Harnessing Deep Learning for TomoSAR stack enhancement

Sergio Alejandro Serafin-Garcia, Matteo Nannini, Gustavo Daniel Martin-del-Campo-Becerra, Ronny Hänsch, and Andreas Reigber
German Aerospace Center (DLR), Oberpfaffenhofen, Germany

Abstract

Synthetic aperture radar (SAR) tomography (TomoSAR) utilizes co-registered SAR images from different tracks (known as a "TomoSAR stack") to create a resolution in the height direction. Spectral analysis retrieves the vertical backscattered Power Spectrum Pattern, enabling 3D imaging (Tomography). Tomograms show ambiguities inversely related to baseline separation, with larger and denser stacks offering better ambiguity rejection. To address the constraints posed by a limited number of acquisitions, we utilize a deep neural network. This network is employed to synthesize Single Look Complex SAR images by introducing an "artificial" baseline that was not part of the original TomoSAR stack.

1 Introduction

The objective of Synthetic aperture radar (SAR) tomography (TomoSAR) is to retrieve the interior distribution (referred to as Power Spectrum Pattern (PSP)) of semi-transparent objects using a set of spare SAR measurements. This set is called "TomoSAR stack" and is constituted by L SAR co-registered images acquired with different Baselines (BLs) in reference to a primary track. Each BL offers a perspective of the illuminated zone from a different Line-Of-Sight (LOS) allowing the synthesis of a resolution in Perpendicular to LOS (PLOS) direction. The reconstruction of PSP is classically obtained via spectrum estimation methods that involve an inversion of measurements [1, 2]. The recovered PSP presents a resolution in PLOS direction oppositely varying to the tomographic aperture (largest BL in the stack, later on called D_{PLOS} in **Figure** 1) [1, Eq. 8]. In addition, ambiguities caused by the subsampling, appear in a position inversely proportional to the separation between BLs (labeled as d in **Figure** 1) [1, Eq. 9]. Hence, with an increase in the size, reducing the distance between tracks, of the TomoSAR stack, the quality of the resulted tomograms improves in terms of ambiguities rejection. In real-world situations, the dimensions of the TomoSAR stack are limited by the revisit time, due to potential temporal decorrelation issues [3]. Another constraint on the number of tracks in the stack belongs to the practicality of executing singular missions to collect them.

We propose the use of a Deep Learning (DL) architecture to synthesize Single Look Complex (SLC) SAR images from an "artificial" baseline, i.e. a sensor path with a LOS not acquired in a specific TomoSAR stack. Deep learning is a subgroup of machine learning (ML) that employs Neural Networks (NN) with multiple layers to automatically learn and extract complex patterns from data. It allows handling intricate tasks like image recognition, natural language processing, and decision-making [6]. The TomoSAR focusing problem has already been addressed in terms of DL, see [4, 5]. These works use additional information, simulated data, and LiDAR respectively, as a way to have ground truth data to train the network to map from the TomoSAR stack to vertical profiles. This has the disadvantage of being attached to the quality and limitations of the additional information. In our case we use DL to improve the TomoSAR stack itself, so the classical methods to reconstruct the PSP can perform better.

