TomoSAR-ALISTA: Efficient TomoSAR Imaging via Deep Unfolded Network

Muhan Wang[1,2,3,4,5], Zhe Zhang[2,3,5]*, Yue Wang[6], Silin Gao[1,4,5], Xiaolan Qiu[1,2,3,5]

[1]*Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China*
[2]*Key Laboratory of Intelligent Aerospace Big Data Application Technology, Suzhou 215123, China*
[3]*Suzhou Aerospace Information Research Institute, Suzhou 215123, China*
[4]*School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China*
[5]*Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China*
[6]*Electrical and Computer Engineering Department, George Mason University, Fairfax, VA 22030, USA*
*E-mail: zhangzhe01@aircas.ac.cn

Abstract

Synthetic aperture radar (SAR) tomography (TomoSAR) has attracted remarkable interest for its ability in achieving three-dimensional reconstruction along the elevation direction from multiple observations. In recent years, compressed sensing (CS) technique has been introduced into TomoSAR considering for its super-resolution ability with limited samples. Whereas, the CS-based methods suffer from several drawbacks, including weak noise resistance, high computational complexity and complex parameter fine-tuning. Among the different CS algorithms, iterative soft-thresholding algorithm (ISTA) is widely used as a robust reconstruction approach, however, the parameters in the ISTA algorithm are manually chosen, which usually requires a time-consuming fine-tuning process to achieve the best performance. Aiming at efficient TomoSAR imaging, a novel sparse unfolding network named analytic learned ISTA (ALISTA) is proposed towards the TomoSAR imaging problem in this paper, and the key parameters of ISTA are learned from training data via deep learning to avoid complex parameter fine-tuning and significantly relieves the training burden. In addition, experiments verify that it is feasible to use traditional CS algorithms as training labels, which provides a tangible supervised training method to achieve better 3D reconstruction performance even in the absence of labeled data in real applications.

1 Introduction

SAR tomography (TomoSAR) [1], as an enabling RADAR imaging technique, has attracted remarkable interest in recent years for its ability in achieving three-dimensional resolution along the elevation direction from multiple observations. TomoSAR plays a significant role in broad applications, such as urban observation, forestry remote sensing and target identification [2, 3].

Traditionally, TomoSAR problems can be solved via spectrum estimation algorithms, which however usually experience poor performance under limited observations and low SNR circumstances. In typical sceanrios of TomoSAR applications, scatters are usually distributed sparsely along the elevation direction, and meanwhile only a few significant scatterers fall into a range-azimuth pixel. Thus, compressed sensing (CS) [4] based methods are widely used to solve the TomoSAR inversion problem as the state-of-the-art approach.

On the other hand, the CS-based methods suffer from several drawbacks, including weak noise resistance, high computational complexity and complex parameter fine-tuning. To overcome the above issues, different data-driven deep learning methods have been introduced into CS-based signal processing applications [5–7]. As a typical sparse signal processing application, deep learning methods have also been introduced and applied in TomoSAR imaging. For example, Budillon et al. [8] forumulated the TomoSAR inversion problem as a typical classification problem and utilized the neural networks to detect a single scatterer and estimate the corresponding elevation. [9] proposed an efficient line spectral estimation algorithm based on deep neural networks to tackle the TomoSAR inversion, which can distinguish overlaid scatterers and achieve diesirable estimation performance.

Thanks to the enhanced interpretability, fast implementation, and high robustness to model mismatch, network unfolding technique has recently be proposed to develop fast neural network approximation for sparse optimization problems. Specifically, sparse reconstruction algorithms of CS can be represented by an unfolded deep network and solved via canonical training methods over a given dataset [10]. Compared with traditional sparse reconstruction algorithms, the deep unfolded network is usually more robust, converges faster and does not require parameter fine-tuning. Kun Qian et al. developed a sparse unfolding network LISTA-CPSS [11] to solve the TomoSAR problem under the name of γ-net. However, its computational cost is still relatively high, and it relies on a very

large volume of dataset in the training stage, which are both remaining obstacles in practice of TomoSAR.

