Advantages of Polarimetry and Interferometry for Semantic Segmentation of Urban SAR Images with Consideration of the Layover

Louis Newman, Flora Weissgerber, Aurélien Plyer, Élise Colin

Advantages of Polarimetry and Interferometry for Semantic Segmentation of Urban SAR Images with Consideration of the Layover

1st Louis Newman
DTIS, ONERA
Université Paris-Saclay
Palaiseau, France

2nd Flora Weissgerber
DTIS, ONERA
Université Paris-Saclay
Palaiseau, France

3rd Aurélien Plyer
DTIS, ONERA
Université Paris-Saclay
Palaiseau, France

4th Élise Colin
DTIS, ONERA
Université Paris-Saclay
Palaiseau, France

***Abstract*—In order to make the most of SAR and its different modalities (PolSAR, InSAR, PolInSAR) for urban area monitoring, a semantic segmentation of dual-polarimetric TanDEM-X images into ground and above-ground classes is carried out. To account for the layover, a pixel may be assigned to both classes at once. Ground truth labels are constructed by projecting 3D LiDAR point clouds onto the azimuth/range reference frame of the SAR image. The impact of PolSAR and InSAR is studied through a comparison to models trained on a single polarization SAR image. PolSAR data are found to be beneficial, except when training and testing on different polarimetric couples. InSAR data are highly beneficial if their ambiguity height is small, and detrimental if it is much larger than the typical building height.**

***Index Terms*—SAR, polarimetry, interferometry, layover, semantic segmentation, deep learning, urban areas**

I. Introduction

The automatic monitoring of man-made structures on a large scale in remote sensing imagery plays an important role in several fields, including urbanization monitoring and disaster response. Thanks to their all-weather capacity, Synthetic Aperture Radar (SAR) images are of particular interest for temporal monitoring. SAR data are difficult to interpret, especially because of their unusual geometry. In order to facilitate the interpretation of high resolution (HR) SAR images of an urban area, we propose a semantic segmentation of ground and above-ground classes. A pixel can contain both classes, which is especially common in SAR images due to the layover. This is particularly relevant for dense urban areas, as buildings often appear overlaid on the ground.

The TanDEM-X data were provided by DLR under the scientific proposal OTHER0103. This work would not have been possible without the collaboration of Paola Rizzoli and Jose Luis Bueso Bello in DLR, whom we thank very sincerely. This fruitful collaboration was initiated in the framework of the ONERA-DLR virtual research center for cooperation in "AI and applications in aerospace engineering".

SAR encompasses several modalities such as polarimetry (PolSAR) and interferometry (InSAR), which are more constraining and not necessarily available. Knowledge of their impact on the quality of the predictions of deep learning based semantic segmentation models can help in the selection of appropriate data for a given application. PolSAR and InSAR images have been previously used to train deep learning based segmentation models. The work [1] introduced a real length-six vector representation of the complex polarization coherency matrix. It was then applied to train a convolution neural network (CNN) for semantic segmentation of the low resolution NASA/JPL AIRSAR San Fransisco dataset. In [2], the same dataset was used to train a model taking as input the Pauli decomposition of the scattering matrix, in the form of RGB images. Two deep learning architectures for semantic segmentation of HR airborne polarimetric images were developed in [3]. Average Intersection over Union (mIoU) scores of 44% for a U-Net and 50% for a fully convolutional network were obtained. In [4], the authors tested various encoders to perform semantic segmentation of buildings and non-buildings on airborne PolSAR images, and obtained an mIoU score of 67%. In [5], a U-Net was trained to segment HR airborne SAR images of urban areas into "building", "road" or "other" classes. Mono-pass InSAR data were compared to a single SAR image, and the authors also investigated the information contents of the modulus and the phase of an image. The phase was found to be detrimental, and an mIoU of 74.67% was reached without it. The layover was treated as an artefact of the image, and the building footprints were used as labels. In a similar spirit, [6] designed a complex-valued neural network for semantic segmentation of buildings in InSAR images. This architecture was compared to other state of the art models on HR airborne images. In contrast to [5], the layover was detected as its own semantic segmentation class.

