Potential of Deep Learning in SAR Tomographic Inversion of Very Small Interferometric Stacks

Kun Qian[a], Yuanyuan Wang[a,b], and Xiaoxiang Zhu[a,b]
[a]Data Science in Earth Observation, Technical University of Munich, Munich, Germany
[b]Department of EO Data Science, Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Germany

Abstract

SAR tomography (TomoSAR) has been extensively applied in 3-D reconstruction in dense urban areas. Compressive sensing (CS)-based algorithms are generally considered as the state-of-the-art methods in super-resolving TomoSAR, in particular in the single-look case. TomoSAR algorithms, including the CS-based ones, usually require a fairly large number of images to achieve a reliable reconstruction, because large error and especially bias occur in low number of measurements. In addition, CS-based algorithms are extremely computationally expensive due to their sparse reconstruction. These factors hinder their practical use. This paper demonstrates the potential of a novel and computationally efficient deep learning algorithm for TomoSAR on very small interferometric stacks. Investigation of the super-resolution ability shows that the proposed algorithm outperforms the state-of-the-art CS-based TomoSAR algorithm by a fair margin when limited acquisitions are available. Test on real TanDEM-X data with 6 interferograms also shows high-quality 3-D reconstruction.

1 Introduction

Synthetic aperture radar tomography (TomoSAR) is an advanced SAR interferometric technique and has been widely employed for 3-D urban mapping. It is capable of retrieving 3-D spatial information of scatterers and the elevation reflectivity profile. Compressive sensing [8] (CS-) based sparse reconstruction algorithms are considered the state of the art in TomoSAR processing because only a few significant scatterers are overlaid in each resolution unit in urban areas [1]. The strong super-resolution power and high estimation accuracy of CS-based methods contribute to achieving the best performance using high-resolution SAR data like TerraSAR-X. However, CS-based methods have two drawbacks, and cannot be applied in practical urban processing. First, CS-based methods suffer from a heavy computational burden since the sparse reconstruction cannot be solved analytically, and time-consuming iterative solvers need to be employed. Second, CS-based methods usually require a fairly large number of images to achieve reliable reconstruction. It is studied in [7] that by using the state of the art, SL1MMER, a minimum number of 11 acquisitions are necessary for reasonable reconstruction. Whereas usually only a few (<10) acquisitions are available for most cities. [9] proposed M-SL1MMER, which applies joint sparsity to achieve satisfactory reconstruction using only 6 acquisitions. Unfortunately, this method can only be applied in areas, where geographic information data is available, because precise geometric prior is demanded. In [10], a non-local filter is added to CS-TomoSAR. The non-local filter increased the SNR level dramatically, thus improving the reconstruction significantly, whereas the elevation will be shifted after non-local filtering.

In this paper, we introduced a computationally efficient TomoSAR inversion algorithm in [2][3], which unrolls iterative shrinkage thresholding algorithm (ISTA) [4] as a complex-valued feedforward neural network with side connection to mimic a CS solver for sparse reconstruction. We evaluated the performance of the proposed deep learning-based algorithm for TomoSAR inversion using only small interferometric stacks to inspect its potential for large-scale TomoSAR processing, where usually limited acquisitions are available.

2 TomoSAR imaging model

In this section, we first introduce the TomoSAR imaging model, which can be expressed as follows (we do not consider deformation term in this paper):

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

where $\gamma(s)$ depicts the reflectivity profiles along the elevation direction. $\xi_n = -2b_n/(\lambda r)$ denotes the elevation frequency at the aperture position b_n. λ and r refer to the wavelength and the range, respectively.

After discretizing the reflectivity profile along the elevation direction s, the canonical form of discrete TomoSAR imaging model is given by (in the presence of noise $\boldsymbol{\varepsilon}$):

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (2)$$

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. N and L are the number of images and discretization level, respectively.
In urban areas, it is shown in [1] that usually only a few (0-4) scatterers are overlaid in a single resolution unit, meaning that the reflectivity profile $\boldsymbol{\gamma}$ is usually consisted of a few non-zero elements with dominant magnitude. In this vein, the reflectivity profile whose entries are predominantly zero can be estimated using compressive sensing (CS)-based sparse reconstruction 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 factor balancing the sparsity and data-fitting terms. It should be adjusted according to the noise level as well as the desired sparsity level.

