Deep Learning based View Interpolation towards Improved TomoSAR Focusing

Sergio Alejandro Serafín-García, *Graduate Student, IEEE,* Matteo Nannini, Ronny Hänsch, *Senior Member, IEEE,* Gustavo Daniel Martín-del-Campo-Becerra, and Andreas Reigber, *Fellow, IEEE*

***Abstract*—Synthetic Aperture Radar Tomography (TomoSAR) uses several co-registered images from different perspectives to reconstruct a power spectrum pattern perpendicular to the line of sight, enabling the estimation of a 3D representation of the area. Classical estimators exhibit ambiguities and other undesired effects that are stronger for sparser and smaller stacks. To mitigate the limitations arising from a restricted number of acquisitions, we propose using a deep neural network to synthesize artificial tracks (i.e., images not contained in the original stack). The presented method utilizes a convolutional neural network with an encoder-decoder architecture. We evaluate the proposed approach on real TomoSAR data from an airborne campaign over a forest region. The view estimation improves the tomographic results, offering robustness to scenarios affected by temporal decorrelation, which other classical methods, like cubic convolution, do not provide.**

***Index Terms*—Deep Learning (DL), Interpolation, Synthetic Aperture Radar (SAR), Tomography.**

I. Introduction

SYNTHETIC Aperture Radar Tomography (TomoSAR) reconstructs the internal distribution of semi-transparent targets (i.e., the Power Spectrum Pattern (PSP)) by using a sparse collection of co-registered SAR acquisitions, so-called TomoSAR stack. The SAR measurements are taken with different baselines (BL) regarding a primary pass; thus, they offer different perspectives of the area. The spatial diversity of perspectives in the stack allows the synthesis of a resolution in the direction Perpendicular to the Line Of Sight (PLOS), which is usually achieved by the inversion of the stack [1].

The number and spatial distribution of tracks within the stack have two main effects on the inversion. First, the resolution of the retrieved PSP is inversely proportional to the total BL span (called D_{PLOS} in Fig. 1) [1]. Second, ambiguities caused by subsampling are inversely proportional to the variance of BLs, depicted by d in Fig. 1 for regularly spaced acquisitions [1]. Thus, making the TomoSAR stack larger and denser increases the ambiguity rejection and the resolution of the estimated tomograms.

In practice, each image in the tomographic stack is acquired by different flights, making large stacks unpractical and expensive. Also, the temporal offset between tracks increases for each new acquisition, causing temporal decorrelation and limiting the potential size of the stack [2].

In this study, we use a deep Neural Network (NN) to interpolate SAR images, whose BLs are not contained in the original TomoSAR stack. The tomographic stack becomes denser and the focused tomograms achieve alleviated ambiguities and reduced side lobes. This facilitates the identification of different layers within the PSP, e.g., canopy and ground.

The authors are with the Microwave and Radar Institute, German Aerospace Center (DLR), 82234 Wessling, Germany (email: Sergio.SerafinGarcia@dlr.de).

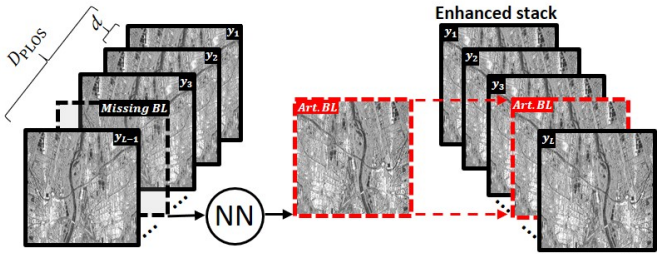

Fig. 1. TomoSAR stack improvement throw DL-based view interpolation.

