vegas, whose baseline distribution is illustrated in Fig. 5. The slant-range resolution was 0.6m and the azimuth resolution was 1.1m. The elevation aperture size of about 270m resulted in the inherent elevation resolution ρ_s to be about 40m, i.e. approximately 20m resolution in height since the incidence angle here was 31.8°. An optical image and the SAR mean intensity image of the test site are shown in Fig. 6

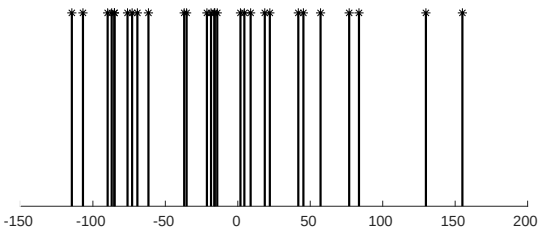


Fig. 5: Effective baselines of the 29 TerraSAR-X high-resolution spotlight images.

As for the D-TomoSAR system model, a time wrap operation assuming only sinusoidal seasonal motion was adopted as in [16] because no long-term linear motion was observed

during the acquisition period of the test area.

In Fig. 7, we compare the estimated elevation and amplitude of the seasonal motion of the detected single scatterers and the top layer of the detected double scatterers. From Fig. 7(a), we can see a smooth gradation of the elevation estimates from the building bottom top, which suggests a reasonable elevation estimation by HyperLISTA-ABT. Moreover, we can see that there is no significant difference between the results of HyperLISTA-ABT and SL1MMER, implying that HyperLISTA-ABT had similar performance to SL1MMER. In addition, Fig. 8 shows the layover separation ability of HyperLISTA-ABT. As can be seen, the two layers of double scatterers were detected and separated by HyperLISTA-ABT. The top layer was mainly caused by signals from the roof and facade of the high rise building while the bottom layer was caused by signals from the ground structures.

We also conducted some numerical comparisons of both algorithms. First, we compared the percentage of pixels detected as zero, one, and two scatterers by both algorithms. Compared to SL1MMER, we found that HyperLISTA-ABT detected more pixels as coherent sactterers. This does not necessarily mean that HyperLISTA-ABT had a better detection ability since there was no ground truth. We believe HyperLISTA-ABT detected more scatterers because HyperLISTA-ABT tends to maintain weak signals, which could be reflections of dark scatterers but also outliers caused by noise interference. The false detection of noise as coherent scatterers causes a speckle-like noise in the reconstruction result. Model order selection and post-processing techniques like spatial filtering can further mitigate such outliers.

For further evaluation, we compared the elevation estimates differences of scatterers detected by both algorithms. A histogram of the elevation estimates differences is shown in Fig. 9. It can be observed that most of the elevation estimates differences were within 1 meter. This observation indicates that both algorithms yielded comparable results in terms of elevation estimation, instilling confidence in their reliability and reasonableness. Furthermore, this similarity in estimation accuracy suggests that HyperLISTA-ABT performed on par with SL1MMER. Moreover, it is worth mentioning that it took more than three weeks for SL1MMER to finish the D-TomoSAR processing over the test site, whereas it only took HyperLISTA-ABT several hours to complete the processing.

TABLE II: Percentage of scatterers detection for the two algorithms.

Algorithm	Percentage of detection as		
	0 scatterer	1 scatterer	2 scatterers
HyperLISTA-ABT	48.48 %	44.09 %	7.43 %
SL1MMER	49.41 %	43.63 %	6.96 %

B. Large demonstration

In this section, we applied HyperLISTA-ABT to TerraSAR-X high-resolution spotlight data over a large area surrounding the convention center in Las Vegas. The stack was composed of 29 acquisitions covering a time period from July 2009 to June 2010, during which the test area was undergoing a pronounced subsidence centered at the convention center. Therefore, the acquisitions were characterized by a multi-component nonlinear motion combining linear and thermal-dilation-induced seasonal motion. Fig. 10 shows us an optical image and the SAR mean intensity map. The red box indicates the "epiccenter" undergoing subsidence around the convention center.