3 Deep learning based SAR Tomography

3.1 Learned iterative shrinkage thresholding algorithm (LISTA)

To solve Eq. (3), ISTA [4] is known as a popular method due to its simplicity and efficiency. Each iteration of ISTA is defined by

$$\begin{aligned} \hat{\boldsymbol{\gamma}}_i &= \eta_{st}(\hat{\boldsymbol{\gamma}}_{i-i} + \beta\mathbf{R}^H\mathbf{b}_{i-1}, \theta_i)\,, \\ &\text{with} \quad \mathbf{b}_i = \mathbf{g} - \mathbf{R}\hat{\boldsymbol{\gamma}}_i \end{aligned} \tag{4}$$

where β is the stepsize and η_{st} indicates the soft-thresholding function. By rewriting Eq. (4) as the following form:

$$\hat{\boldsymbol{\gamma}}_i = \eta_{st}\left\{\mathbf{W}_1^i\mathbf{g} + \mathbf{W}_2^i\hat{\boldsymbol{\gamma}}_{i-1}, \theta_i\right\} \tag{5}$$

where $\mathbf{W_1^i} = \beta\mathbf{R}^H$ and $\mathbf{W}_2^i = \mathbf{I} - \beta\mathbf{R}^{\mathrm{H}}\mathbf{R}$, we have the formulation of the i^{th} layer of LISTA [5], which unrolls a recurrent neural network and truncates it into several iteration, thus leading to a feed-forward neural network with side-connection. Figure 1 illustrates the learning architecture of a K-layer LISTA. As can be seen, each layer of CV-LISTA replicates the computation of an ISTA iteration, despite the fact that the conventional soft-thresholding function is replaced by the piecewise linear function. The major difference is that the weights matrices $\mathbf{W}_1^i$, $\mathbf{W}_2^i$ as well as the threshold θ_i in each layer of CV-LISTA are not pre-determined. Those parameters can be learned by minimizing the following loss function over the training data $\{(\mathbf{g}_i, \boldsymbol{\gamma}_i)\}_{i=1}^T$

$$\underset{\boldsymbol{\Psi}}{\text{minimize}}\ \mathcal{L}(\boldsymbol{\Psi}) = \frac{1}{T}\sum_{i=1}^{T} ||\hat{\boldsymbol{\gamma}}(\boldsymbol{\Psi}, \mathbf{g}) - \boldsymbol{\gamma}||_2^2, \tag{6}$$

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 trainable parameters.

3.2 Complex-valued (CV-)LISTA for TomoSAR

To apply LISTA to solve the TomoSAR inversion, we need to extend LISTA to complex-valued domain. 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 Eq. (5)

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

where

$$\begin{aligned} \widetilde{\mathbf{W}}_j^i &= \begin{bmatrix} \Re(\mathbf{W}_j^i) & -\Im(\mathbf{W}_j^i) \\ \Im(\mathbf{W}_j^i) & \Re(\mathbf{W}_j^i) \end{bmatrix}, \\ \tilde{\mathbf{g}} &= \begin{bmatrix} \Re(\mathbf{g}) \\ \Im(\mathbf{g}) \end{bmatrix}, \\ \tilde{\boldsymbol{\gamma}} &= \begin{bmatrix} \Re(\hat{\boldsymbol{\gamma}}) \\ \Im(\hat{\boldsymbol{\gamma}}) \end{bmatrix} \end{aligned} \tag{8}$$

with $j = 1, 2$ and $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary operators, respectively. In addition, we replaced the soft-thresholding function by the piecewise linear function for shrinkage, which is formally defined by

$$\eta_{pwl}(\hat{\boldsymbol{\gamma}}, \boldsymbol{\theta}) = \begin{cases} \theta_3\hat{\boldsymbol{\gamma}}, & |\hat{\boldsymbol{\gamma}}| \le \theta_1 \\ e^{i\cdot\arg(\hat{\boldsymbol{\gamma}})}[\theta_4(|\hat{\boldsymbol{\gamma}}| - \theta_1) & \\ \quad +\theta_3\theta_1], & \theta_1 < |\hat{\boldsymbol{\gamma}}| \le \theta_2 \\ e^{i\cdot\arg(\hat{\boldsymbol{\gamma}})}[\theta_5(|\hat{\boldsymbol{\gamma}}| - \theta_2) & \\ \quad +\theta_4(\theta_2 - \theta_1) + \theta_3\theta_1], & |\hat{\boldsymbol{\gamma}}| > \theta_2 \end{cases}. \tag{9}$$