The use of Deep Learning (DL) in the context of TomoSAR focusing has been explored in [3], [4]. These studies use supplementary data as ground truth (e.g., simulated PSPs and LiDAR measurements) to train supervised DL models. The method presented in [3] demonstrates that lightweight NNs can be trained to perform inversion with a single feed-forward pass, resulting in fast reconstructions that can efficiently scale to future missions handling large volumes of data. However, since beamformed PSPs are employed as input to the NN, biases can be introduced. The input data may present focusing issues, such as mixed targets due to poor resolution or spectral ambiguities, which can affect the performance of the NN. On the other hand, [4] retrieves forest height and underlying topography using single- and dual-polarimetric data. Nonetheless, the usage of LiDAR information to train the NN may be misleading. The penetration of low-frequency SAR signals into the canopy, results in backscattering phase center heights that differ from the ones measured by LiDAR. Moreover, an incorrect handling of the TomoSAR stack potentially affects the accuracy and reliability of the results. Finally, it is imperative to consider that both [3], [4] may be limited due to uncertain generalization, where discrepancies between the supplementary information and the actual vertical profiles (model mismatching) can lead to poor focusing.

The main contributions of this work are:

- Use of a DL model to synthesize views in a TomoSAR stack, aimed at improving the focused tomograms. The interpolated SLCs present robustness againts temporal decorrelation and larger BLs.
- Use of flattened interferograms as input to the DL-model,

which represent the phase information of the SLCs. The interferograms are calibrated and flattened (employing a known digital elevation model) in a pre-processing step.

II. Methodology

We envision a TomoSAR mission over a Region Of Interest (ROI). Part of the ROI is acquired with the whole set of BLs, whereas the remaining area omits some tracks. The excluded passes are synthesized by the DL-based interpolator. To address this approach, we first explain how the TomoSAR stack produces a PSP. Subsequently, we elaborate on the DL-model employed to synthesize the SLCs.

A. TomoSAR signal model

The TomoSAR inverse problem is modeled via [1], [2], $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}$. We consider a geometry with L passes, each of them with a different LOS. For a single azimuth range position, $\mathbf{y}$ represents the L pre-processed observations. The M complex reflectivity values for every position $\{z_m\}_{m=1}^{M}$ in the PLOS direction are stored in vector $\mathbf{s}$. The steering matrix $\mathbf{A}$ denotes in its columns the steering vectors $\{\mathbf{a}(z_m)\}_{m=1}^{M}$ [2]. These vectors represent the interferometric phase information corresponding to the backscattering sources along the PLOS positions $\{z_m\}_{m=1}^{M}$. The co-registered SLC images are combined coherently.

The main objective of TomoSAR is recovering the PSP in PLOS, defined by $\mathbf{b} = \left\{\left\langle \left| s_m \right|^2 \right\rangle\right\}_{m=1}^{M}$. For example, the classical Fourier-based Matched Spatial Filter (MSF) [2] attains $\mathbf{b}_{\text{MSF}} = \mathbf{A}^{+}\mathbf{Y}\mathbf{A}$, where $\mathbf{Y}$ is the sample covariance matrix [2, Eq. 4]. It is noteworthy that our DL-based interpolator works independently of the chosen spectral estimator.

B. Deep learning model

DL models use cascaded non-linear processing units to extract dataset features via example-based learning. We refer to a modified U-net [5]. This architecture consists of two parts, the left part of the "U", called contracting path, and the right part, referred to as expansive path.

The contracting path comprises five encoder blocks, each doubling the number of filters and halving the spatial size. Conversely, the expansive path includes five decoder blocks, each increasing the number of features and halving the spatial dimension. Encoder blocks consist of two 3×3 convolutional layers with padding 1, followed by a ReLU activation and a 2×2 maximum pooling function. Decoder units consist of a transposed convolutional layer with stride 2, concatenation of up-sampled feature maps with those from the contracting path, and two 3×3 convolutional layers with padding 1 and ReLU activation. The concatenation integrates low-level and high-level information. A 1×1 convolutional layer in the expansive path restores the initial sizes and defines the regression model.