Aiming at efficient TomoSAR imaging, we focus on a holistic integration of advanced deep unfolding techinques with sparse signal processing algorithm. Along this line, a novel sparse unfolding network, called analytic learned ISTA (ALISTA), has been recently proposed as the unfolded network of the ISTA algorithm [12], which is a popular sparse reconstruction algorithm and has been broadly utilized in TomoSAR problems [13]. In ALISTA, the key parameters of ISTA are learned from training data via deep learning. Furthermore, the convergence speed of ALISTA is much faster than the standard ISTA. In this paper, as the first work of applying ALISTA to the TomoSAR problem, we customize ALISTA design to fit in the TomoSAR imaging model, which can avoid complex parameter fine-tuning and reduce computational complexity. In addition, an effective supervised training method is proposed for the real applications where the ground truth is not known as the training set.

The remainder of this paper is organized as follows. The problem statement of TomoSAR Imaging is first formulated in Section 2. Then, our TomoSAR imaging algorithm based on ALISTA unfolding network is proposed in Section 3. in Section 4 presents numerical results on image reconstruction by OMP, IHT, ISTA, and proposed ALISTA network on synthetic and real data, followed by conclusions in Section 5.

2 TomoSAR Imaging Model

Tomographic SAR technique uses a stack of SAR images collected from different cross-track angles of the same observation area to reconstruct the scattering information along the elevation direction and achieve the three-dimensional resolving ability. The basic principle of SAR tomography is illustrated in Fig. 1.
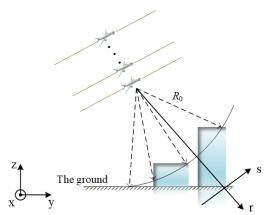

Fig. 1. TomoSAR imaging model diagram.

Assume that the same target is observed N times from different viewing angles to obtain N single look complex SAR images (SLCs). The values of the same pixel among N SLCs yields a sequence . If the SLCs are perfectly aligned, each pixel value can be expressed as the integration of the upward scattering rate distribution along the elevation weighted by sinusoids as (1).

$$y_n = \int \gamma(s) \exp\left(-j2\pi\xi_n s\right) \mathrm{d}s \quad (n = 1, 2, ..., N) \tag{1}$$

where $\xi_n = 2b_n/\lambda r$ is the spatial frequency, b_n is the baseline length the $n-$th channel and λ, r is the wavelength and range. $\gamma(s)$ is the reflectivity function along the elevation. After discretizing the continuous elevation and considering the additive noise, Eq. (1) can be written in a matrix-vector form as (2).

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

where $\mathbf{R}$ plays as a $N \times L$ sensing matrix

$$R_{nl} = \exp\left(-j2\pi\xi_n s_l\right) \tag{3}$$

Now, TomoSAR imaging boils dwown to a signal recovery problem, where the goal is to obtain the reflectivity profile $\boldsymbol{\gamma}$ of each range-azimuth cell by solving (2).

3 TomoSAR via ALISTA

3.1 ISTA and LISTA

TomoSAR imaging usually becomes a sparse reconstruction problem [14], and can be solved via sparse reconstruction algorithms of CS such as OMP [15], IHT [16]and ISTA [17].

Among the different CS algorithms, Iterative soft-thresholding algorithm or ISTA is a widely used option, thanks to its robustness against noise and good reconstruction performance. In ISTA, the estimate of $\hat{\boldsymbol{\gamma}}$ is achieved under an iterative manner as

$$\boldsymbol{\gamma}_{k+1} = h_{(\alpha/L)}\left(\boldsymbol{\gamma}_k + \frac{1}{L}\mathbf{R}^{\mathrm{H}}(\boldsymbol{y} - \mathbf{R}\boldsymbol{\gamma}_k)\right) \tag{4}$$

where L is a parameter controlling the iteration step size and $L > \lambda_{\max}(\mathbf{R}^{\mathrm{H}}\mathbf{R})$, $\lambda_{\max}(\cdot)$ is an operation to calculate the maximum eigenvalue, $\theta = \alpha/L$ is the threshold parameter, and $h_\theta(\mathbf{X})$ is the soft-thresholding function defined as

$$h_\theta(\mathbf{X}) = \mathrm{sign}(\mathbf{X}) \max(|\mathbf{X}| - \theta, 0) \tag{5}$$

However, the parameters α, L in the ISTA algorithm are manually chosen, which usually requires a time-consuming fine-tuning process to achieve the best performance. Furthermore, these parameters are not adaptive i.e., they are fixed from one scene to another.

