FOREST MAPPING WITH TANDEM-X INSAR DATA AND SELF-SUPERVISED LEARNING

Jose-Luis Bueso-Bello[1], *Benjamin Chauvel*[1,2], *Daniel Carcereri*[1], *Ronny Hänsch*[1], *Paola Rizzoli*[1]

[1]German Aerospace Center (DLR), Microwaves and Radar Institute, 82234 Wessling, Germany
[2]ENSTA Bretagne, 29200 Brest, France

ABSTRACT

Deep learning models trained in a fully supervised way have shown encouraging capabilities for mapping forests with TanDEM-X interferometric data, being able to generate time-tagged forest maps at large-scale over tropical forests. These maps have been generated at 50 m resolution to reduce the computation burden. In this work, we now aim to exploit the high-resolution capabilities of the TanDEM-X interferometric dataset, processed at only 6 m resolution. In order to cope with the lack of reliable reference data at such high resolution, we focus on the investigation of self-supervised learning approaches. The availability of a reference map over Pennsylvania, USA, based on Lidar acquisitions at 1 m resolution, allows us to compare different deep learning approaches. First promising results show the possibility to extend the proposed self-supervised learning approach over areas where the lack of reference data prevent us from using fully supervised deep learning methods.

***Index Terms*—** Synthetic Aperture Radar, rainforest, deforestation monitoring, deep learning, convolutional neural network, U-Net, autoencoder

1. INTRODUCTION

Forests play an essential role in supporting life on our planet. Their monitoring is of key importance for assessing changes in forest coverage and biomass. This can be done by generating and analyzing large-scale forest maps using spaceborne remote sensing. Previous global large-scale forest maps have been generated using optical or hyperspectral data, such as the 30 m resolution world forest coverage map derived from Landsat [1]. More recent global maps of land cover have been generated at 10 m resolution using optical data and including a tree cover layer, such as the ESA WorldCover 2021 map [2]. However, optical-based approaches can be hindered by cloud coverage, particularly over tropical regions and north latitude areas, which are characterized by long rainy seasons that obscure the ground from view for several months per year.

Synthetic Aperture Radar (SAR) systems offer an attractive solution to monitor these areas thanks to their ability to acquire data independently from weather and daylight conditions. The first global forest map based on SAR images was generated from L-band ALOS PALSAR satellite data, using cross-polarization backscatter images, and was provided at a posting of 25 m [3]. Recent studies have also demonstrated the usefulness of Interferometric Synthetic Aperture Radar (InSAR) systems for monitoring vegetated areas, particularly the added value of the interferometric coherence, as provided by the TanDEM-X (TerraSAR-X add-on for Digital Elevation Measurement) bistatic mission [4]. The interferometric coherence, defined as the normalized complex correlation coefficient between the two InSAR acquisitions, gives information about the amount of noise in the interferograms and is sensitive to different decorrelation sources, such as the limited signal-to-noise ratio and volume scattering mechanisms. This last aspect is quantified by the volume decorrelation factor (γ_{Vol}), which was the main input feature for the generation of the global TanDEM-X Forest/Non-Forest (FNF) map, generated using a machine learning (ML) fuzzy clustering algorithm and released at a resolution of 50×50 m [4].

The potential of deep learning (DL) approaches for forest mapping with TanDEM-X images is demonstrated in [5], which compares three state-of-the-art Convolutional Neural Networks (CNNs) architectures, showing very promising performance. A U-Net architecture [6] resulted to be the most effective one and was the starting point for a DL approach with further model generalization capabilities, where an adhoc training strategy was developed to distinguish forests in TanDEM-X images acquired with different acquisition geometries over the Amazon rainforest [7]. The classification improvements reached by applying DL methods on TanDEM-X data, have allowed for the generation of time-tagged mosaics at 50 m resolution over tropical forests by utilizing all nominal TanDEM-X acquisitions between 2011 and 2017, skipping the weighted mosaicking of overlapping images used in the clustering approach for achieving a good final accuracy. Furthermore, the trained U-Net over the Amazon rainforest has been used to extend the forest mapping to other tropical forests over Africa and Asia, also showing a high accuracy and a good agreement with other land cover maps [7].

The objective of the present study is to extend the previous work by exploiting the full-resolution TanDEM-X InSAR dataset. By applying sophisticated InSAR processing techniques [8], it is possible to process the TanDEM-X single-look slant-range complex images, acquired in stripmap single-polarization mode, to an independent pixel spacing of only 6 m. With such high-resolution data, we aim at improving the forest mapping accuracy and detecting forest degradation phenomena. Deforestation paths in the middle of dense forested areas, which are not visible at 50 m resolution, can be successfully detected using 6 m resolution images. Moreover, a finer contour delimitation of the deforested areas is expected. However, the lack of reliable reference data at such high resolution to train a fully-supervised (FL) DL approach moved us to the investigation of a self-supervised learning (SSL) DL approach. An available forest reference map of 2010 at 1 m resolution over Pennsylvania, USA, allowed us to investigate the performance of SSL DL methods. The investigations and first results are now presented in this paper.