Fig. 11 illustrates us the reconstructed elevation estimates as well as the estimated amplitude maps of the two different motions. As we can see from the surface model generated from the elevation estimates in Fig. 11(a), we can capture the shapes of individual buildings and the surrounding infrastructures, like roads, at a detailed level. In addition, Fig. 11(b) shows that clear deformation caused by seasonal motion can be observed in the metallic building structures since they were affected by thermal dilation more seriously compared to surrounding infrastructures. Furthermore, as illustrated in Fig. 11(c), it could be observed that the magnitude of the linear subsidence increased as the scatterer getting closer to the "epicenter". These results are consistent with the fact, thus validating the effectiveness of HyperLISTA-ABT for multi-component nonlinear motion estimation and giving confidence that HyperLISTA-ABT can be applied in large-scale D-TomoSAR processing.

VI. Conclusion

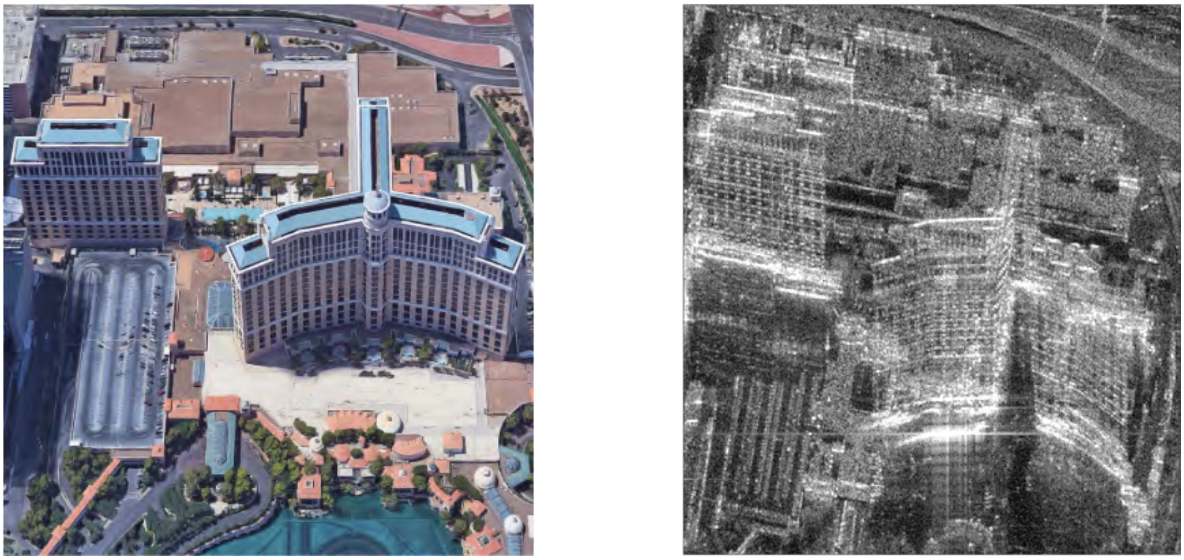
This paper proposes HyperLISTA-ABT to address the gap in applying deep neural networks for solving D-TomoSAR inversion. Unlike traditional methods that learn weights directly from data, HyperLISTA-ABT computes the weights with an analytical optimization technique by minimizing generalized mutual coherence. Additionally, HyperLISTA-ABT introduces an adaptive blockwise thresholding scheme that applies block coordinate techniques to accelerate the algorithm. Moreover, it conducts thresholding in local blocks to retain weak expressions of reflection from dark scatterers and consider more local features. Laboratory experiments for 3D reconstruction confirmed the efficiency of HyperLISTA-ABT in estimation. Moreover, tests on real data over a large area demonstrated that HyperLISTA-ABT can reconstruct high-quality 4D point clouds, making it an efficient and accurate algorithm for future large-scale or even global D-TomoSAR processing.