Image	I1	I2	I3
date [dd/mm/yyyy]	10/08/2021	21/08/2021	27/06/2021
polarizations	HH/HV	HH/VV	HH/HV
h_{amb}	39.6m	41.5m	320m

Table I: Relevant metadata of the three TDM acquisitions in the dataset

Following these findings, the choice was made to take the layover into account using multi-label encoding, as explained in section II. Section III presents the results of experiments comparing models trained on PolSAR and InSAR data to models trained on a single polarization SAR image.

II. Method

A. Dataset Creation

The considered dataset comprises three ascending mono-pass interferometric TanDEM-X (TDM) dual pol acquisitions, over the city of Toulouse, France. Two are in HH/HV polarization, and one in HH/VV polarization. Table I presents the corresponding scenarios, where h_{amb} denotes the TDM interferometric ambiguity height.

The first two acquisitions have similar ambiguity heights and different polarizations. The third acquisition contains the same polarizations as the first but an ambiguity height 10 times larger. The images were roughly cropped to their common footprint (around 3500×7000 pixels) and partitioned into 1024×1024 pixels tiles. The tiles from the first date were then randomly assigned to a training or an evaluation set, with the objective of obtaining identical class representations between these two sets and a $70\% - 30\%$ split. This partition was then reused to create similar sets with the two other acquisitions.

Ten real channels were created from the complex SAR data: firstly, the logarithms of two polarimetric amplitudes HH_m and XV_m (X is H or V depending on the acquisition) for the main satellite, and the phase difference between these polarizations ψ_m. Secondly, these same parameters for the secondary satellite: HH_s, XV_s, ψ_s. Finally, four additional channels for interferometry: two ϕ^k_{XY} phases for each available polarization XY, and the two coherences in absolute value γ^k_{XY}. The k index indicates the number of pixels of the boxcar filter involved in the coherence calculation. All channels were normalized to $[0, 1]$.

The labels were created by projecting airborne LiDAR data acquired in March and April 2019 by *Toulouse Métropole*. The LiDAR points had been assigned by *Toulouse Métropole* to the "ground" or the "above-ground" class, with the "above-ground" class typically containing vegetation or buildings. These points were projected onto the reference frame of the SAR SLC images, using their metadata. Two binary raster images of "ground" and "above-ground" were then derived, depending on whether the pixel contains a point of the concerned class. Ground truth labels are shown in Fig. 1. One can see that a pixel can belong to both binary ground and above-ground labels. This either corresponds to a mix of ground and above-ground points or a layover, distinctly noticeable on buildings. In high vegetation areas, noisy and arbitrary labels are introduced by the volumetric nature of LiDAR point clouds and the extreme nature of the ground truth computation: a pixel with one point of ground and 80 points of over-ground is labeled as belonging to both classes. These two-channel labels were used to train a deep neural network for multi-label semantic segmentation.

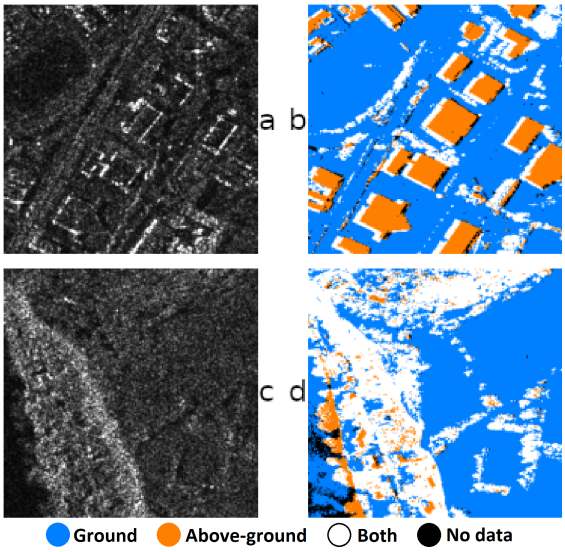

Figure 1: SAR image intensity patches (a, c) and corresponding ground truth labels (b, d). The bottom pair shows the noisy and arbitrary nature of the labels in high vegetation areas.