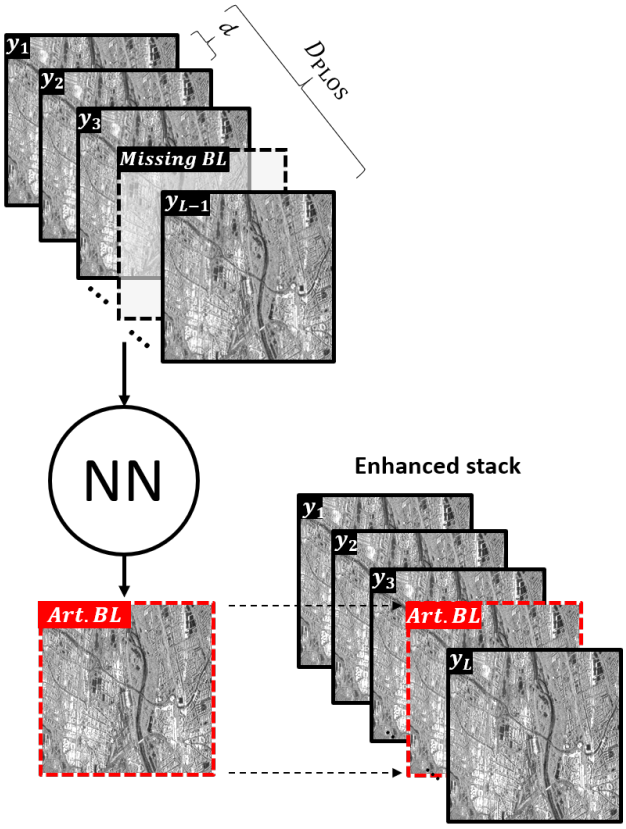

Figure 1 DL-based SLC synthesis process. A subset of known BLs are used to train a NN to generate one from an unknown BL position.

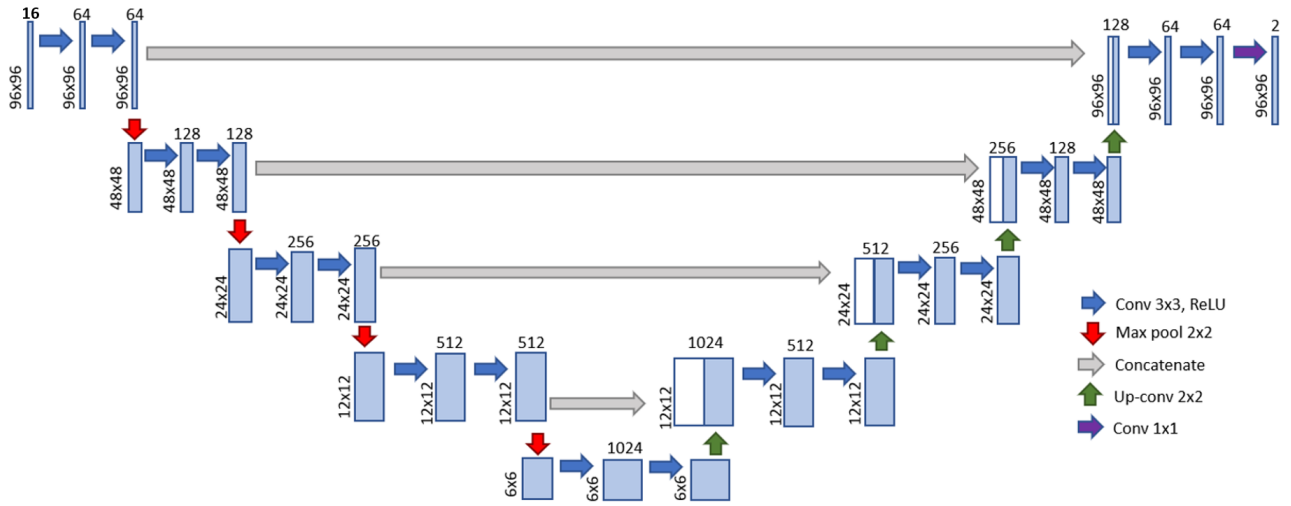

Figure 2 Proposed DL architecture based on classical U-net [7]. The network employs a contracting path for context capture and an expansive path for detailed feature incorporation, facilitated by skip connections.

As a proof of concept, the next experiment is proposed. A TomoSAR mission dataset is cropped in the azimuth direction in two subsets; one that is used for training and the other for testing. The training subset accounts for the L BLs while the test one contains only $L-1$. The reason is to imitate a scenario where the training area was acquired with the whole tracks and the testing zone was flown with fewer passes. Following that, a DL model is fitted with the data in the training subset to generate the missing SLCs in the test crop. This process is shown in **Figure** 1. Experiments were performed using the F-SAR dataset obtained in the 17SARTOM experiment over the test site of Traunstein, Germany in 2016.