Appendix

A. Efficient algorithm to calculate weight analytically

As discussed in [17], it is difficult to solve Eq. (13) directly and Eq. (13) can be reformulated as minimizing the Frobenius norm of $\mathbf{W}^H\mathbf{R}$ over a linear constraint. Defining $\mathbf{W} = \mathbf{G}^H\mathbf{G}\mathbf{R}$ ($\mathbf{G} \in \mathbb{C}^{N\times N}$ is named as the Gram matrix), the minimization of the Frobenius norm reads:

$$\min_{G} \left\| R^H \boldsymbol{G}^H \boldsymbol{G} \boldsymbol{R} - \boldsymbol{I} \right\|_F^2, \quad \text{s.t. } \operatorname{diag}\left(\boldsymbol{R}^H \boldsymbol{G}^H \boldsymbol{G} \boldsymbol{R} \right) = \mathbf{1}. \tag{18}$$



(a) (b)

Fig. 6: Test site. (a): optical image from Google Earth, (b): SAR mean intensity image

However, it is hard to handle the constraint in the above problem (18). As a solution, a matrix $\mathbf{D} = \mathbf{GR} \in \mathbb{C}^{N\times L}$ is introduced and we use the following method as an alternative:

$$\min_{\boldsymbol{G},\boldsymbol{D}} \left\|\boldsymbol{D}^T\boldsymbol{D} - \boldsymbol{I}\right\|_F^2 + \frac{1}{\alpha}\|\boldsymbol{D} - \boldsymbol{GR}\|_F^2, \text{ s.t. } \operatorname{diag}\left(\boldsymbol{D}^T\boldsymbol{D}\right) = \mathbf{1}. \tag{19}$$

With a proper $\alpha > 0$, the solution to Eq. (19) approximates Eq. (18) and we obtain the optimized weights accordingly. The steps for solving the optimization problem (19) are described as follows.

First, $\mathbf{G}$ is fixed and we update $\mathbf{D}$ with the projected gradient descent (PGD):

$$\boldsymbol{D} \leftarrow \mathcal{P}\left(\boldsymbol{D} - \zeta\boldsymbol{D}\left(\boldsymbol{D}^H\boldsymbol{D} - \boldsymbol{I}\right) - \frac{\zeta}{\alpha}(\boldsymbol{D} - \boldsymbol{GR})\right) \tag{20}$$

where $\mathcal{P}$ denotes the projection operator on the constraint $\operatorname{diag}\left(\boldsymbol{D}^T\boldsymbol{D}\right) = \mathbf{1}$, so that each column of $\mathbf{D}$ will be normalized; and ζ is the stepsize. Hereafter, we fix D and update the minimizer of G with

$$\boldsymbol{G} \leftarrow \mathbf{D}\mathbf{R}^+ \tag{21}$$

where $\mathbf{R}^+$ represents the Moore-Penrose pseudoinverse of the steering matrix $\mathbf{R}$. Then, we repeat the procedure until $\mathbf{D} \approx \mathbf{GR}$. The whole algorithm is summarized in **Algorithm 2**.

Algorithm 2 Efficient algorithm for analytical weight optimization

Input: the steering matrix $\mathbf{R}$
Init: $\mathbf{D} = \mathbf{R}$, $\mathbf{G} = \mathbf{I}$, $\zeta = \alpha = 0.1$
for iter$= 1, 2, \cdots$ until **convergence do**
  update $\mathbf{D}$ with (20)
  update $\mathbf{G}$ with (21)
  Compute $f_1 = \left\|\boldsymbol{D}^H\boldsymbol{D} - \boldsymbol{I}\right\|_F^2$
  Compute $f_2 = \left\|(\boldsymbol{GR})^H\boldsymbol{GA} - \boldsymbol{I}\right\|_F^2$
  if two consecutive f_1s are close enough **then**
    $\zeta = 0.1\zeta$
    $\alpha = 0.1\alpha$
    if f_1 and f_2 are close enough **then**
      break
    end
  end
end
Output: $\mathbf{W} = \mathbf{G}^H\mathbf{GR}$

REFERENCES

[1] X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4296–4308, 2010, 00125.

[2] D. Reale, G. Fornaro, A. Pauciullo, X. Zhu, and R. Bamler, "Tomographic imaging and monitoring of buildings with very high resolution SAR data," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 661–665, 2011.

[3] X. X. Zhu and M. Shahzad, "Facade reconstruction using multiview spaceborne tomosar point clouds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3541–3552, 2014.