B. CNN-Based Semantic Segmentation Model

A U-Net with an efficientnet-b5 backbone and pre-trained weights on Imagenet from [7] was selected, since [8] showed that it is sufficient to obtain very accurate results. Each label channel being binary, a sigmoid layer and a binary cross-entropy loss were used. A stochastic gradient descent optimizer with a learning rate scheduled from 0.05 to 0.001 on 150 epochs was used. No data augmentation or hyperparameter optimization was performed. The training of each model took approximately 30 minutes on an NVIDIA RTX 2080Ti GPU.

III. Results and Discussion

Multiple experiments were carried out in order to study the impact of PolSAR and InSAR data on deep learning

	T_1E_1	T_1E_2	T_1E_3	T_2E_1	T_2E_2
HH_m	*61.26*	61.29	61.01	61.37	*61.48*
HH_s	*61.24*	61.05	60.68	61.21	*61.07*
XV_m	*60.99*	**59.50**	60.06	**59.78**	*61.67*

Table II: Average IoU [%] for mono-channel models.

	T_1E_1	T_1E_2	T_1E_3	T_2E_1	T_2E_2
HH_m XV_m ψ_m	*64.30*	**61.76**	63.57	**60.31**	*63.94*
HH_m ψ_m	*60.24*	53.96	59.80	56.32	*61.18*
HH_m XV_m	*64.42*	61.75	63.81	62.48	*63.24*

Table III: Average IoU [%] for PolSAR models.

based segmentation of urban areas. Models were trained on the training set of one acquisition, and evaluated on multiple evaluation sets. In the following tables, T_jE_k corresponds to the evaluation on **Ik**'s evaluation set of a model trained on **Ij**'s training set. The metric of choice is an average IoU, defined as the unweighted average of the IoU of the positive and negative classes of both the ground and above-ground channels (four terms in total).

A. Mono-Channel Experiment

In a first experiment, models were trained on a single channel in order to establish a baseline, as well as compare different polarizations. The results presented in Table II show that training on a single channel and testing on the same image yields similar performance across acquisitions, polarizations and main or secondary images. Training on one polarization and testing on another results in a performance loss, of slightly more than one point (in bold in the table).

B. PolSAR Experiment

With a baseline established, a second experiment was carried out to study the impact of additional polarimetric data, and its results are displayed in Table III. When training on the moduli and phase difference of two polarizations, gains of about three points were observed for both HH/HV and HH/VV combinations. Training on one polarization couple and testing on another yields results approximately three points worse than when testing on an identical couple (in bold in the table). The polarimetric phase is detrimental, but does not greatly reduce performance when used alongside both moduli.

C. InSAR Experiment

A similar experiment was then conducted for interferometric data, and its results can be found in Table IV. The introduction of the TDM interferometric information increases performance by approximately five points for a small h_{amb}, compared to a mono-channel model. Training with the larger h_{amb} is much less effective, resulting in improvements of only one point. This is in line with the fact that for a large h_{amb}, the object

	T_1E_1	T_1E_2	T_1E_3	T_3E_1	T_3E_3
$HH_{m,s}$ $\{\phi,\gamma\}^{25}_{HH}$	*66.11*	66.01	**49.23**	59.94	*62.46*
$XV_{m,s}$ $\{\phi,\gamma\}^{25}_{XV}$	*66.46*	61.23	49.68	59.85	*62.38*
$HH_{m,s}$ ϕ^{25}_{HH}	*66.01*	65.94	51.66	61.16	*62.25*
$HH_{m,s}$ γ^{25}_{HH}	*64.34*	64.04	51.33	59.54	*61.45*
$HH_{m,s}$	*63.13*	63.03	54.66	61.17	*61.22*
$HH_{m,s}$ ϕ^{1}_{HH}	*65.66*	65.32	53.17	61.15	*61.89*

Table IV: Average IoU [%] for InSAR models.