2 TomoSAR signal model

In the framework of the direction of arrival estimation theory [3], the linear equation of observation

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{1}$$

is used to describe the TomoSAR inverse problem. This considers an acquisition geometry with L tracks, they all offering distinct LOS. Each pass yields a single co-registered SAR image, which is then coherently combined via SAR interferometric methods. When considering that co-registration is not dependent on height, these L passes are handled as a linear array. In equation (1), $\mathbf{y}$ collects the L processed signals for each pass at certain azimuth range position. Vector $\mathbf{s}$ contains the M values of the reflectivity at each elevation position $\{z_m\}_{m=1}^{M}$ in the PLOS direction. The interferometric phase information corresponding to the backscattering sources found along the height positions $\{z_m\}_{m=1}^{M}$ holds on matrix $\mathbf{A}$. This matrix is referred to as the "steering matrix" and its columns store the steering vectors $\{\mathbf{a}(z_m)\}_{m=1}^{M}$ whose definition is as in [3].
The PSP (whose reconstruction is the main objective of TomoSAR) in the PLOS direction is represented in discrete form by vector $\mathbf{b} = \{b_m\}_{m=1}^{M} = \{\langle |s_m|^2 \rangle\}_{m=1}^{M}$, the second-order statistics of the complex reflectivity vector $\mathbf{s}$. One of the most classical methods to solve the inverse problem stated in (1), is via the Fourier-based Matched Spatial Filter (MSF) [3],

$$\mathbf{b}_{\text{MSF}} = \mathbf{A}^{+}\mathbf{Y}\mathbf{A}. \tag{2}$$

This makes use of a sampled covariance matrix [3] depicted as:

$$\mathbf{Y} = \frac{1}{J}\sum_{j=1}^{J}\mathbf{y}_{(j)}\mathbf{y}_{(j)}^{+}, \tag{3}$$

here J states the number of independent looks of the signal acquisitions. In practical cases, TomoSAR is considered as an ergodic process. This means that the multi-looking is obtained via averaging adjacent values. MSF is very interesting in the context of our research because the position of the ambiguities can be computed assuming equidistant BLs [1, Eq. 9] via

$$V_{\text{PLOS}} = \frac{\lambda r_1}{2d}, \tag{4}$$

where r_1 represents the slant-range distance to the primary, λ the sensor wavelength and as previously stated d the distance between passes.

3 Deep learning model

DL methods use a cascade of nonlinear processing units to obtain features of a dataset under the idea of learning by example. This knowledge is later used to map an input into a desired output. A modified version of the U-net encoder-decoder network architecture [7] is employed. As can be seen in **Figure** 2, it consists of five encoder blocks and five decoder blocks. The contractive path (encoders) doubles the number of features and half them spatially at each step. While in the expansive path (decoders) at stepping, the dimensional size doubles and half the number of

filters. Each of the encoder blocks follows the next configuration. Two 3×3 convolution layers with padding 1, each of them activated by a Rectified Linear Unit (ReLU) function. After this, a 2×2 Maximum Pooling layer with stride 2 is applied. On the other side, every step in the expansive path is constituted by the next sequence. A 2×2 transposed convolutional layer with stride 2 followed by the concatenation of the up-sampled feature map with the feature map from the contracting path. That is to combine both low-level and high-level information. Subsequently, two 3×3 convolution layers with padding 1 and activated by a ReLU function are implemented. After the five blocks of the expansive path, a 1×1 convolution layer activated by a linear function is employed to return to the input size and state our regression model.

Our model takes as an input eight of the SLCs available in the stack. Each of the complex input images is represented with 2 channels: real part, and imaginary part. Both channels are normalized with the amplitude, this redundant information is incorporated to aid the NN to assimilate the magnitude of the complex vector that aims to approximate. Considering 96x96 input patches then our input can be defined as $\mathbf{X}^{96\times96\times16}$.

The remaining SLC in the stack is used as a target for the training, meaning that this is the BL that the network will learn to estimate. Since one BL is being approximated and the output data of the network is portrayed by two channels: amplitude and phase. Then, the output of the network can be presented as $f(\mathbf{X}) = \mathbf{W}^{96\times96\times2}$, where $f : \boldsymbol{X} \rightarrow \mathbf{W}$ represents the action of the NN mapping the input in to the output.