[4] G. Fornaro, F. Lombardini, A. Pauciullo, D. Reale, and F. Viviani, "Tomographic processing of interferometric SAR data: Developments, applications, and future research perspectives," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 41–50, 2014.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[7] X. Zhu and R. Bamler, "Tomographic SAR inversion by l_1 -norm regularization—the compressive sensing approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, 2010.

[8] 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 Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 6148–6158, 10 2018.

[9] X. X. Zhu, N. Ge, and M. Shahzad, "Joint sparsity in SAR tomography for urban mapping," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 8, pp. 1498–1509, 2015.

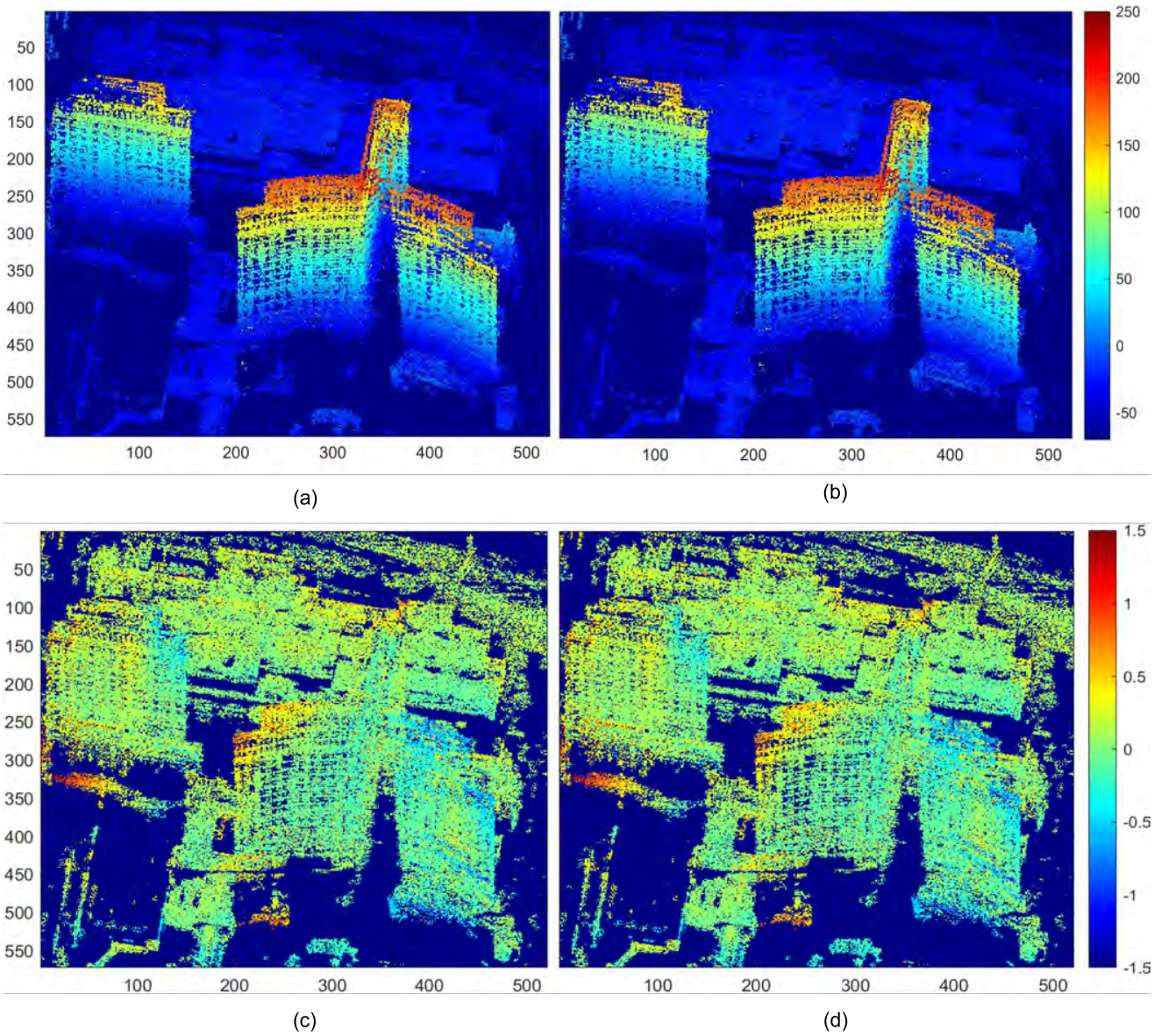


Fig. 7: Color-coded reconstruction results of the test site. (a) Elevation estimates using HyperLISTA-ABT in meters, (b) elevation estimates using SL1MMER in meters, (c) estimated amplitude of seasonal motion using HyperLISTA-ABT in centimeters, (d) estimated amplitude of seasonal motion using SL1MMER in centimeters.

[10] A. Budillon, A. C. Johnsy, G. Schirinzi, and S. Vitale, "SAR tomography based on deep learning," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 3625–3628.

[11] J. R. Hershey, J. L. Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," 2014. [Online]. Available: https://arxiv.org/abs/1409.2574

[12] K. Qian, Y. Wang, Y. Shi, and X. X. Zhu, "γ-net: Superresolving SAR tomographic inversion via deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[13] K. Qian, Y. Wang, P. Jung, Y. Shi, and X. X. Zhu, "Basis pursuit denoising via recurrent neural network applied to super-resolving SAR tomography," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[14] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 224–237, 2009.

[15] F. Lombardini, "Differential tomography: a new framework for SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 37–44, 2005.

[16] X. X. Zhu and R. Bamler, "Let's do the time warp: Multicomponent nonlinear motion estimation in differential SAR tomography," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 735–739, 2011.

[17] J. Liu, X. Chen, Z. Wang, and W. Yin, "ALISTA: Analytic weights are as good as learned weights in LISTA," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=B1lnzn0ctQ

[18] X. Chen, J. Liu, Z. Wang, and W. Yin, "Hyperparameter tuning is all you need for LISTA," in *Advances in Neural Information Processing Systems*, 2021.

[19] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 224–237, 2009.