	T_1E_1	T_1E_2	T_1E_3	T_2E_1	T_2E_2
10 channels	*68.28*	**65.91**	53.02	**66.16**	*67.47*
10 channels \$\psi_{m,s}$	*68.17*	65.21	52.21		

Table V: Average IoU [%] for PolInSAR models.

details in the interferometric phase are hidden in the noise, which can be seen in Fig. 2. Additionally, training and testing with different ambiguity heights yields worse results than a mono-channel model (in bold in the table). Removing the coherence slightly decreases performance, while removing the phase leads to a drop in the IoU score of almost two points, suggesting that the phase contains almost all the information. Using the moduli only further reduces the scores (underlined in Table IV), and yields worse results than a model trained on the moduli of two polarizations (third line in Table III). This shows that the interferometric information is richer than two correlated speckle realizations. In [5], the mono-pass interferometric phase was found to worsen performance. This is in line with our results, since their antennas were mounted on one airplane, leading to a large h_{amb}. Training and testing on TDM pairs in different polarizations is worse than training and testing on pairs in the same polarization, by about five points.

D. PolInSAR Experiment

In a fourth experiment, the combination of polarimetric and interferometric data was compared to the previous results. The corresponding data are presented in Table V. When training and testing on the same image, using all 10 input channels increases performance by one to four points compared to the best results obtained on TDM or polarimetric data only. The 10 channel model performs well when tested on an acquisition with a different polarimetric channel, with a one point drop in mIoU relative to the best results (in bold in the table). On the contrary, the 10 channel model performs very poorly for acquisitions with different ambiguity heights, as evidenced by worse results than a mono-channel model (underlined in the table).

E. Qualitative Results

In Fig. 2, semantic segmentation results are shown for the 10 channel model trained on **I1**'s training set (first line of Table V). For **I1** and **I2**, the interferometric phase varies