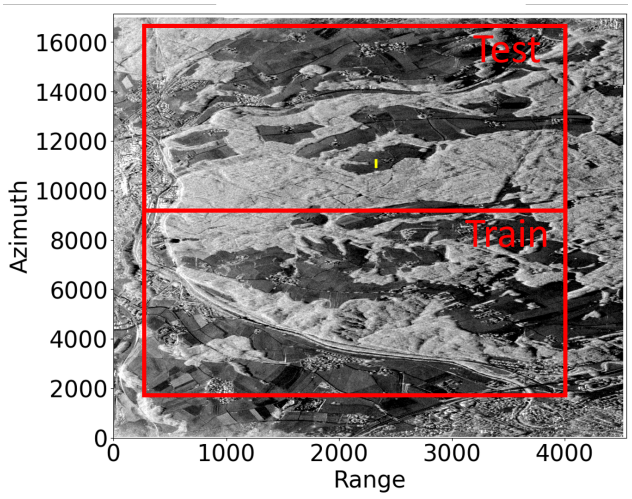

Figure 3 Test and train subset. The yellow line in the test subset is the slice where the tomographic experiments are performed.

The loss function used for the training is

$$Loss = Loss_{\emptyset} + Loss_{amp}. \tag{5}$$

Where the loss of the phase is given by,

$$Loss_{\emptyset} = \frac{1}{\mathrm{VAR}(\mathbf{W}_{\emptyset})n} \sum_{i=1}^{n} sub_{ang}(\mathbf{W}_{\emptyset_i}, \widehat{\mathbf{W}}_{\emptyset_i})^2. \tag{6}$$

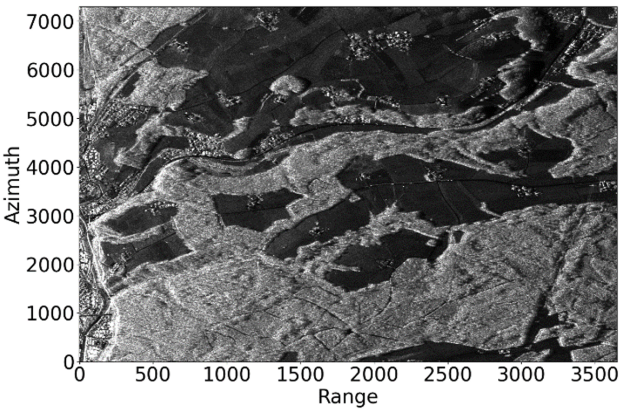

Figure 4 Original SLC (reflectivity), test subset.

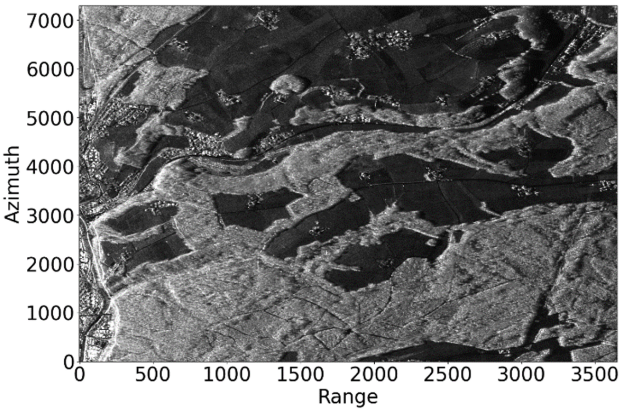

Figure 5 Artificial SLC (reflectivity), test subset.

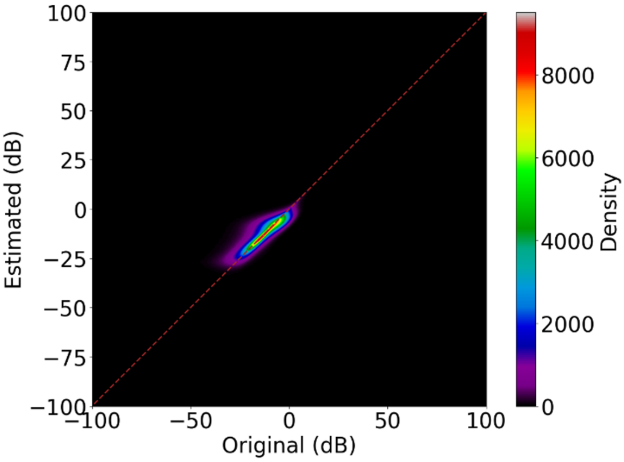

Figure 6 Scatter plot amplitude. In the horizontal axis the amplitude of the original SLC and in the vertical axis the artificial one.