[20] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition

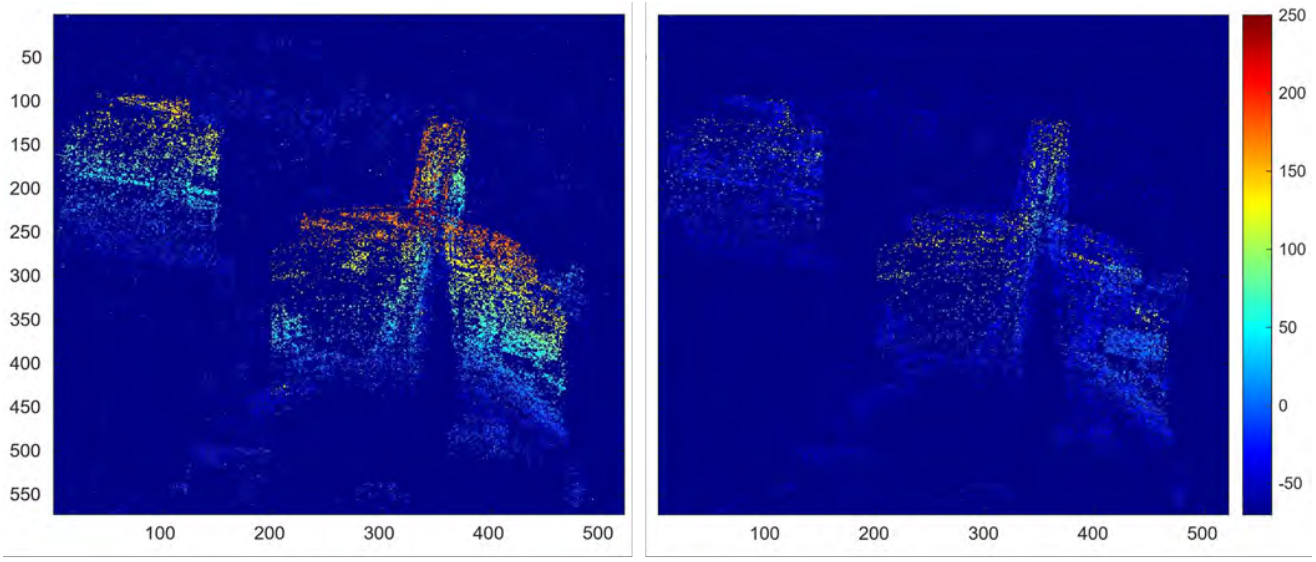


Fig. 8: Color-coded elevation estimates of the top and bottom layers of detected double scatterers using HyperLISTA-ABT. (a) Top layer, mostly caused by reflections from the building roof and facade, (b) bottom layer, mostly caused by reflections from low infrastructures and the ground.

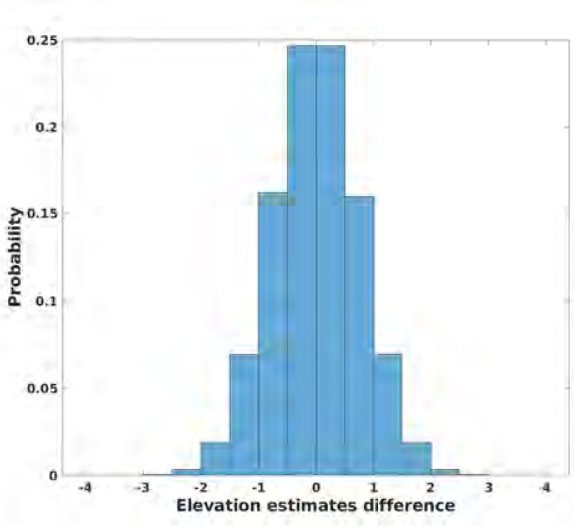


Fig. 9: Histogram of elevation estimates differences between HyperLISTA-ABT and SL1MMER.

by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, p. 129–159, Jan. 2001.
[21] Y. Xu and W. Yin, "Block stochastic gradient iteration for convex and nonconvex optimization," *SIAM Journal on Optimization*, vol. 25, no. 3, pp. 1686–1716, 2015.
[22] Z. Peng, T. Wu, Y. Xu, M. Yan, and W. Yin, "Coordinate-friendly structures, algorithms and applications," *Annals of Mathematical Sciences and Applications*, vol. 1, no. 1, pp. 57–119, 2016.
[23] X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic sar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 247–258, 2012.
[24] X. X. Zhu and R. Bamler, "Demonstration of super-resolution for tomographic SAR imaging in urban environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 8, pp. 3150–3157, 2012.

Kun Qian received double B.Sc. degree in Remote Sensing and Information Engineering from Wuhan University, Wuhan, China and Aerospace Engineering and Geodesy from University of Stuttgart, Stuttgart, Germany, respectively, in 2016, and M.Sc. degree in Aerospace Engineering and Geodesy from University of Stuttgart, Stuttgart, Germany in 2018. He is pursuing the Ph.D. degree with Data Science in Earth Observation, Technical Unversity of Munich, Munich, Germany. His research focus includes data-driven methods, deep unfolding algorithms and their application in multi-baseline SAR tomography.

Yuanyuan Wang (S'08–M'15) received the B.Eng. degree (Hons.) in Electrical Engineering from The Hong Kong Polytechnic University, Hong Kong, China in 2008, and the M.Sc. and Dr.-Ing. degree from the Technical University of Munich, Munich, Germany, in 2010 and 2015, respectively. In June and July 2014, he was a Guest Scientist with the Institute of Visual Computing, ETH Zürich, Zürich, Switzerland. He is currently a Guest Professor at the German Interantional AI Future Lab: AI4EO, at the Technical University of Munich. He is also with the Department of EO Data Science, in the Remote Sensing Technology Institute of the German Aerospace Center, where he leads the working group Big 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, uncertainty quantification and mitigation in machine learning, 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. He is a Member of the IEEE.

Fig. 10: Demonstration of the large test area. (a) Optical image from Google Earth, (b) SAR mean intensity map in dB. The red box in (b) indicates the area undergoing subsidence.

