COMPLEX-VALUED SPARSE LONG SHORT-TERM MEMORY UNIT WITH APPLICATION TO SUPER-RESOLVING SAR TOMOGRAPHY

Kun Qian[1], Yuanyuan Wang[1,2], Peter Jung[3], Yilei Shi[1] and Xiao Xiang Zhu[1,2]

(1) Data Science in Earth Observation, Technical University of Munich, Munich, Germany
(2) Remote Sensing Technology Institute, Germany Aerospace Center, Wessling, Germany
(3) Communications and Information Theory Chair, Technical University of Berlin, Berlin, Germany

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883246

ABSTRACT

To achieve super-resolution synthetic aperture radar (SAR) tomography (TomoSAR), compressive sensing (CS)-based algorithms are usually employed, which are, however, computationally expensive, and thus is not often applied in large-scale processing. Recently, deep unfolding techniques have provided a good combination of physical model-based algorithms and the ability of neural networks to learn from data. In this vein, iterative CS-based algorithms can usually be unrolled as neural networks with only 10 to 20 layers. When trained, it shows great computational efficiency for further TomoSAR processing. However, the learning architecture of neural networks built in this approach tends to result in error propagation and information loss, thus degrading the performance. In this paper, we propose to employ complex-valued sparse long short-term memory (CV-SLSTM) units to tackle this problem by incorporating historically updating information into the optimization procedure and preserving full information. Simulations are carried out to validate the performance of the proposed algorithm.

Index Terms— SAR tomography, Super-resolution, Complex-valued neural network, deep learning

1. INTRODUCTION

1.1. TomoSAR imaging model

SAR tomography (TomoSAR) is known as a powerful technique for 3-D reconstruction in dense urban areas due to its strong capability of separating overlaid scatterers in the same resolution unit. In the presence of noise $\boldsymbol{\epsilon}$, the discrete TomoSAR imaing model can be expressed as follows:

$$\mathbf{g} = \mathbf{R}\gamma + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{g} \in \mathbb{C}^{N\times 1}$ is the complex-valued 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 $R_{nl} = \exp(-j2\pi\xi_n s_l)$. $\xi_n = -2b_n/(\lambda r)$ denotes the elevation frequency.

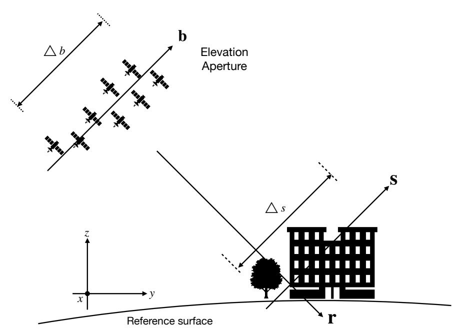

Fig. 1: The SAR imaging geometry. The elevation synthetic aperture is built up by SAR data acquired from slightly different viewing angles. Flight direction is orthogonal into the plane.

1.2. CS-based TomoSAR inversion

It was investigated in [1] that usually only a few scatterers (0-4) are overlaid along the elevation direction in each resolution unit in dense urban areas, meaning that the reflectivity profile should be sparse in the space domain. Therefore, one usually resorts to compressive sensing [1] [2] (CS)-based sparse reconstruction methods to distinguish multiple scatterers overlaid in an individual resolution unit along the elevation direction. With CS-based methods, the reflectivity profile $\hat{\boldsymbol{\gamma}}$ can be estimated by:

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

CS-based methods are known to have unprecedented super-resolution ability and estimation accuracy. However, CS-based methods are extremely computationally expensive due to costly L_2-L_1 mix norm minimization using typical basis pursuit denoising and are hard to be applied to practical large-scale TomoSAR processing.

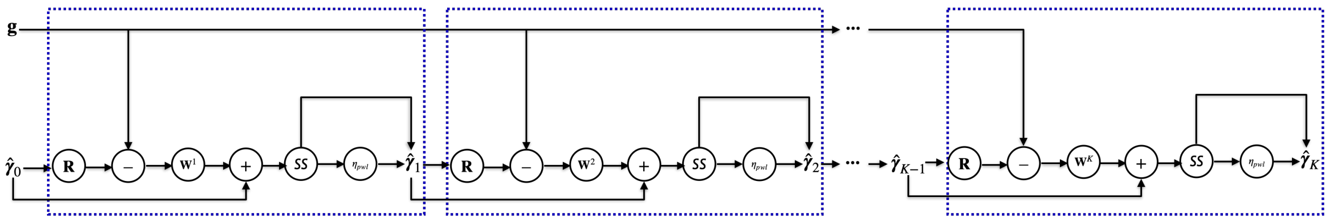

Fig. 2: Illustration the learning architecture of a K-layer γ-Net.