Consider a TomoSAR stack $\{\mathbf{SLC}_{[i]}\}_{i=1}^{L}$, $\mathbf{SLC}_{[1]}$ is selected as primary and the remaining $\{\mathbf{SLC}_{[i]}\}_{i=2}^{L}$ as secondaries. The input SLCs are represented with $2(L-1)$ channels: one half being the flattened interferograms [6] of the primary with each secondary $\{\Gamma(\mathbf{SLC}_{[1]}, \mathbf{SLC}_{[i]})\}_{i=2}^{L}$ and the other half attained with $\{log(amp(\mathbf{SLC}_{[i]})/amp(\mathbf{SLC}_{[1]}) + 1)\}_{i=2}^{L}$. The log is used to manage a broad range of values. Then, the model input is defined as $\mathbf{X}^{pl \times pw \times 2(L-1)}$, with $pl \times pw$ patches.

The flattened interferograms represent the phase characteristics of the images. As such, the detection of patterns is facilitated, making the fitting of the model possible. Since no multilooking is considered in the computation of the interferograms, we can calculate back the original SLCs.

The SLC that the network learns to estimate, $\mathbf{SLC}_{target}$, is portrayed by the next two channels: $\Gamma(\mathbf{SLC}_{[1]}, \mathbf{SLC}_{target})$ and $log(amp(\mathbf{SLC}_{target})/amp(\mathbf{SLC}_{[1]}) + 1)$. Then, the output of the network is presented as $f(\mathbf{X}) = \mathbf{W}^{pl \times pw \times 2}$, where $f : \mathbf{X} \rightarrow \mathbf{W}$ represents the NN mapping the input into the output. The loss function $Loss = Loss_{\emptyset} + Loss_{amp}$ is considered during training, where

$$Loss_{\emptyset} = \frac{1}{\text{VAR}(\mathbf{W}_{\emptyset})n} \sum_{i=1}^{n} sub_{ang}(\mathbf{W}_{\emptyset_i}, \widehat{\mathbf{W}}_{\emptyset_i})^2, \quad (1)$$

$$Loss_{amp} = \frac{1}{\text{VAR}(\mathbf{W}_{amp})n} \sum_{i=1}^{n} (\mathbf{W}_{amp_i} - \widehat{\mathbf{W}}_{amp_i})^2. \quad (2)$$

This is a modified version of the mean square error, computed channel-wise (amplitude and phase) between $\mathbf{SLC}_{target}$ and the approximations made by the NN. The function $sub_{ang}(a, b)$ represents the subtraction of angles a and b considering its circular disposition, giving us a realistic distance of the two phases. The quantity of samples is denoted with n. Expressions (1) and (2) are scaled with the variance (VAR) aiming that $Loss_{\emptyset}$ and $Loss_{amp}$ are in the same numeric range.

III. Experiments

As a demonstration of the feasibility of the DL-based SLC interpolator, the next experiment is conducted. The dataset is divided in the azimuth direction into two subsets: one designated for training and the other one for testing. The training subset includes the entire set of L BLs, while the test subset contains data from $L-1$ BLs. This division is intended

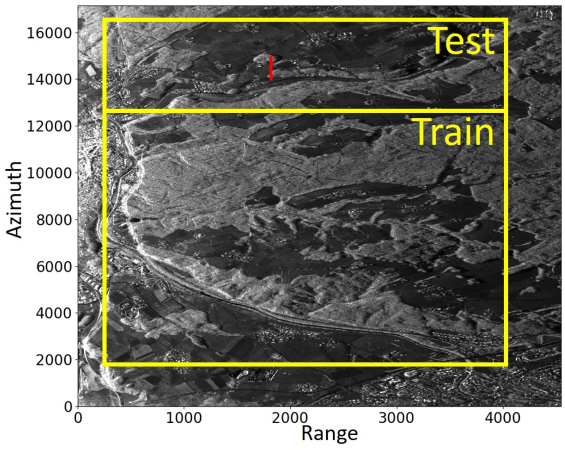

Fig. 2. Test and train subset. The red line in the test subset is the slice where the tomographic experiments are performed.