As mentioned in [10], to let a neural network layer mimic the operation of signal processing algorithm, (4) can be rewritten in a way of neuron's activities:

$$\hat{\boldsymbol{\gamma}_{k+1}} = h_{\theta_{k+1}}(\mathbf{W}_1^{k+1}\boldsymbol{y} + \mathbf{W}_2^{k+1}\hat{\boldsymbol{\gamma}_k}) \qquad (6)$$

where $\mathbf{W}_1^{k+1} = \frac{1}{L}\mathbf{R}^{\mathrm{H}}$ and $\mathbf{W}_2^{k+1} = 1 - \frac{1}{L}\mathbf{R}^{\mathrm{H}}\mathbf{R}$.

If we treat the soft-thresholding function $h_\theta(\cdot)$ as the activation function, then (7) is in the same form as the k+1th layer of a recurrent neural network (RNN). Inspired by the connection between ISTA and RNN, we can unfold the RNN into a deep network to leverage the benefits of deep learning, namely Learned ISTA (LISTA) as proposed in [10]. LISTA is a sparse unfolded network that uses the learnability of deep learning to train the parameters from data: matrix $\mathbf{W}_1^k, \mathbf{W}_2^k$ and threshold θ_k, where k represents the number of layers in the deep network.

3.2 ALISTA Formulation for TomoSAR

Although LISTA enables us to learn parameters from data autonomously, the weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are usually huge in real SAR signal processing applications. Hence, we need a vast dataset to train the enormous amount of parameters and the training stage is temporally and spatially complex.

To solve such challenges, we propose to customize Analytic LISTA (ALISTA) for TomoSAR, where both weight matrices are computed as the solution of a data-free convex optimization problem, while only the step size and threshold parameters are determined by data-driven learning. It significantly simplifies the training stage in both spatial and temporal domains. Here, the data-free optimization of the weight matrices is conducted based on coherence minimization. [12] indicates that ALISTA retains the benefits of LISTA in terms of optimal linear convergence rate and acheives a performance comparable to LISTA.

Instead of training $\mathbf{W}_1, \mathbf{W}_2$, in ALISTA $\mathbf{W}$ is pre-determined by solving the following convex optimization problem:

$$\mathbf{W} = \underset{\mathbf{W} \in \mathbb{C}^{M \times N}}{\arg\min} \|\mathbf{W}_T\mathbf{R}\|_F^2 \;\; \mathrm{s.t.}(\mathbf{W}_{:,i}\mathbf{R}_{:,i}, i = 1, ..., n) \quad (7)$$

which plays as a mutual coherence minimizer between $\mathbf{W}$ and $\mathbf{R}$. This is motivated by the tenet in CS that a dictionary with smaller coherence possesses better sparse recovery performance.

Then, let $\mathbf{W}^k = \mu^k\mathbf{W}$, (7) can be rewritten as the following:

$$\hat{\boldsymbol{\gamma}}_{k+1} = h_{\theta_k}\{\hat{\boldsymbol{\gamma}}_k + \eta_k\mathbf{W}^T(\boldsymbol{y} - \mathbf{R}\hat{\boldsymbol{\gamma}}_k)\} \qquad (8)$$

where only two scalar parameters $\{\theta_k, \eta_k\} \in \mathbb{R}$ are learned from end to end.

Thus, we can significantly reduce the number of training parameters (from LISTA: $\mathcal{O}(KM^2 + K + MN)$ down to ALISTA: $\mathcal{O}(K)$) and system complexity, which reduce the training burden and required training dataset size. The basic workflow of the proposed algorithm is shown in Algorithm 1.

Algorithm 1 TomoSAR-ALISTA

Obtain training data;
   Generate/Acquire reflectivity profile $\boldsymbol{\gamma}$ as labels;
   Acquire TomoSAR observation signal $\boldsymbol{y}$;
   Finish The generation of training data $\{(\boldsymbol{y}_i, \boldsymbol{\gamma}_i)\}_{i=1}^{M \times N}$