The paper is organized as follows: in Section 2, the current DL approaches for the exploitation of the full-resolution TanDEM-X bistatic dataset are presented. In Section 3 the data used in the training, validation and testing of the different DL approaches are introduced. Preliminary results of this study are presented in Section 4. Finally, in Section 5 the conclusions are drawn.

2. METHODOLOGY

The main strategy of our work is depicted in Figure 1, where SSL is used to improve the final classification with supervised learning (SL). For the SSL part, the goal is to train a model (e.g. an autoencoder) that maps an image to a representation of visual contents without the necessity of human annotation, expecting that the extracted features will benefit the forest mapping downstream task in the SL part, by reducing the amount of necessary labeled reference data to train a U-Net for semantic segmentation.

2.1. Self-supervised learning

The purpose of autoencoders is to efficiently encode the input data by learning the most informative features in the data rather than every single small detail. While there are plenty of SSL methods used for remote sensing applications [9], only very few studies have applied them to single-pass InSAR data. In the current study, we evaluate a standard and a masked autoencoder, denoted here as identity and inpainting tasks. While both aim at reconstructing the original input image, the masked autoencoder has to tackle the additional challenge that part of the input is artificially occluded. In our case several requirements drove the design of the autoencoder. Since the weights of the encoder need to be transferable to the U-Net, it needs to have the same structure as the U-Net encoding path. The decoder part is only made of transposed convolutions without skip connections from the encoder.

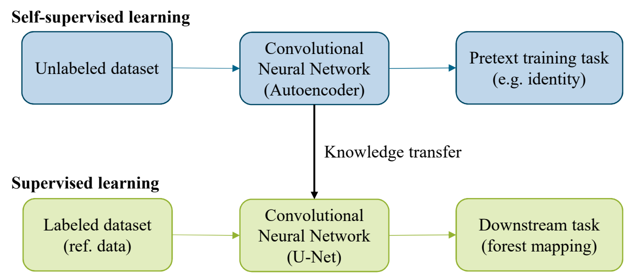

Fig. 1. Strategy to combine DL-based approaches for forest mapping with TanDEM-X high-resolution data.

2.2. Supervised learning

For the following SL part, we built on the previous works in [5, 7]. A U-Net-shaped CNN is considered in our approach, since it showed the best performance for forest classification using TanDEM-X data [5].

2.3. Scenarios

We defined different scenarios to assess the impact of a self-supervised pre-training on the downstream task, as well as to find a compromise between the final performance and the amount of reference data required to reach it. The best case scenario, which represents our baseline, consists of training a U-Net model in a fully supervised fashion with as much reference data as possible. This scenario is defined as FL100, meaning fully-supervised learning with 100% of the labeled data. Different competing scenarios are created based on: a) the pretext task used in the SSL part (identity or inpainting); b) the type of training after transferring the weights from SSL to FL. Two possibilities have been tested: Freezing the transferred encoder weights of the U-Net and training only its decoder (D) part or using the weights for initialization of the encoder part of the U-Net, but afterwards training the whole U-Net (D + E); c) usage of a reduced amount of labeled data in the SL part (1.5%, 8%, and 22%), selected from the ones used for the FL100 case and being representative of the different TanDEM-X acquisition geometries.

3. DATA

To overcome the lack of reliable reference data at resolutions < 10 m, which are useful to properly train FL DL methods as in [7], we now investigate SSL techniques with TanDEM-X data acquired over the state of Pennsylvania, USA, where a reliable and high-resolution forest map is available.

3.1. Test area and reference map

The state of Pennsylvania, USA, has been selected as test area thanks to the availability of a high-resolution and reli-

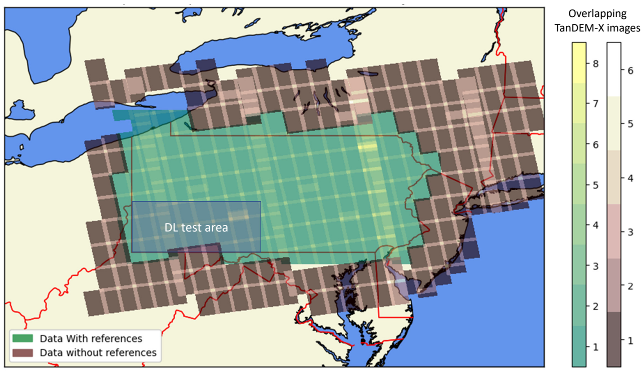

Fig. 2. TanDEM-X images acquired in 2011 over Pennsylvania, USA. For the TanDEM-X images indicated in green, a high-resolution reference forest/non-forest map is available.