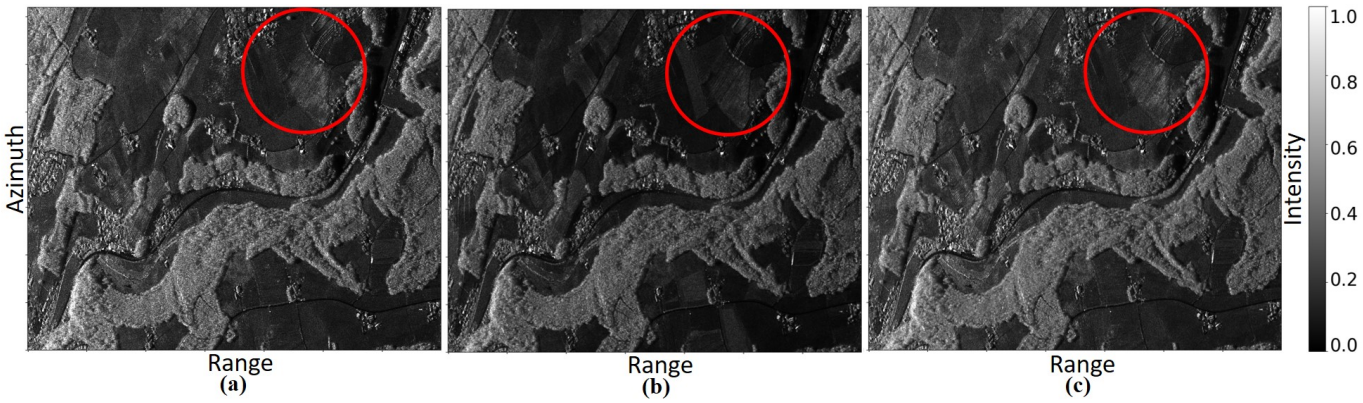

Fig. 3. SLC intensity. **(a)** Original. **(b)** CC interpolated. **(c)** DL-based interpolated. The circles indicate the EA.

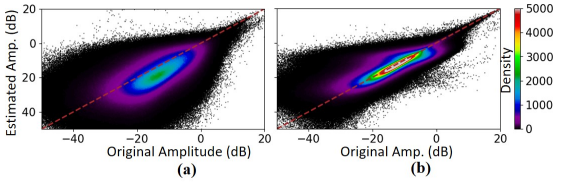

Fig. 4. Scatter plot between the original and estimated amplitudes in dB. **(a)** Original and CC interpolated. **(b)** Original and DL-based interpolated.

to simulate a scenario where the training area is acquired with all the tracks, and the testing area is covered with fewer passes. Subsequently, a DL model is trained using the training subset to generate the missing SLC image in the test subset. To assess the capabilities of the proposed method, results are compared against those obtained using a not DL-based method, i.e., the classical Cubic Convolution (CC) [7]. According to [7], CC is considered one of the most effective SLC interpolators.

A. Specifications

The experiments are conducted using a crop of the F-SAR dataset acquired during the 17SARTOM mission over the Traunstein test site in Germany in 2017. The mission utilized a L-Band sensor with a wavelength of 0.226m, featuring multipolarimetric capabilities. In our case only VV polarization is considered. The sensor operated at a nominal height of 3720m, with a range resolution of 1.3m and an azimuth resolution of 0.6m. The training and test subsets comprise crops of 11400×3500 and 3600×3500 pixels, respectively, as seen in Fig. 2. The following experiments consider 7 secondaries with horizontal BLs at 6.6m, 22.3m, 30.2m, 42.3m, 90.1m, 120.9m and 200.0m, with respect to the primary track at 0m.

The proposed DL-model was fitted with the training subset using 75% of the data for training and 25% for validation. We consider patches with size 96×96. Mini-batches of size 256 patches were employed with an Adam optimizer using a learning rate of $3 \cdot 10^{-4}$ during 110 epochs.

B. Results

Results are addressed in terms of intensity (III-B1), phase (III-B2) and its usability to improve the tomograms (III-B3). Results in subsections III-B1 and III-B2 consider a target BL at 68.5m, while III-B3 consider target BLs at 14.6m, 68.5m, 105.8m, and 158.6m. The Exemplary Areas (EA) in the upcoming results highlight the differences between cases.