Training of ALISTA
   Generate mapping matrix $\mathbf{R}$ with

      $\mathbf{R}_{nl} = \exp(-\frac{j4\pi b_n s_l}{\lambda r})$

   Precomputing $\mathbf{W}$ by solving a convex optimization problem

      $\mathbf{W} = \underset{\mathbf{W} \in \mathbb{C}^{M \times N}}{\arg\min} \|\mathbf{W}_T\mathbf{R}\|_F^2 \;\; \mathrm{s.t.}(\mathbf{W}_{:,i}\mathbf{R}_{:,i}, i = 1, ..., n)$

   The loss function of ALISTA over the training data $\{(\boldsymbol{y}_i, \boldsymbol{\gamma}_i)\}_{i=1}^{M \times N}$ is defined as the mean square error loss:

      $\underset{\boldsymbol{\Psi}}{\min} L(\boldsymbol{\Psi}) = \frac{1}{MN} \sum_{i=1}^{MN} \|\hat{\gamma}(\boldsymbol{\Psi}, \mathbf{y}) - \boldsymbol{\gamma}\|_2^2$

      $\hat{\boldsymbol{\gamma}}_{k+1} = h_{\theta_k}\{\hat{\boldsymbol{\gamma}}_k + \eta_k\mathbf{W}^T(\boldsymbol{y} - \mathbf{R}\hat{\boldsymbol{\gamma}}_k)$

      where $\boldsymbol{\Psi} = [\theta, \eta]$

   Complete the training, obtain the trained model and optimization parameters

---

4 Experiment results

4.1 Simulations

A three-dimensional building structure is generated as the simulation target scene shown in Fig. 2. And the simulation settings are given in Table 1.
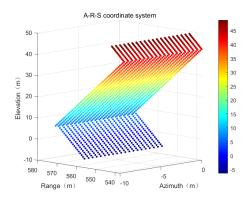

Fig. 2. Simulated 3D building

We generate the training dataset in two ways. Method 1 uses the ground-truth value in simulation as the annotated label. Method 2 is to select some scattering points with good recovery accuracy and high SNR condition after reconstruction by traditional CS methods as data labels, considering that we usually lack ground truth in real applications. The training was carried out based on Python 3.7 using Tensorflow 1.12.0. In the training procedure, we gradually increase the number of layers from 1 to 15 to determine an optimal network structure. Fig. 3 illustrates the performance of ALISTA with the different number

3

Table 1 Simulation parameters

Parameters	Values
Number of channels	8
Wavelength (m)	0.003125
Carrier frequency (GHz)	5.5
Baseline interval (m)	0.1
Average look angle (^{o})	45
Distance of adjacent array elements (m)	0.1
Slant range of the middle trajectory (m)	400
SNR (dB)	20

of layers. It can be seen that the NMSE first decreases rapidly, and the benefit becomes marginal as the layer number reaches 10. The increment of layer number leads to a heavier computational burden. Therefore, ALISTA network employed in this article contains just ten layers.
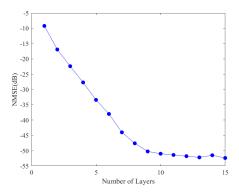

Fig. 3 The performance of ALISTA w.r.t the number of layers

The training result in terms of three-dimensional point clouds is given in Fig. 4.5. Reconstruction results of canonical OMP and ISTA algorithms are also provided as the benchmark. The trained step-size parameter η was 0.02696917, and the threshold parameter θ was 0.01076145.

We can see that both versions of ALSITA reconstruction results significantly outperform all three CS-based methods. Furthermore, the ALSITA result trained from the IHT label has a very similar visual effect as that trained from ground truth, revealing the possibility of utilizing IHT to generate annotated labels in real applications when we lack the ground truth. We also use NMSE to evaluate the accuracy of reconstruction results, which is shown in Table 2. The running time of OMP, IHT, traditional ISTA and ALISTA methods are also given in Table 3.

Table 2 NMSE of different algorithms

Algorithm	OMP	IHT	ISTA	ALISTA-GT	ALISTA-IHT
NMSE (-dB)	8.9594	17.4396	14.2579	68.9500	50.6792

Table 3 The running time of different algorithms

Algorithm	OMP	IHT	ISTA	ALISTA
Time cost (s)	0.25819	0.24026	1.57813	0.26302

(a)

(b)