As one can see from the learning architecture of the CV-LISTA, the output of the intermediate layer in the CV-LISTA is generated exclusively on the output of the previous layer, thus leading to "error propagation" phenomenon. Specifically speaking, minor errors will be propagated and amplified in the upcoming layers in the feedforward phase. More seriously, once useful information is discarded in the first few layers, the upcoming layers are no longer possible to utilize the discarded information. The conventional soft-thresholding function simply prunes element with small magnitude and it is very likely to result in information loss. On the contrary, the piecewise linear function just further shrinks elements with small magnitude so that it is able to maintain information and execute shrinkage in the meanwhile. We illustrate the both function in figure 2 for an intuitive view of comparison.

4 Experiments

4.1 Simulation setup

To evaluate the performance of the proposed algorithm, we implemented experiments on simulated data. We used the same system parameters of a small stack, whose baselines are listed in Table 1, to simulate two-scatterer mixtures. The elevation aperture size of about 940m results in about 12m inherent elevation resolution, i.e. Rayleigh resolution ρ_s. The double scatterers were set to have identical phase and amplitude, i.e. the worst case for the TomoSAR processing.

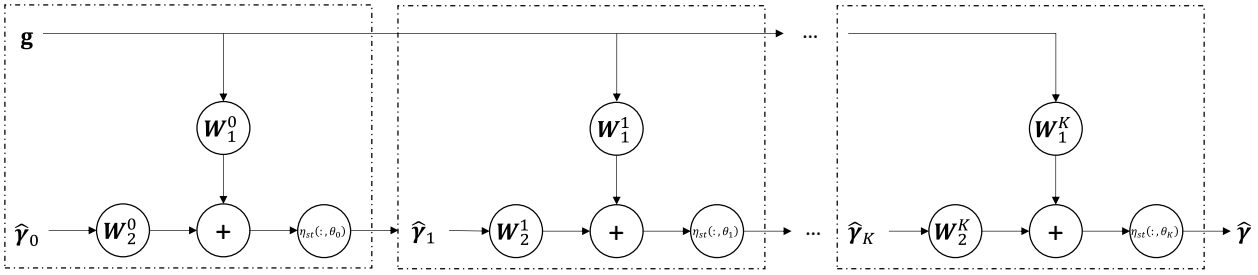


Figure 1 Unfolded LISTA architecture. A K-layer LISTA unrolls the RNN and truncates it into K iterations, thus leading to a side-connected feedforward neural network.

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

Table 1 Detailed introduction of spatial baselines we used for data simulation

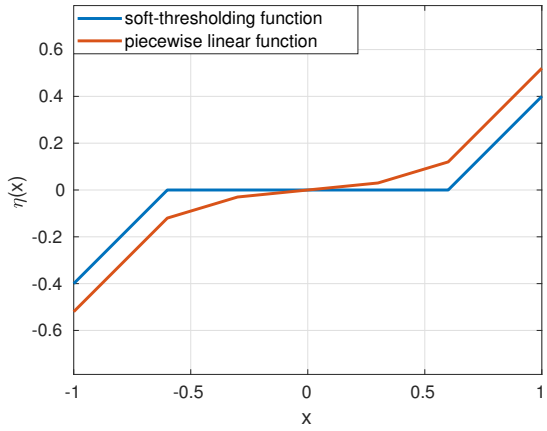


Figure 2 Comparison between the piecewise linear function and soft-thresholding function. Instead of pruning the elements with small magnitude, the piecewise linear function just further minifies them, thus possibly avoiding the information loss.