1) Intensity estimation assessment: Consider the original SLC in Fig. 3-**a** as a reference. Observe that both CC and DL-based methods (Fig. 3-**b** and Fig. 3-**c**, respectively) managed to successfully recognize most of the patterns that constitute the scene. Yet, the agricultural fields in the EA are missed in Fig. 3-**b** and not in Fig. 3-**c**. Additionally, Fig. 3-**a** and Fig. 3-**c** share the same brightness, while Fig. 3-**b** is darker.

To evaluate the recreation of the intensity attained by the interpolators, Fig. 4 shows a scatter plot of the amplitudes. In Fig. 4-**b** (DL interpolator), the majority of pixel density lies above the diagonal red line. In contrast, Fig. 4-**a** (CC interpolator) shows that most of the density lies beyond the red line and is more widely spread. This indicates that the DL-based method achieves a more accurate recreation of the original data in terms of intensity.

2) Phase estimation assessment: We refer to flattened interferograms between $\mathbf{SLC}_{target}$ and $\mathbf{SLC}_{[1]}$. Consider as a reference the interferogram computed using the original SLCs, shown in Fig. 5-**a**. For the CC interpolation in Fig. 5-**b**, some features in the EA are missed. Contrariwise, the DL-based interpolation is able to retrieve such features (Fig. 5-**c**).

Fig. 6 depicts the interferometric coherence between the original SLC and those interpolated. As seen in the EA of Fig. 6-**a**, it is more evident that the CC interpolation exhibits inferior performance in contrast to the DL-based interpolation in Fig. 6-**b**. The latter shows a coherence very close to 1 in most of the EA. Overall, the coherence achieved by the DL-based interpolated SLC is higher than the one of CC. Note that both methods are not accurate enough for forest areas.

3) Tomographic assessment: Tomograms in Fig. 7 were computed with MSF. Three distinct EAs are highlighted in the tomograms, tagged as Z1, Z2 and Z3. Fig. 7-**a** considers 8 original BLs, as 4 targeted BLs are taken out. The reference tomogram in Fig. 7-**b** is produced with 12 original BLs. Fig. 7-**c** and Fig. 7-**d** consider also 12 BLs, but 4 of them obtained via interpolation, CC and DL-based, respectively. The red dots

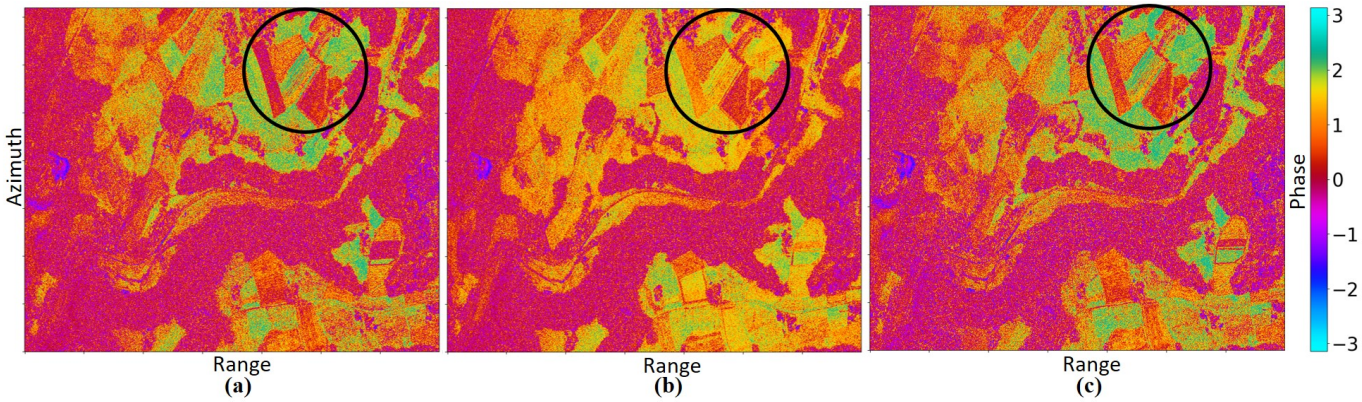

Fig. 5. Flattened interferograms between primary SLC and **(a)** original targeted SLC, **(b)** CC interpolated SLC and **(c)** DL-based interpolated SLC. The circles indicate the EA.