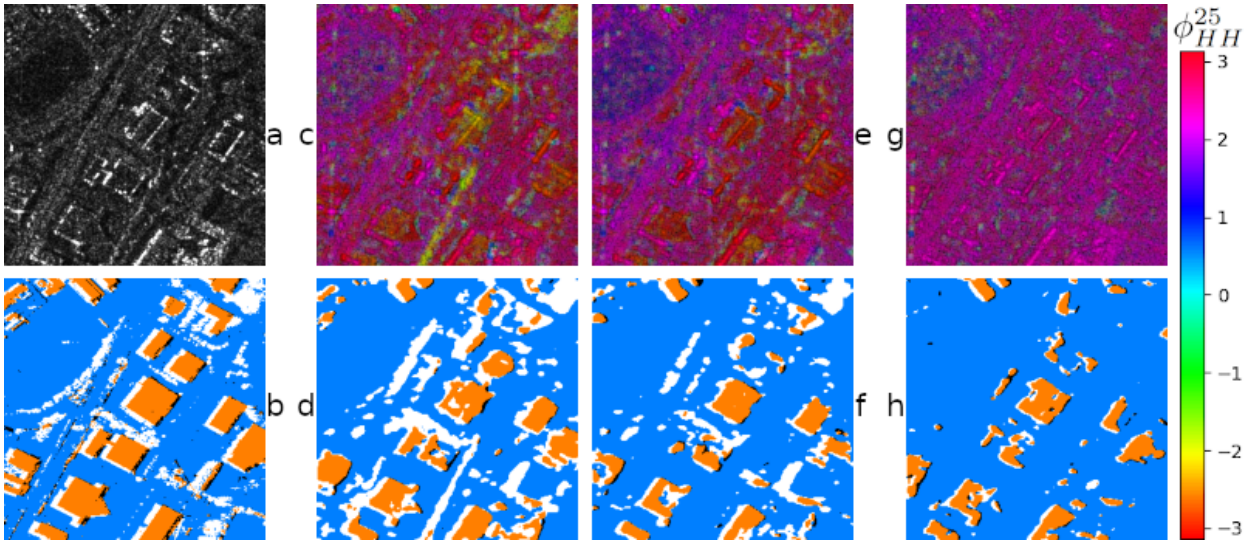

Figure 2: Results for the 10 channel model trained on **I1**'s training set: HH_m SAR patch intensity (a), ground truth labels (b), HH interferograms (HSV representation of ϕ^{25}_{HH}, γ^{25}_{HH}, HH_m) (c, e, g) and predicted labels (d, f, h) for all three evaluation sets (from left to right, **I1**, **I2**, **I3**).

noticeably between the ground and the top of above-ground objects, as a consequence of their small h_{amb}. The corresponding predictions are relatively accurate. Conversely, due to its large h_{amb}, the phase on **I3** is much more uniform, and the predictions are much poorer.

The worst results were mainly obtained on patches with high vegetation, such as the second row of Fig. 1, for which the ground truth labels are relatively noisy and arbitrary.

IV. Conclusion

In this work, we present a solution for semantic segmentation of SAR images that takes into account the layover effect through a multi-label representation. We explore the impact of the performance of the different components of the PolSAR and InSAR modalities of the TDM mission. We have observed that InSAR data are beneficial for this task, but only for a small h_{amb}. Additionally, having a different h_{amb} for the training and the test interferometric data leads to a large domain shift and worse performance than a mono-channel model. Separately, performance gains were also achieved with PolSAR data. Finally, the combination of both modalities outperformed all previous results, demonstrating non redundancy between them.

Acquiring polarimetric data comes at the expense of a coarser spatial resolution. It will thus be interesting to train a model on data from a nonpolarimetric acquisition over the same area, and compare the results against the PolSAR experiment results. Another point of investigation will be the use of multi-pass interferometric data.

References

[1] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, "Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 1935–1939, Dec. 2016.

[2] S.-H. Wang, J. Sun, P. Phillips, G. Zhao, and Y.-D. Zhang, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *Journal of Real-Time Image Processing*, vol. 15, pp. 631–642, Oct. 2018.

[3] W. Wu, H. Li, X. Li, H. Guo, and L. Zhang, "PolSAR Image Semantic Segmentation Based on Deep Transfer Learning—Realizing Smooth Classification With Small Training Sets," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, pp. 977–981, June 2019.

[4] T. Kumari, F. H. Khan, T. Halder, R. K. Gayen, A. M. Ray, and D. Chakravarty, "Semantic Segmentation of Urban Areas in Polarimetric SAR Imaging using Deep Neural Networks and Decision Trees," in *2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS)*, pp. 479–482, Dec. 2021.

[5] X. Wang, L. Cavigelli, M. Eggimann, M. Magno, and L. Benini, "HR-SAR-Net: A Deep Neural Network for Urban Scene Segmentation from High-Resolution SAR Data," in *2020 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, Mar. 2020.

[6] J. Chen, X. Qiu, C. Ding, and Y. Wu, "CVCMFF Net: Complex-Valued Convolutional and Multifeature Fusion Network for Building Semantic Segmentation of InSAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[7] P. Iakubovskii, "Segmentation models pytorch." https://github.com/qubvel/segmentation_models.pytorch, 2019.

[8] J. Xia, N. Yokoya, B. Adriano, L. Zhang, G. Li, and Z. Wang, "A Benchmark High-Resolution GaoFen-3 SAR Dataset for Building Semantic Segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5950–5963, 2021.