1.3. Deep learning for TomoSAR inversion

Recently, an emerging technique coined ***deep unfolding*** [3] was proposed to provide a concrete and systematic connection between iterative physical model based algorithms and deep neural networks, especially for sparse coding. Inspired by this, the TomoSAR community started to investigate the potential of deep unfolding techniques for TomoSAR inversion [4]. The author proposed γ-Net in [4] by unrolling iterative shrinkage thresholding algorithm (ISTA) as a complex-valued recurrent neural network and integrating weight coupling [5] and support selection [5] into the network structure. Fig. 2 illustrates us the learning architecture of γ-Net. It was demonstrated in [4] that γ-Net has competitive performance w.r.t super-resolution power and estimation accuracy with the state-of-the-art CS-based TomoSAR method *SL1MMER*. However, γ-Net still suffers from error propagation and information loss since it inherits the learning architecture of the learned ISTA (LISTA) despite modifications made by the authors to accommodate to TomoSAR inversion. To be specific, the errors in the first few layers will be propagated and further amplified in the upcoming layers because the output in the current layer is generated exclusively on the previous output. More seriously, once useful information is discarded in the previous layers, it is no longer possible for the upcoming layers to recover and utilize the discarded information. As a consequence, large estimation errors can be expected.

1.4. Contribution of this paper

To address this issue, we proposed complex-valued sparse long short-term memory (CV-SLSTM) units in this paper by introducing two gated units to preserve full information. With the assistance of the CV-SLSTM units, the long-term dependence of the previous outputs can be captured and maintained. Simultaneously, the CV-SLSTM unit automatically accumulates important information and forgets redundant information in the dynamics of the network.

2. COMPLEX-VALUED SPARSE LONG SHORT-TERM MEMORY UNIT

2.1. Sparse long short-term memory unit

In the optimization community, it has been extensively studied and proved [6, 7, 8] that incorporating historically updating information contributes to improving the algorithm performance. Inspired by the previous researches about incorporating historically updating information, the author proposed sparse long short-term memory (SLSTM) unit in [9] to integrate and make use of historically updating information to preserve full information by introducing forget and input gates, termed as $\mathbf{f}$ and $\mathbf{i}$, respectively, into each layer of LISTA.

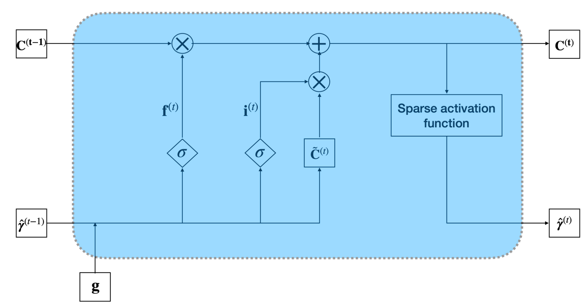

Fig. 3: SLSTM unit. Each unit refers to a layer of the proposed deep RNN.

Fig. 3 illustrates us the learning structure of an individual SLSTM unit. To clarify, SLSTM unit does not have the output gate like conventional LSTM units. Specific formal definition of a SLSTM unit is expressed as follows:

$$
\begin{aligned}
\mathbf{i}^{(t)} &= \sigma\left(\mathbf{W}_{i2}^{(t)}\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_{i1}^{(t)}\mathbf{g}\right) \\
\mathbf{f}^{(t)} &= \sigma\left(\mathbf{W}_{f2}^{(t)}\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_{f1}^{(t)}\mathbf{g}\right) \\
\tilde{\mathbf{C}}^{(t)} &= \mathbf{W}_2\hat{\boldsymbol{\gamma}}^{(t-1)} + \mathbf{W}_1\mathbf{g} \\
\mathbf{C}^{(t)} &= \mathbf{f}^{(t)} \odot \mathbf{C}^{(t-1)} + \mathbf{i}^{(t)} \odot \tilde{\mathbf{C}}^{(t)} \\
\hat{\boldsymbol{\gamma}}^{(t)} &= \eta_{dt}\left(\mathbf{C}^{(t)}\right)
\end{aligned} \tag{3}
$$

where $\mathbf{W}_*$ denotes the weight matrices to be learned in each SLSTM unit. It is worth mentioning that the weight matrices

$\mathbf{W}_1$ and $\mathbf{W}_2$ are shared for all SLSTM units in a network. $\sigma(\cdot)$ indicates the conventional sigmoid function. $\eta_{dt}(\cdot)$ is the double hyperbolic tangent function and acts as the sparse activation function employed in the SLSTM to promote sparse codes. It is defined as follows:

$$\eta_{dt}(\hat{\gamma}) = s \cdot [\tanh(\hat{\gamma} + \theta) + \tanh(\hat{\gamma} - \theta)] \quad (4)$$

where s and θ denote two trainable parameter. It is worth noting that the double hyperbolic tangent function can be viewed as a smooth and continuously differentiable alternative of the conventional soft-thresholding function.