4.2 Performance evaluation

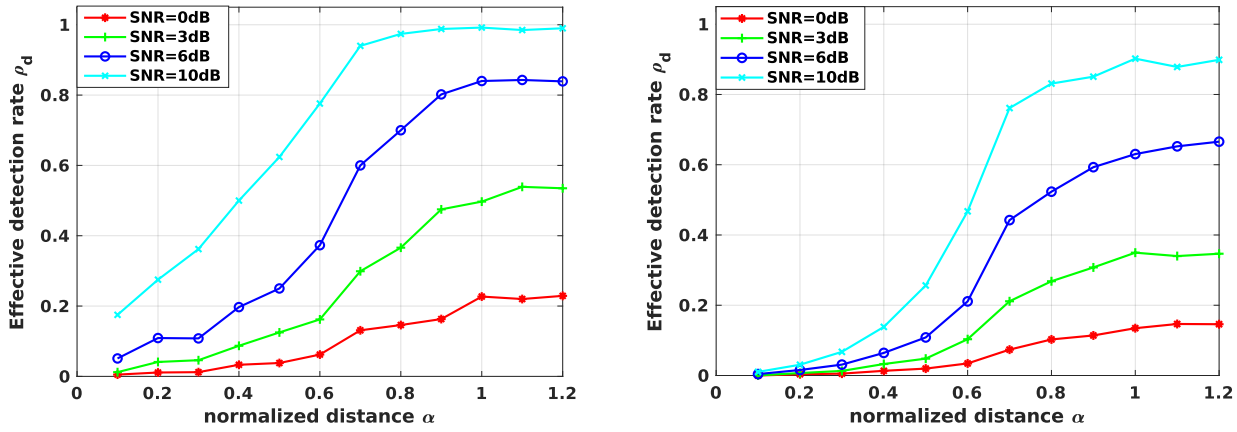
In the experiment, we performed a well-known TomoSAR benchmark test [1][6]. We simulated overlaid double scatterers with increasing elevation distance d_s to mimic a facade-ground interaction. Four different scenarios were taken into consideration with SNR$\in \{0, 3, 6, 10\}$ dB. We compared the performance of the proposed algorithm with the state-of-the-art TomoSAR method SL1MMER [7] w.r.t the super-resolution power and estimation accuracy for small stack processing. We used the ***effective detection rate*** to fairly evaluate the super-resolution power. An effective detection of double scatterers is defined as:

1. the hypothesis test correctly decides two scatterers for a double-scatterers signal;
2. the estimated elevation of ***both*** detected double scatterers are within ± 3 times CRLB w.r.t their true elevation;
3. both elevation estimates are also within $\pm 0.5\ d_s$ w.r.t their true elevation.

The third criterion is seldom seen in the literature. However, it is necessary, because in extremely super-resolving cases, 3 times CRLB will become much larger than the elevation distance. Hence, it cannot be used as a accountable measure for reasonable estimates. $\pm 0.5\ d_s$ is a much stricter constraint in such cases, which will reflect the true performance of the algorithm. We plotted the effective detection rate as a function of the normalized elevation distance α, which is defined as the ratio between the distance between double scatterers and the Rayleigh resolution ρ_s. Figure 4 demonstrates us the comparison of the two methods. As we can see, the effective detection rate of the proposed algorithm is about 20% higher than SL1MMER in moderate super-resolving cases at 6 dB SNR. In addition, at high SNR level, i.e. 10 dB, we can see that the proposed algorithm outperforms SL1MMER by an obvious margin, especially in extremely super-resolving cases where the double scatterers are spaced within 0.5 ρ_s. In extremely noisy cases, i.e. at 0 dB SNR level, we can see that the proposed algorithm still performs slightly better, despite the fact that the two algorithms have quite low effective detection rate in super-resolving cases.

4.3 Real-data experiment

In this section, we worked with a stack of six TanDEM-X high resolution interferograms acquired in pursuit monostatic mode with 0.6m slant-range resolution and 0.25m azimuth resolution. Fig. 4 shows an optical image from Google Earth and the SAR intensity image of the test site. Due to negligible temporal baselines, the atmospheric effects and deformation are ignored in this experiment. The DLR's integrated wide area processor (IWAP) was employed for image co-registration and phase calibration. We simulated a training dataset with 4 million samples using the real baseline distribution and trained the CV-LISTA with the simulated training data. After training, a well-trained CV-LISTA can be directly applied to real data.
The reconstruction results using CV-LISTA is demon-



(a) proposed

(b) SL1MMER

Figure 3 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. The effective detection rate of the proposed algorithm is about 20% higher than SL1MMER in moderate super-resolving cases at 6 dB SNR. In addition, at high SNR level, i.e. 10 dB, we can see that the proposed algorithm outperforms SL1MMER by an obvious margin especially in extremely super-resolving cases, where the double scatterers are spaced within 0.5 ρ_s. In extremely noisy cases, i.e. at 0 dB SNR level, we can see that the proposed algorithm still performs slightly better, despite the fact that the two algorithms have quite low effective detection rate in super-resolving cases.