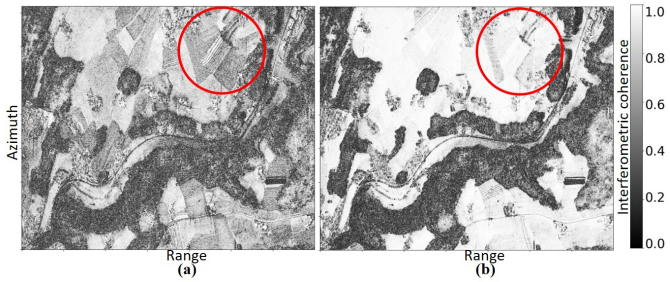

Fig. 6. Interferometric coherence between original and estimated SLCs. **(a)** CC interpolated. **(b)** DL-based interpolated. The circles indicate the EA.

at right hand of each tomogram indicate the interpolated BLs, while the black dots refer to original BLs. On the left, Fig. 7-**e** displays the superimposed vertical profiles from an exemplary azimuth position, each of them referenced by color.

At first glance, the tomogram in Fig. 7-**d** (DL interpolator) and the tomogram in Fig. 7-**b** (full stack) show similar features. The Z1 EA encircles ambiguities, stronger in Fig. 7-**a** and Fig. 7-**c**. For the DL interpolation (Fig. 7-**d**), these are mostly suppressed. These differences are more prominent in the vertical profiles depicted in Fig. 7-**e**. The Z2 EA points out the ground layer, where we can see the side lobes of the spectrum. These are stronger for the CC interpolator w.r.t. the full stack and the DL interpolator. One can also observe an attenuation of the spectrum for the CC interpolator w.r.t. the full stack and the DL reconstructions. The Z3 EA encircles forest, it is noticeable that the ground layer, beneath the forest, is less intense for the CC case in Fig. 7-**c**, being so drastic that the ground and canopy layers are no longer distinguishable. In contrast, this does not occur with the estimation made by the DL method in Fig. 7-**d**.

C. Discussion

Consider the average interferometric coherence between the original SLCs and those estimated in the tomographic experiment (see Fig. 8). Overall, the DL-based interpolator demonstrates better performance compared to CC, particularly as the size of the gap between BLs increases. Nonetheless, for the gap of 30.8m (BL at 105.8m) CC performs marginally better. In order to understand this behavior, the temporal information of the acquisitions is studied.

The 17SARTOM acquisitions occurred over two days: D1 on May 17th, 2017, and D2 on May 18th, 2017. The TomoSAR stack includes BLs from both days, as marked in pink and yellow in Fig. 8. Most of the BLs used as input for the interpolators are from D1; namely, six belong to D1 and two are from D2. Among the four estimated BLs (red dots in Fig. 8), three are from D2 and only one (at +105.8m) was acquired on D1. The two nearest BLs (at +90.1m and +120.9m) to the targeted BL at +105.8m are from D1. For the other three estimated BLs, the two nearest BLs are from a different day. Since CC is a classical kernel interpolator [7], the two nearest points greatly influence the result. This suggests that in cases where the inputs are from the same day, CC may exhibit slightly superior performance.