(c)
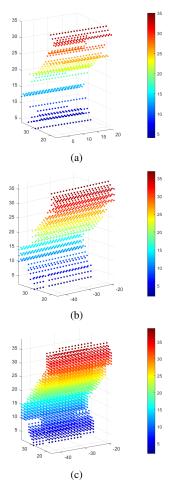
Fig. 4 Traditional results of 3D simulated building reconstruction:(a) OMP with sparsity K=2, (b) IHT with sparsity K=3 (c) traditional ISTA.
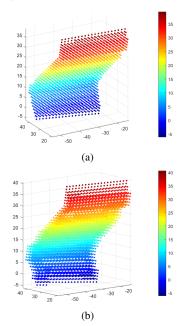

(a)

(b)

Fig. 5 TomoSAR-ALISTA results of 3D simulated building reconstruction:(a) TomoSAR-ALISTA trained from the ground truth label, (b) TomoSAR-ALISTA trained from the IHT label.

4.2 Real Data Experiment

This section adopted real SAR data from the SARMV3D1.0 dataset [18]. It is an airborne array interferometric SAR system and the data is obtained from an urban community in Wan-rong County, Yuncheng city, Shanxi Province, by Aerospace Information Research Institute, Chinese Academy of Sciences (AIRCAS). The array InSAR system has 8 channels. Optical and SAR images of the imaging area are shown in Fig. 6, in which the imaging targets are residential buildings with 14 floors. The height of each floor is around 3.5m, so the height of the building is approximately 52.5 meters.
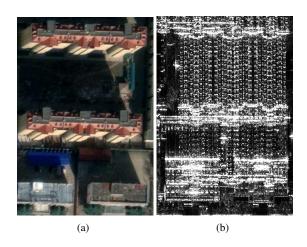

(a)   (b)

Fig. 6 Real data of Yuncheng area. (a) Optical image visualized in Google-Earth, (b) 2D SAR image.

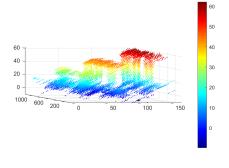### 4.2.1 Experiment setup: parameters and training label

TomoSAR experiment and system configuration is briefly enumerated in Table 4.
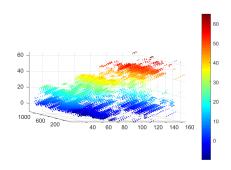
Table 4 Real data parameters

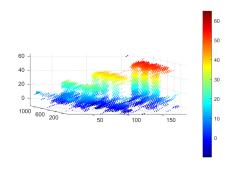| Parameters	Values
Number of channels	8
Wavelength (m)	0.02105
Carrier frequency (GHz)	14.25
Bandwidth (MHz)	500
Baseline interval (m)	0.0832
Average look angle (^{o})	45
Distance of adjacent array elements (m)	0.1
Slant range of the middle trajectory (m)	959.0361
Waveband	Ku

We currently lack precise ground truth of the imaging area at hand. Based on the simulation observations, it is feasible to select some imaging results of conventional CS methods with good reconstruction as the labeled data. Therefore, we adopt some carefully selected IHT reconstruction results as the training set.

4.2.2 TomoSAR 3D Image Reconstruction Results: result and analysis

(a)

(b)

(c)

Fig. 7 Reconstructed and color-coded elevation of real data: (a) IHT, (b) IST, (c) Proposed ALISTA

Traditional CS algorithms and the proposed ALISTA network are utilized, and the 3D cloud points are given as follows:

Processing with MATLAB R2018b and CPU intel i7-1165G7, the time cost of 3D image reconstruction is illustrated in Table 5.

Table 5 The running time of ISTA and propoed ALISTA

Algorithm	ISTA	ALISTA
Time cost (s)	2918.950596	444.421808

Consistent with the simulated results on synthetic data, the reconstruction results of the proposed ALISTA Network are visually better than that of traditional CS algorithms on real data. Under the same experimental conditions and algorithm parameters, the time cost of traditional ISTA is approximately 6.6 times slower than the proposed ALISTA. Height of the

reconstruction results is very close to 52.5m, which is consistent with the ground truth.

In summary, based on 3D image reconstruction results and analysis of simulated building and real data, we can conclude that the proposed method has higher reconstruction accuracy, better convergence rate, and higher algorithm efficiency than traditional CS algorithms under the same experimental conditions in TomoSAR applications.