This is the mean square error of the phase channel with its approximation by the NN, but with some modifications. The function $Sub_{ang}(a, b)$ represents the subtraction of angles a and b considering its circular disposition, giving us the realistic distance of the two phases. The number of samples is expressed with n. Also, the loss is scaled with the variance (VAR) in order to be in the same numeric range as the $Loss_{amp}$. The loss of the amplitude is defined

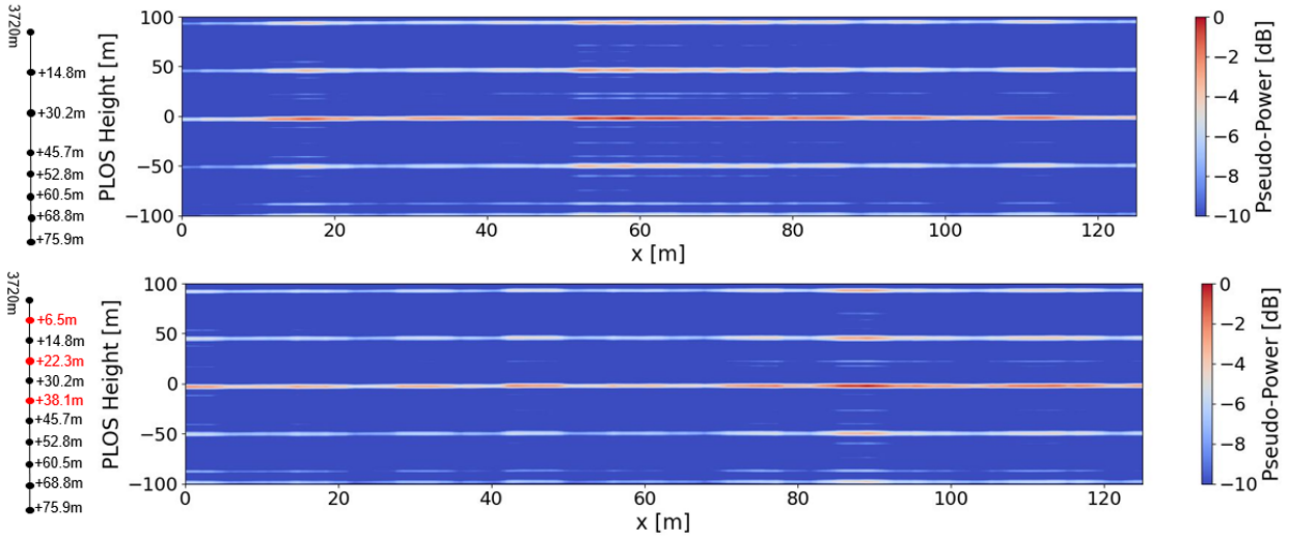

Figure 7 Tomographic results. (Top) MSF using 8 original BLs. (Bottom) MSF using only 8 original BLs + 3 artificial BLs. On the left, the horizontal BLs positions are displayed: in red the ones synthesized and in black the original ones

as,

$$Loss_{amp} = \frac{1}{\text{VAR}(\mathbf{W}_{amp})n} \sum_{i=1}^{n} (\mathbf{W}_{amp_i} - \widehat{\mathbf{W}}_{amp_i})^2, \quad (7)$$

with the scaling factor of the variance being used for the same reason as before.