Fig. 8 suggests that the DL-based method is more robust against temporal decorrelation. To corroborate this statement, we analyze the temporal coherence to evaluate if temporal decorrelation has occurred. Two BLs of 17SARTOM, acquired at different days, are very close to each other (less than 1m). Therefore, these BLs are employed to perform a zero-BL experiment, as done in [8]. In this way, the total coherence only depends on two factors, the SNR and time. SNR can be estimated with the characteristics of the sensor and the noise-equivalent-sigma-zero, measured during the mission (from -39.5 dB to -35.5 dB, depending on the area). After subtracting the SNR contribution from the total coherence, the temporal coherence of the scene is obtained, as shown in Fig. 9. The EA spots an area with high temporal decorrelation. This is the same area where the DL-based method (Fig. 9-**b**) provides a better estimation of the data in comparison to CC (Fig. 9-**a**).

IV. Conclusions

The analysis revealed that the proposed DL-based interpolator accurately estimates SLCs, showing satisfactory results in phase and intensity. The interpolated SLCs improved the tomographic focusing and demonstrated robustness in scenarios with temporal decorrelation and larger BLs, outperforming

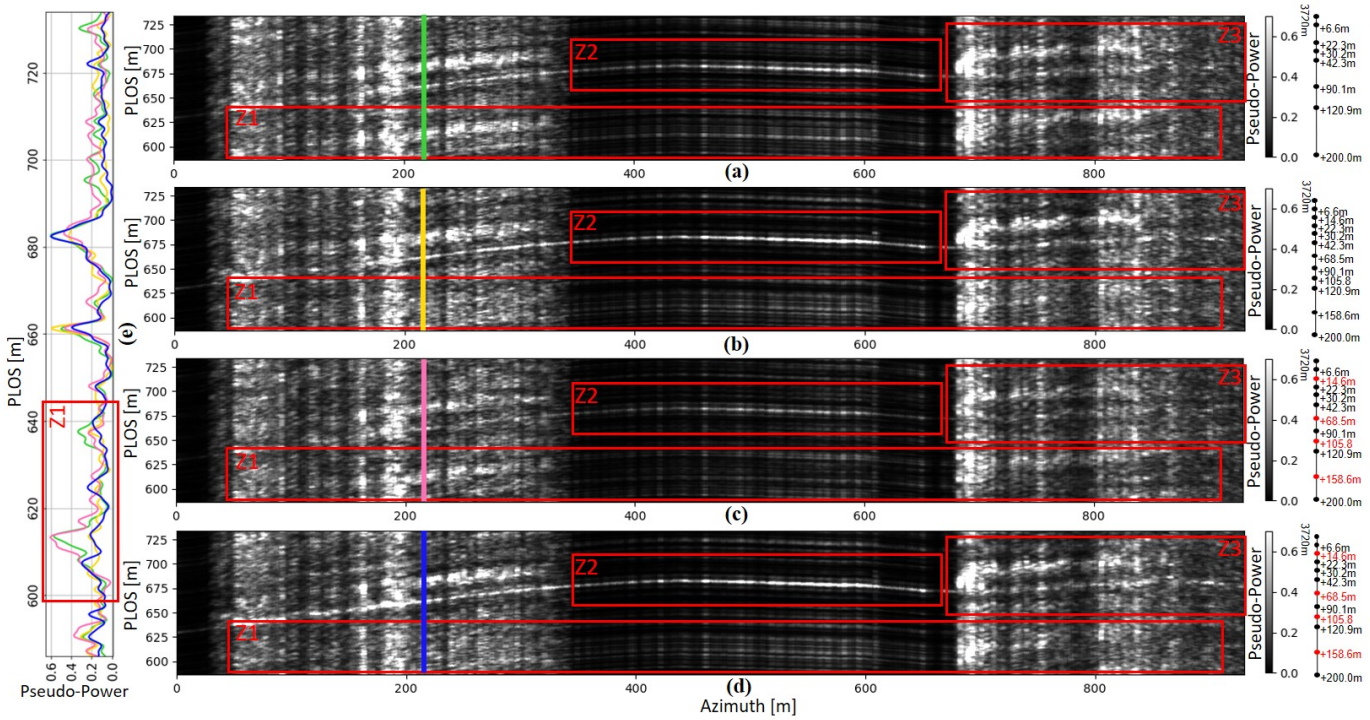

Fig. 7. Tomograms recoverd via MSF. **(a)** Using 8 original BLs. **(b)** Using 12 original BLs. **(c)** Using 8 original BLs + 4 CC estimated BLs. **(d)** Using 8 original BLs + 4 DL-based estimated BLs. **(e)** Exemplary vertical profiles, colors indicate the corresponding tomogram. The positions of the horizontal BLs are indicated on the right: the synthesized BLs marked in red, while the original BLs marked in black. The red rectangles indicate the EAs.