Its advantages are mainly two-fold. On the one hand, its second derivative sustains for a long span so that it is able to effectively address the gradient vanishing problem caused by the cell recurrent connection. On the other hand, it is able to effectively imitate the soft-thresholding function within the interval of $[-\theta, \theta]$. Fig. 4 demonstrates an example of the double hyperbolic tangent function and compares it to the soft-thresholding function.

The RNN built with SLSTM units is termed as ***sc2net*** [9]. 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 from the previous outputs can be captured and maintained. Second, important information will be automatically accumulated, whereas useless or redundant information will be blocked, in the dynamics of ***sc2net***.

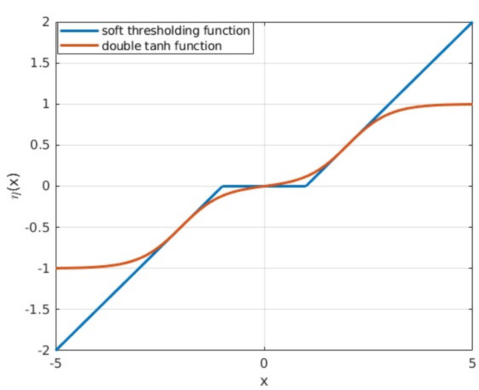

Fig. 4: comparison between double tanh and soft-thresholding

2.2. Complex-valued SLSTM unit

To apply SLSTM units to solve TomoSAR inversion, we need to further extend it to a complex domain since SAR data is complex-valued. The Complex-valued SLSTM (CV-SLSTM) unit has essentially the same structure as the SLSTM unit despite two differences. First, each neuron in the CV-SLSTM unit has two channels indicating the real and imaginary parts of a complex number, respectively. In addition, the complex number cannot be directly activated. Usually, we need to perform the activation on the magnitude of the complex number. Hence, it is no longer appropriate to use the sigmoid function for activation to generate the forget and input gates since the magnitude is always larger than zero and the forget and input gates will always be larger than 0.5 after activation, which is not reasonable and seriously affects the performance. To tackle this problem, we employed the "$\tanh$" function instead of the sigmoid function to guarantee that the value of the forget and input gates varies from 0 to 1 after activation, as it was originally designed. By applying the aforementioned adaptions to Eq. (3), we have the formulation of the CV-SLSTM unit as follows:

$$\begin{aligned}
\mathbf{i}^{(t)} &= \tanh\left(|\mathbf{W}_{i2}^{(t)}\hat{\gamma}^{(t-1)} + \mathbf{W}_{i1}^{(t)}\mathbf{g}|\right) \\
\mathbf{f}^{(t)} &= \tanh\left(|\mathbf{W}_{f2}^{(t)}\hat{\gamma}^{(t-1)} + \mathbf{W}_{f1}^{(t)}\mathbf{g}|\right) \\
\tilde{\mathbf{C}}^{(t)} &= \mathbf{W}_2\hat{\gamma}^{(t-1)} + \mathbf{W}_1\mathbf{g} \\
\mathbf{C}^{(t)} &= \mathbf{f}^{(t)} \odot \mathbf{C}^{(t-1)} + \mathbf{i}^{(t)} \odot \tilde{\mathbf{C}}^{(t)} \\
\hat{\gamma}^{(t)} &= \eta_{cv-dt}\left(\mathbf{C}^{(t)}\right)
\end{aligned} \quad (5)$$

where $\eta_{cv-dt}(\cdot)$ is the complex-valued version of the double hyperbolic function and expressed as follows:

$$\eta_{cv-dt}(\hat{\gamma}) = s \cdot e^{j \cdot \angle(\hat{\gamma})}[\tanh(|\hat{\gamma}| + \theta) + \tanh(|\hat{\gamma}| - \theta)], \quad (6)$$

with j being the imaginary number.

3. EXPERIMENTS

The performance of the proposed algortihm is validated using simulations. We simulated an interferometric stack containing 25 regularly distributed baseline ranging from -135m to 135m. The elevation aperture size of 270m results in a Rayleigh resolution ρ_s of about 42m. Overlaid double scatterers were simulated in each resolution unit. The double scatterers were set to have identical amplitude and phase, which is the worst case in TomoSAR processing.

A well-known TomoSAR benchmark test [1] [10] was carried out to evaluate the super-resolution power as well as the estimation accuracy of the proposed algorithm. In the experiment, we simulated double scatterers with increasing elevation distance between the two layovered scatterers, in order to mimic a facade-ground interaction. Two different scenarios with SNR$\in \{0, 6\}$ dB were taken into consideration, which represent typical SNR levels in a high-resolution spaceborne SAR image.