4 Experimental results

As previously stated, we envision a scenario where multiple TomoSAR stacks are required to cover a region of interest. The goal is to acquire a large stack from only a part of the scene. This data can be subsequently leveraged to train the network to synthesize a subset of the stack. For the remaining parts of the scene, only a smaller stack has to be acquired, as it can be extended by SLCs synthesized by the network. As a proof of concept, we limit ourselves to a single tomographic stack that is divided into train and test regions by a split in azimuth. Experiments are performed using a crop of the F-SAR dataset obtained from the 17SARTOM mission over the test site of Traunstein, Germany in 2016. This campaign has the next specifications: An L-Band sensor (0.226m wavelength), multipolarimetric (only VV was used for our case), nominal height of 3720m, range resolution of 1.3m, and azimuth resolution of 0.6m.

The training and testing subsets of 7300x3650 pixels can be seen in **Figure** 3. The following experiments consider the horizontal BLs (in reference to the master track at 0m) at 14.8m, 22.3m, 30.2m, 45.7m, 52.8m, 60.5m, 68.8m and 75.9m. The one at 75.9m was selected as a target to be synthesized in the next example.

The model was fitted with the training subset using 75% of the data to train and 25% for validation. Mini-batches of size 32 were employed with an Adam optimizer using a learning rate of $1 * 10^{-4}$ for 140 epochs.

For the next results, we will refer to the test subset. In **Figure** 4, it is shown the original SLC (reflectivity). While **Figure** 5 displays the SLC generated by the NN. As it can be seen, the NN successfully recognizes most of the patterns that constitute the area.

In an effort to assess the results, **Figure** 6 displays the scatter plot of the original amplitude (dB) and the artificial amplitude (dB). In an ideal case, these two values should be the same, therefore the closer they are to the red diagonal the better the results are. For our case, it is observable that most of the density of the plot is lying on the red line, which means a good quality of the synthetic SLC, in terms of amplitude.

Finally, MSF was performed in the surface marked with a yellow line in the test subset, as can be seen in **Figure** 3.To highlight the impact of the artificial BLs in the tomogram's computation, three SLCs were estimated using the DL-based method proposed (in reference to the master at 0.0m): 6.5m, 22.3m, and 38.1m. In **Figure** 7 (top) we can see Tomogram reconstructed using eight BLs and in **Figure** 7 (bottom) the one using eleven BLs. It is observable a reduction in the ambiguity at approximately 25m.

5 Conclusions

In the framework of classical methods to solve the TomoSAR inverse problem, the quality of the tomograms, in terms of ambiguity rejection, is directly related to the size and density of the tomographic stack. Therefore, this work offers an option to enhance a spare TomoSAR stack taking advantage of DL.

The correct synthesis of an artificial BL was evaluated qualitatively. The correct recreation of the amplitude can be seen in the scatter plot, **Figure** 6. Finally, the tomograms displayed shows the correct enhancement of the results, in terms of ambiguity suppression.

In summary, the proposed method to improve a TomoSAR stack has proved to be usable specially in scenarios like surfaces or human-made structures. In future work, a deeper analysis and emphasis in the phase estimation needs

to be done in order to asses the performance in other kind of scenes.

6 Literature

[1] A. Reigber and A. Moreira, “First demonstration of airborne SAR tomography using multibaseline L-band data,” *IEEE Trans. Ge-osci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.

[2] 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*, vol. 1, no. 1, pp. 6-43, March 2013.

[3] G. D. Martín-del-Campo-Becerra, S. A. Serafín-García, A. Reigber and S. Ortega-Cisneros, “Parameter selection criteria for Tomo-SAR focusing,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1580–1602, Jan. 2021.

[4] Berenger, Zoé and Denis, Loïc and Tupin, Florence and Ferro-Famil, Laurent and Huang, Yue, “A Deep-Learning Approach for SAR Tomographic Imaging of Forested Areas,” *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1-5, 2023.

[5] Yang, Wenyu and Vitale, Sergio and Aghababaei, Hossein and Ferraioli, Giampaolo and Pascazio, Vito and Schirinzi, Gilda, “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.

[6] A. Shrestha and A. Mahmood, "Review of Deep Learning Algo-rithms and Architectures," in *IEEE Access*, vol. 7, pp. 53040-53065, 2019.

[7] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", *MICCAI*, vol. 9351, pp. 234-241, November 2015.