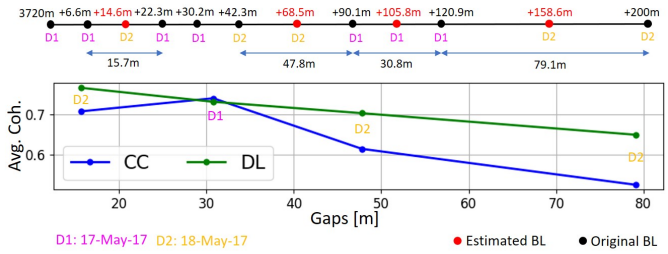

Fig. 8. Average coherence of the estimated SLC as a function of the gap between BLs to be filled up. On the top, the positions of the horizontal BLs: the synthesized ones are marked in red, while the original ones are marked in black.

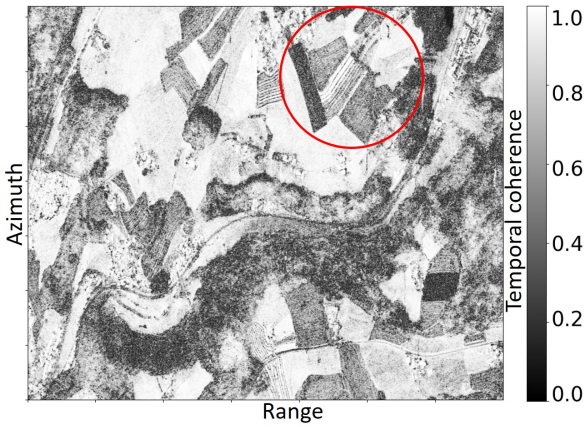

Fig. 9. Temporal coherence of the scene. The circles indicate the EA.

the classical CC interpolator. These findings suggest that a DL-based interpolator is a valid tool for improving TomoSAR results in campaigns with a limited number of flights.

REFERENCES

[1] A. Reigber and A. Moreira, "First demonstration of airborne sar tomography using multibaseline l-band data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2142–2152, 2000.

[2] G. D. Martín-del Campo-Becerra, S. A. Serafín-García, A. Reigber, and S. Ortega-Cisneros, "Parameter selection criteria for tomo-sar focusing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1580–1602, 2021.

[3] Z. Berenger, L. Denis, F. Tupin, L. Ferro-Famil, and Y. Huang, "A deep-learning approach for sar tomographic imaging of forested areas," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[4] W. Yang, S. Vitale, H. Aghababaei, G. Ferraioli, V. Pascazio, and G. Schirinzi, "A deep learning solution for height inversion on forested areas using single and dual polarimetric tomosar," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[5] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241.

[6] P. Rosen, S. Hensley, I. Joughin, F. Li, S. Madsen, E. Rodriguez, and R. Goldstein, "Synthetic aperture radar interferometry," *Proceedings of the IEEE*, vol. 88, no. 3, pp. 333–382, 2000.

[7] Y. Ivanenko, V. T. Vu, A. Batra, T. Kaiser, and M. I. Pettersson, "Interpolation methods with phase control for backprojection of complex-valued sar data," *Sensors*, vol. 22, no. 13, 2022.

[8] M. Simard, S. Hensley, M. Lavalle, R. Dubayah, N. Pinto, and M. Hofton, "An empirical assessment of temporal decorrelation using the uninhabited aerial vehicle synthetic aperture radar over forested landscapes," *Remote Sensing*, vol. 4, no. 4, pp. 975–986, 2012.