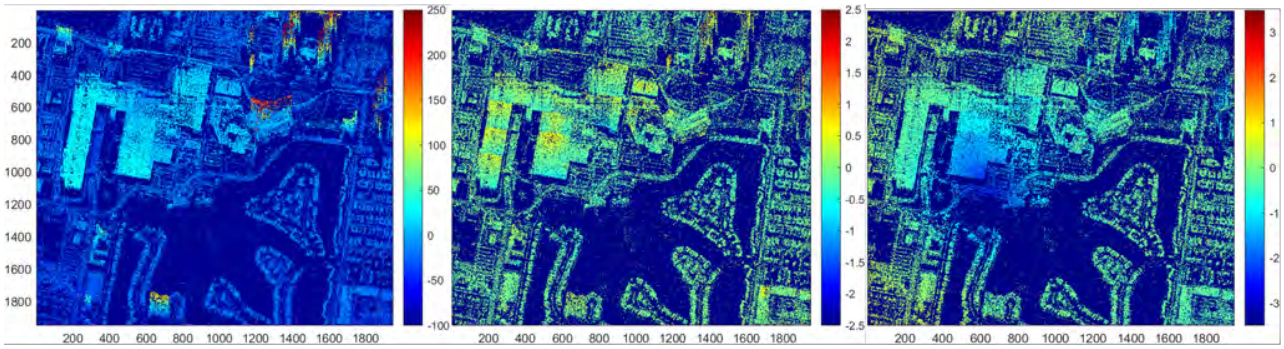


Fig. 11: Demonstration of color-coded elevation estimates and estimated amplitude of multi-component motion. (a) Elevation estimates in meters, (b) estimated amplitude of seasonal motion in centimeters, (c) estimated amplitude of linear motion in centimeters/year.

Peter Jung (Member IEEE) received the Dipl.-Phys. degree in high energy physics from Humboldt University, Berlin, Germany, in 2000, in cooperation with DESY Hamburg, and the Dr.-rer.nat (Ph.D.) degree in Weyl–Heisenberg representations in communication theory with the Technical University of Berlin (TUB), Germany, in 2007. Since 2001, he has been with the Department of Broadband Mobile Communication Networks, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut (HHI), and since 2004 with the Fraunhofer German-Sino Laboratory for Mobile Communications. He is currently working under DFG grants at TUB in the field of signal processing, information and communication theory, and data science. He is also a Visiting Professor with TU Munich and associated with the Munich AI Future Laboratory (AI4EO). His current research interests include the area compressed sensing, machine learning, time–frequency analysis, dimension reduction, and randomized algorithms. He is giving lectures in compressed sensing, estimation theory and inverse problems. He is also a member of VDE/ITG.

Yilei Shi (Member, IEEE) received the Dipl.-Ing. degree in mechanical engineering and the Dr.-Ing. degree in signal processing from 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 SAR and InSAR processing, machine learning and deep learning for variety of data sources, such as SAR, optical images, and medical images, and PDE-related numerical modeling and computing.

Xiao Xiang Zhu (S'10–M'12–SM'14–F'21) received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her "Habilitation" in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively.

She is currently the Chair Professor for Data Science in Earth Observation (former: Signal Processing in Earth Observation) at Technical University of Munich (TUM) and was the Founding Head of the Department "EO Data Science" at the Remote Sensing Technology Institute, German Aerospace Center (DLR). Since 2019, Zhu is a co-coordinator of the Munich Data Science Research School (www.muds.de). Since 2019 She also heads the Helmholtz Artificial Intelligence – Research Field "Aeronautics, Space and Transport". Since May 2020, she is the director of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond", Munich, Germany. Since October 2020, she also serves as a co-director of the Munich Data Science Institute (MDSI), TUM. Prof. Zhu was a guest scientist or visiting professor at 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, United States in 2009, 2014, 2015 and 2016, respectively. She is currently a visiting AI professor at ESA's Phi-lab. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with a special application focus on global urban mapping.

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 IEEE Transactions on Geoscience and Remote Sensing and serves as the area editor responsible for special issues of IEEE Signal Processing Magazine. She is a Fellow of IEEE.