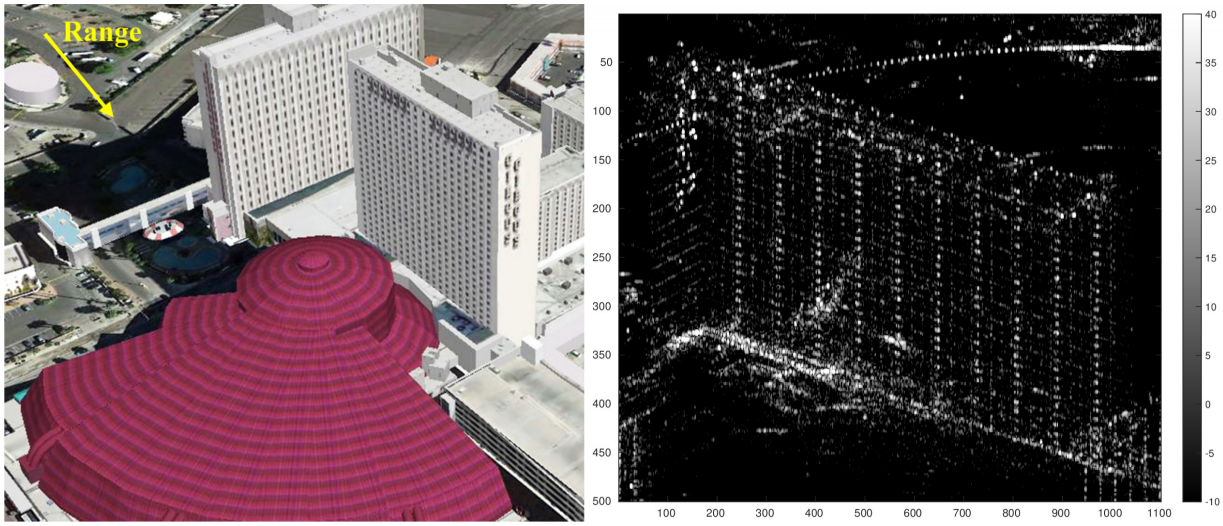
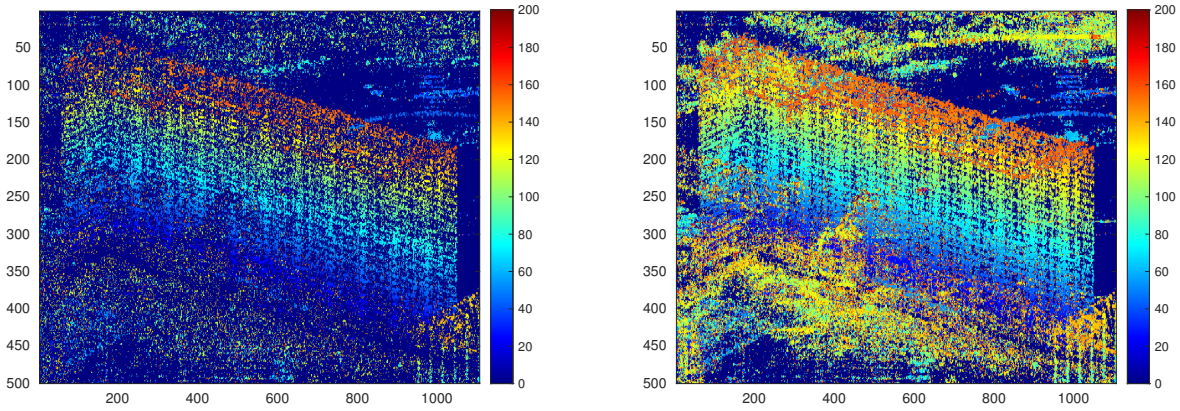


Figure 4 Test site. Left: optical image from Google Earth, right: SAR mean intensity image

strated in Figs. 5 and 6. Fig. 5 depicts the color-coded reconstruction of detected single scatterers and the detected single scatterers combined with the top layer of dectected double scsatterers from left to right. As one can see in Fig. 5, CV-LISTA is able to detect dense double scatterers, which contribute to significant information increment and complete the structure of individual buildings in the test site. Fig. 6 presents the separation of overlaid double scatterers. Closer inspection in Fig. 6 shows that most reflection from roof and facade are overlaid at the top of buildings. By applying CV-LISTA, we can effectively distinguish reflection from roof and facade