We used the ***effective detection rate*** P_d to evaluate the performance. An effective detection should simultaneously 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 the ground truth;

3. both elevation estimates are also within $\pm 0.5\ d_s$ w.r.t. the ground truth.

d_s is the elevation distance between the double scatterers. The third criterion is seldom seen in the literature. However, it is necessary, because in extremely super-resolving cases, 3 times CRLB is a constraint not sufficient to guarantee reasonable estimates. $\pm 0.5\ d_s$ is a much stricter constraint in such cases, which will reflect the true performance of the algorithm. With the effective detection rate, we can simultaneously evaluate the super-resolution power and the estimation accuracy since an effective detection not only means the successful detection of double scatterers, but also guarantee that the elevation estimate has low bias in the meantime.

Fig. 5 compares the effective detection rate P_d of the proposed algorithm and γ-Net [4]. The effective detection rate P_d is plotted as a function of the normalized distance α, which is defined as the ratio between the elevation distance between the double scatterers d_s and the Rayleigh resolution ρ_s. As can be seen in Fig. 5, the proposed algorithm achieves a higher effective detection rate. To be specific, at 6dB SNR, the proposed algorithm offers about 10% to 20% higher effective detection rate than γ-Net in the super-resolving regime, i.e. the double scatterers are spaced closer than 0.6 Rayleigh resolution. In the low SNR case, the effective detection rate of the proposed algorithm is, on average, about 20% higher than that of γ-Net.

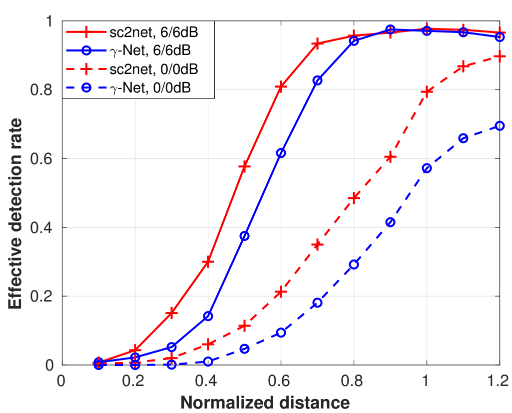

Fig. 5: Effective detection rate as a function of the normalized elevation distance between the simulated facade and ground using the proposed algorithm and γ-Net with SNR = 0dB and 6dB under 0.2 million Monte Carlo trials.

4. CONCLUSION

In this paper, we introduced the SLSTM unit and extended it to the CV-SLSTM unit to solve SAR tomography. By introducing two gated units to each LISTA layer, the CV-SLSTM unit is able to preserve full information and overcome the problem of error propagation and information loss caused by the learning architecture of LISTA. Realistic simulation demonstrates that a RNN built with CV-SLSTM units is able to deliver significantly better super-resolution power in TomoSAR than the state of the art. The encouraging result opens up a new prospect for SAR tomography using deep learning and motivates us to further investigate the potential of the RNN with gated units in practical TomoSAR processing.

5. REFERENCES

[1] Xiaoxiang Zhu and Richard Bamler, “Tomographic sar inversion by l1-norm regularization – the compressive sensing approach,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, 2010.

[2] David Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] John Hershey, Jonathan Le Roux, and Felix Weninger, “Deep unfolding: Model-based inspiration of novel deep architectures,” *Computer Science*, 2014.

[4] Kun Qian, Yuanyuan Wang, Yilei Shi, and Xiao Xiang Zhu, “γ-net: Superresolving sar tomographic inversion via deep learning,” *arXiv preprint, arXiv:2112.04211*, 2021.

[5] Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin, “Theoretical linear convergence of unfolded ista and its practical weights and thresholds,” *arXiv preprint arXiv:1808.10038*, 2018.

[6] Ning Qian, “On the momentum term in gradient descent learning algorithms,” *Neural Networks*, vol. 12, no. 1, pp. 145–151, 1999.

[7] Matthew Zeiler, “Adadelta: An adaptive learning rate method,” *arXiv preprint, arXiv:1212.5701*, 2012.

[8] John Duchi, Elad Hazan, and Yoram Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” *J. Mach. Learn. Res.*, vol. 12, no. null, pp. 21212159, jul 2011.

[9] Joey Tianyi Zhou, Kai Di, Jiawei Du, Xi Peng, Hao Yang, Sinno Jialin Pan, Ivor W. Tsang, Yong Liu, Zheng Qin, and Rick Siow Mong Goh, “Sc2net: Sparse lstms for sparse coding,” in *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, Feb. 2018, AAAI, pp. 4588–4595.

[10] Xiaoxiang Zhu and Richard 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.