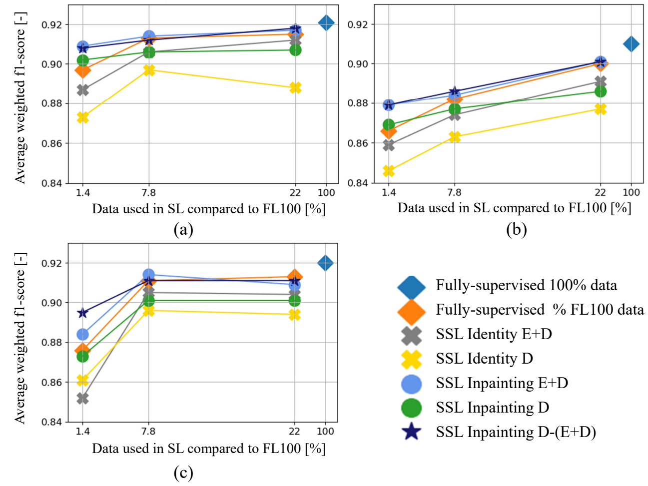

Fig. 3. Performance results obtained for the TanDEM-X images acquired over the test area reserved for the evaluation of the different implemented DL approaches. a) Test images with $h_{\text{amb}} < 40$ m; b) Test images with $h_{\text{amb}} \in [40-50]$ m; c) Test images with $h_{\text{amb}} > 65$ m.

able forest map. The used reference data is based on optical and Lidar data, acquired over Pennsylvania up to 2010, which were combined to generate a forest/non-forest classification map with a ground resolution of 1×1 m [10]. For our work, we downsampled the original resolution to match that of the TanDEM-X images. By counting the input pixels within a cell of 6×6 m, the majority class (forest or non-forest) has been set as reference for the map at 6 m resolution and used for the following investigations.

3.2. TanDEM-X InSAR dataset

The use of Φ-Net, a developed DL strategy for InSAR parameter estimation and denoising [8], allows for generating TanDEM-X InSAR products with a 6 m resolution. As in the previous work [7], we rely on the backscatter, the interferometric coherence, and the volume decorrelation factor as main input features from the TanDEM-X InSAR dataset. To describe the acquisition geometry, the height of ambiguity (h_{amb}) and the local incidence angle are selected as inputs for our DL approaches, too. To minimize the time span between reference data and TanDEM-X acquisitions, mainly acquisitions of 2011 have been used for training purposes only. Some TanDEM-X data of 2012 have been utilized as well to extend the range of h_{amb} seen by the CNN in the training and validation processes. In all cases, the input dataset for training, validation and testing, is divided into patches of 128×128 pixels with the 5 channels defined by the considered TanDEM-X input features.

Figure 2 shows the TanDEM-X images acquired in 2011 over Pennsylvania, USA. In green are depicted TanDEM-X acquisitions which overlay the area with reference data. These acquisitions are employed by SL approaches as well as to maximize the extent of the training dataset for the SSL part. The TanDEM-X images indicated in brown are used only in the SSL investigations, since no reference data are available. To transfer knowledge as relevant as possible from the autoencoder to the U-Net, it is necessary that the autoencoder learns from the complete range of possible acquisition geometries of TanDEM-X considering both h_{amb} and orbit directions. We use acquisitions from 2011, 2012, 2013 and 2018 for the SSL training and validation. A test area for the different DL scenarios, representative of the Pennsylvania's landscape, has been selected for testing the models (blue shadowed area in Figure 2). These images are under no circumstances used for any learning task, in order to provide a rigorous performance evaluation of the model. For the purpose of testing on the whole variability of h_{amb} we build different sets of test acquisitions in which we distinguish three ranges of h_{amb}, named low, medium, and high, corresponding to $h_{\text{amb}} < 40$ m, $h_{\text{amb}} \in [40-50]$ m and $h_{\text{amb}} > 65$ m.

4. RESULTS

The obtained results for the test dataset are presented in Figure 3. Up to three simulations have been run for each DL approach. The average f1-score, weighted by the number of pixels belonging to each class (forest/non-forest), is depicted. In general for the whole test dataset and for all scenarios, the performance improves when using more labeled data in the SL part. Using the inpainting pretext task in the SSL training, better results are achieved. With respect to the trainability of the SL part, the most competitive results are obtained when just initializing the weights of the encoder and training both encoder and decoder parts of the U-Net.