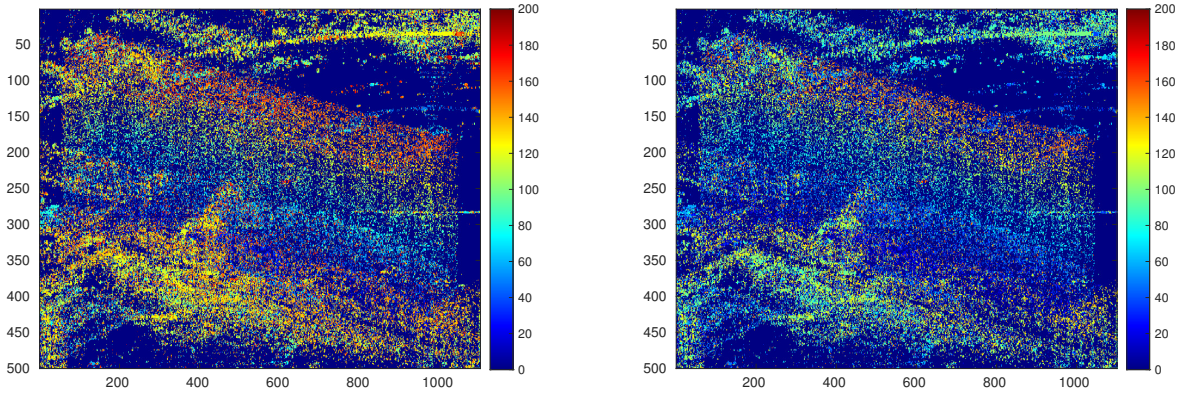
5 Conclusion

In this paper, we introduced a complex-valued neural network CV-LISTA for SAR tomography using small stacks. The proposed CV-LISTA is constructed by unrolling ISTA as a feed-forward neural network with side-connections and replacing the conventional soft-thresholding function with the piecewise linear function Realistic simulations demonstrate that the proposed CV-LISTA outperforms the state of the art by a fair margin, when only limited acquisitions are available. For instance, the proposed algorithm outperforms SL1MMER by about 20% effective detection rate in moderate super-resolving cases at 6dB SNR.



(a) (b)

Figure 5 Reconstructed and color-coded elevation of the test site. (a) detected single scatterer, (b) detected single scatterer + top layer of detected double scatterers.



(a) (b)

Figure 6 Reconstructed and color-coded elevation of double scatterers. (a) top layer of detected double scatterers, (b) bottom layer of detected double scatterers.

At high SNR level, i.e. 10 dB, the proposed algorithm provides 20%-30% higher effective detection rate in extremely super-resolving cases, where the double scatterers are spaced closer than 0.5 ρ_s. Therefore, we see a high potential of the proposed algorithm in large-scale urban TomoSAR processing, where usually only a few acquisitions are available on average for each city.

6 Literature

[1] X. X. Zhu and R. Bamler, "Tomographic SAR Inversion by L_1 -Norm Regularization—The Compressive Sensing Approach," in IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 10, pp. 3839-3846, Oct. 2010, doi: 10.1109/TGRS.2010.2048117.

[2] Kun Qian, Yuanyuan Wang, Yilei Shi and Xiao Xiang, Zhu "Super-resolving SAR Tomography using Deep Learning", In: International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1-5. IEEE. IGARSS 2021, Brussels, Belgium.

[3] Kun Qian, Yuanyuan Wang, Yilei Shi and Xiao Xiang, Zhu "γ-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.

[4] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problemswith a sparsity constraint,” Communications on Pure and Applied Mathematics, vol. 57, no. 11, pp. 1413–1457, 2004.

[5] Karol Gregor and Yann LeCun,“Learning fast approximations of sparse coding,” in Proceedings of the 27th International Conference on International Conference on Machine Learning, Madison, USA, 2010.

[6] X.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

[7] 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.

[8] D. L. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] X. X. Zhu, N. Ge and M. Shahzad, "Joint Sparsity in SAR Tomography for Urban Mapping," in IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 8, pp. 1498-1509, Dec. 2015, doi: 10.1109/JSTSP.2015.2469646.

[10] Y. Shi, X. X. Zhu and R. Bamler, "Nonlocal Compressive Sensing-Based SAR Tomography," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 5, pp. 3015-3024, May 2019, doi: 10.1109/TGRS.2018.2879382.