5 Conclusion

In this paper, an efficient TomoSAR imaging method based on sparse unfolded network namely ALISTA is proposed. Compared with LISTA, ALISTA network reduces the parameter complexity of training from $\mathcal{O}(KM^2 + K + MN)$ down to $\mathcal{O}(K)$, which significantly relieves the training burden in practice. At the same time, we also compare the network training effects by using different schemes in training set labels. The experiments show that it is feasible to use reconstruction results of traditional CS algorithms as training labels, which provides a tangible supervised training method for achieving better 3D reconstruction performance even in the absence of labelled data in real applications.

The future work includes: further improved ALISTA network; introducing the adaptive threshold method; exploring more domain knowledge in the iterative process of threshold shrinkage to achieve faster convergence and higher reconstruction accuracy; and using LiDAR/optical data as the training ground truth.

6 Acknowledgements

References

[1] A. Moreira, P. Prats-Iraola, M. Younis, et al.: 'A tutorial on synthetic aperture radar', IEEE Geoscience and Remote Sensing Magazine, 2013, 1. (1), pp. 6-43

[2] C. Ding, X. Qiu, F. Xu, et al.: 'Synthetic aperture radar three-dimensional imaging—From TomoSAR and array InSAR to microwave vision', J. Radars, 2019, vol. 8, no. 6, pp. 693–709

[3] Z. Jiao, C. Ding, X. Qiu, L, et al.: 'Urban 3D imaging using airborne TomoSAR: Contextual informationbased approach in the statistical way', ISPRS J. Photogramm. Remote Sens., 2020, vol. 170, pp. 127–141

[4] D. L. Donoho,: 'Compressed sensing',' IEEE Transactions on Information Theory, 2006, 52, (4), pp. 1289-1306

[5] M. Wang et al.: 'RMIST-net: Joint range migration and sparse reconstruction network for 3-D mmW imaging', IEEE Trans. Geosci. Remote Sens.,2022, vol. 60, pp. 1–17

[6] 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., 2020, vol. 13, pp. 4476–4492

[7] R. Fu, Y. Liu, T. Huang, et al.: 'Structured LISTA for multidimensional harmonic retrieval', IEEE Trans. Signal Process., 2021, vol. 69, pp. 3459–3472

[8] A. Budillon, A. C. Johnsy, G. Schirinzi, et al.: 'SAR tomography based on deep learning', Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 2019, 7. pp. 3625–3628

[9] J. Gao, Y. Ye, S. Li, et al.: 'Fast super-resolution 3D SAR imaging using an unfolded deep network', Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP), 2019, 12. pp. 1–5

[10] X. Chen, J. Liu, Z. Wang, et al.: 'Theoretical linear convergence of unfolded ista and its practical weights and thresholds', Proc. 32nd Int. Conf. Neural Inf. Process. Syst. Red Hook, NY, USA: Curran Associates, 2018, pp. 9079–9089

[11] K. Qian, Y. Wang, Y. Shi, et al.: 'γ-Net: Superresolving SAR Tomographic Inversion via Deep Learning', IEEE Transactions on Geoscience and Remote Sensing, 2022, vol. 60, pp. 1-16

[12] J. Liu, X. Chen, Z. Wang, et al.: 'ALISTA: Analytic Weights Are As Good As Learned Weights in LISTA', International Conference on Learning Representations, 2019.

[13] D. Han, L. Zhou, Z. Jiao, et al.: 'Efficient 3D Image Reconstruction of Airborne TomoSAR Based on Back Projection and Improved Adaptive ISTA', IEEE Access, 2021, vol. 9, pp. 47399-47410

[14] 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, 2010, 48, (10), pp. 3839-3846

[15] J. A. Tropp and A. C. Gilbert.: 'Signal recovery from random measurements via orthogonal matching pursuit', IEEE Trans. Inf. Theory, 2007, 12. vol. 53, no. 12, pp. 4655–4666

[16] J. D. Blanchard, M. Cermak, D. Hanle, et al.: 'Greedy Algorithms for Joint Sparse Recovery', IEEE Transactions on Signal Processing, 2014, vol. 62, no. 7, pp. 1694-1704

[17] J. Zhang and B. Ghanem.: 'ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing' 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 1828-1837

[18] X. Qiu, Z. Jiao, L. Peng, et al.: 'SARMV3D-1.0: Synthetic aperture radar microwave vision 3D imaging dataset', Journal of Radars, 2021, 10, (4), pp. 485-498