First results are shown in Figure 4, where the detected forest map with TanDEM-X images at 6 m resolution is overlaid to an optical image obtained from Google Earth. A high

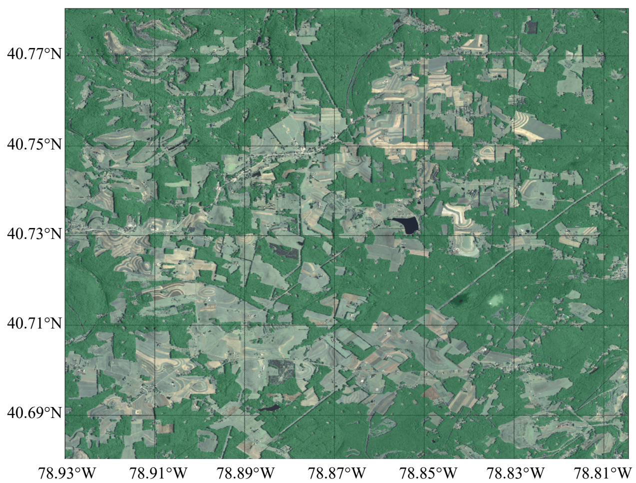

Fig. 4. Overlay of the detected forest areas with TanDEM-X at 6 m resolution over an optical image, obtained from Google Earth.

agreement is observed with respect to the optical image. The borders of the forested areas are very well delimited and single lines of trees are well detected. Narrow roads in between the forests are visible as well.

5. CONCLUSIONS AND OUTLOOK

In this study we have successfully demonstrated the effectiveness of deep convolutional neural networks for mapping forests using TanDEM-X bistatic InSAR acquisitions at a resolution of 6 m. To address the challenge of limited referenced data at such a high resolution, we proposed and evaluated different self-supervised pre-training approaches. Specifically, the use of inpainting and sufficient data representing all TanDEM-X acquisition geometries showed considerable benefits, such as a better performance and stability during training than the other competitive scenarios. The successful implementation of the self-supervised pre-training strategy is extremely promising, particularly in regions like the Amazon rainforest, where reference labelled data is scarce and challenging to obtain. This approach opens up new possibilities for accurate forest mapping with TanDEM-X bistatic images leading to improved environmental monitoring and conservation efforts over such areas. In the final paper, we will present more consolidated results and validation.

6. REFERENCES

[1] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehamn, S. J. Goetz, T. R. Loveland, and J. R. G. Kommareddy, "High-resolution global maps of 21st century forest coverage change," *Science*, vol. 342, pp. 850–853, Nov. 2013.

[2] D. Zanaga, R. Van De Kerchove, D. Daems, W. De Keersmaecker, C. Brockmann, G. Kirches, J. Wevers, O. Cartus, M. Santoro, S. Fritz, M. Lesiv, M. Herold, N.E. Tsendbazar, P. Xu, F. Ramoino, and O. Arino, "ESA WorldCover 10 m 2021 v200," 2022.

[3] M. Shimada, T. Itoh, T. Motooka, M. Watanabe, T. Shiraishi, R. Thapa, and R. Lucas, "New global forest/non-forest maps from ALOS PALSAR data (2007-2010)," *Remote Sensing of Environment*, vol. 155, pp. 13–31, 2014.

[4] M. Martone, P. Rizzoli, C. Wecklich, C. Gonzalez, J.-L. Bueso-Bello, P. Valdo, D. Schulze, M. Zink, G. Krieger, and A. Moreira, "The Global Forest/Non-Forest Map from TanDEM-X Interferometric SAR Data," *Remote Sensing of Environment*, vol. 205, pp. 352–373, Feb. 2018.

[5] A. Mazza, F. Sica, P. Rizzoli, and G. Scarpa, "TanDEM-X Forest Mapping Using Convolutional Neural Networks," *Remote Sensing*, vol. 11, no. 24, pp. 2980, 2019.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Munich, Germany 5 – 9 October*, N. Navab, J. Hornegger, W.M. Wells, and A.F. Frangi, Eds., Cham, Switzerland, 2015, pp. 234–241, Springer International Publishing.

[7] J.L. Bueso-Bello, D. Carcereri, M. Martone, C. González, P. Posovszky, and P. Rizzoli, "Deep Learning for Mapping Tropical Forests with TanDEM-X Bistatic InSAR Data," *Remote Sensing*, vol. 14, no. 16, pp. 3981, 2022.

[8] F. Sica, G. Gobbi, P. Rizzoli, and L. Bruzzone, "Φ-Net: Deep Residual Learning for InSAR Parameters Estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 3917–3941, 2021.

[9] Y. Wang, C.M. Albrecht, N.A.A. Braham, L. Mou, and X.X. Zhu, "Self-Supervised Learning in Remote Sensing: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 4, pp. 213–247, 2022.

[10] J. O'Neil-Dunne, S. MacFaden, A. Royar, M. Reis, R. Dubayah, and A. Swatantran, "An object-based approach to statewide land cover mapping," in *Proceedings of ASPRS 2014 annual conference*, 2014, pp. 23–28.