Deep Learning-based Approaches for Forest Mapping with TanDEM-X Interferometric Data

Jose-Luis Bueso-Bello[a], Benjamin Chauvel[a,b], Daniel Carcereri[a], Ronny Hänsch[a], and Paola Rizzoli[a]
[a]Microwaves and Radar Institute, German Aerospace Center (DLR), 82234 Wessling, Germany
[b]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, for forest mapping purposes. In order to cope with the lack of reliable reference data at such a 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. The obtained 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.

1 Introduction

Forest areas, occupying around 30% of the Earth's land surface, play an essential role in supporting life on our planet. They are aprimary player to reduce greenhouse gas concentrations, allowing control of climate change, and they serve as natural habitats for a variety of animal species, preserving biodiversity and ensuring healthy ecosystems. Moreover, they also provide valuable resources such as biomass, food, and livelihoods for many people worldwide. For this reason, the generation of large-scale forest maps is of key importance for monitoring changes in forest coverage and biomass. Previous global forest maps have been generated using optical or hyperspectral data, such as the 30 m resolution world forest coverage map derived from Landsat optical data between 2010 and 2015 [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 Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) map from 2017 [2] and the ESA WorldCover 2021 map [3].

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 coverage 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 [4]. Recent studies have also demonstrated the usefulness of interferometric SAR (InSAR) systems for monitoring vegetated areas, particularly the added value of the interferometric coherence [5].

In this context, the TanDEM-X (TerraSAR-X add-on for Digital Elevation Measurement) mission maps the Earth's surface providing single images at high-resolution from the recorded backscattered signal as well as highly accurate unique InSAR products acquired in bistatic mode [6]. The advantage of flying two satellites in close formation, constituting a single-pass InSAR system, adds valuable information to the amplitude data, such as the the interferometric phase and coherence. 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 fuzzy clustering algorithm and released at a resolution of 50×50 m [5].

The potential of Deep Learning (DL) for feature extraction and forest mapping from TanDEM-X SAR images has been demonstrated in [7] and [8]. The classification improvements applying DL methods on TanDEM-X data have allowed for the generation of time-tagged mosaics at 50 m resolution over the tropical forests by utilizing the nominal TanDEM-X acquisitions between 2011 and 2017, with no need for sophisticated mosaicking strategies and for the averaging of multiple coverages in order to achieve a satisfying accuracy. The objective of the present study is to extend the previous work by exploiting the full-resolution TanDEM-X InSAR dataset. Promis-

ing preliminary results were presented in [7] and [9], both based on a limited set of single TanDEM-X full-resolution images over a temperate forest in Pennsylvania (USA). In [9], the TanDEM-X interferometric acquisitions were processed with a non-local filtering technique and they were classified using a similar clustering approach as the one used for the global TanDEM-X FNF map [5]. By applying further sophisticated InSAR processing techniques [10], it is possible to process the TanDEM-X single-look slant-range complex images acquired at 3 m resolution (stripmap single-polarization mode) to an independent pixel spacing of 6 m. With such high-resolution data, we aim at improving the forest mapping accuracy and to detect forest degradation phenomena. Deforestation paths in the middle of dense forested areas, which were not visible at 50 m resolution, as well as a finer contour delimitation of the deforested areas, are expected to be successfully detected using 6 m resolution images. However, the lack of reference data at such a high-resolution to train a Supervised Learning (SL) DL approach, moved us to the investigation of a Self-Supervised Learning (SSL) DL approach with TanDEM-X InSAR data. The investigations and results are presented in this paper.

The paper is organized as follows: in Section 2, past investigations for large-scale forest mapping with TanDEM-X with DL techniques are summarized. The current DL developments for the exploitation of the full-resolution TanDEM-X bistatic dataset, including first results, are presented in Section 3. Finally, in Section 4 the conclusions are drawn.

2 State of the art in forest mapping with DL and TanDEM-X

To generate binary classification maps through semantic segmentation for forest classification using TanDEM-X bistatic InSAR data and DL techniques, a study was conducted in [7]. Three state-of-the-art Convolutional Neural Networks (CNNs) architectures were investigated, namely Residual Network (ResNet), Dense Network (DenseNet), and U-Shaped Network (U-Net). Different input features, including the backscatter, the interferometric coherence, and the volume decorrelation coefficient, were used to train the different CNNs, with the aim of evaluating their impact on the final classification accuracy. The study was performed on a small dataset consisting of individual TanDEM-X images at full-resolution (12 m) covering a temperate forest area in Pennsylvania, USA. A U-Net architecture [11], was able to achieve the best overall performance out of all the analyzed models. However, the results lacked adequate generalization of the model for large-scale mapping, as a larger and more diverse training dataset was needed.

Based on this work, a deep convolutional neural network approach for large-scale mapping of forests and water surfaces using TanDEM-X bistatic InSAR acquisitions at 50 m resolution was presented in [8]. The proposed approach extends the generalization capabilities of the model by training on a large amount of different geometries. It employed TanDEM-X data across the Amazon region for the training process. The basic idea involved utilizing the height of ambiguity (h_{amb}) as the main characterizing feature of the TanDEM-X acquisition geometry. This was combined with the local incidence angle, in order to account for variabilities in the slant-range direction. Furthermore, a specific training approach was implemented, considering distinct ranges of acquisition incidence angles. More than 50,000 TanDEM-X images acquired over the tropical forests between the end of 2010 and the middle of 2017 were classified with the trained CNN. In general, thanks to its capabilities in understanding two-dimensional patterns, the CNN performs much better with respect to the clustering approach, as it is able to generate forest maps with continuous forested and non-forested areas, which are less noisy. This allowed for forest mapping over tropical regions regardless of the geometry of acquisition. Up to three time-tagged mosaics have been generated over the Amazon rain forest, and two over the other tropical forest areas located in Africa and South-East Asia, respectively. The results showed a good agreement with external reference maps and a significant accuracy improvement with respect to the baseline clustering approach utilized for the generation of the global TanDEM-X FNF map.

3 Exploiting the TanDEM-X Full-resolution Dataset

The exploitation of the full-resolution TanDEM-X interferometric dataset is possible also thanks to newly developed DL strategies for InSAR parameter estimation and denoising [10]. In particular, the new processing skills introduced by the use of the Φ-Net allow to generate TanDEM-X InSAR products with a 6 m resolution. With the generation of forest maps at this resolution we aim at mitigating previous limitations, such as the detection of narrow roads and small clear-cuts within the vegetated areas, which occurred when considering mid-resolution data at 50 m.

To overcome the lack of reliable reference data at resolutions < 10 m, which are useful to properly train SL DL methods as in [8], we now investigate SSL techniques. The main purpose of our study is to assess the relevance of SSL approaches to reduce the amount of referenced data needed for training a DL architecture to effectively map forests in different environments using TanDEM-X data.

3.1 Test area and input dataset

The state of Pennsylvania, USA, has been selected as test area due to the availability of a high-resolution and reliable forest map, courtesy from the University of Maryland [12]. In **Figure 1** the green area indicates where reference data are available. Optical and Lidar data, acquired over this region up to 2010, were combined to generate a forest/non-forest classification map with a ground resolution of 1 m × 1 m [12]. For our work, we downsampled the original resolution to match that of the TanDEM-X images. By counting the input pixels within a

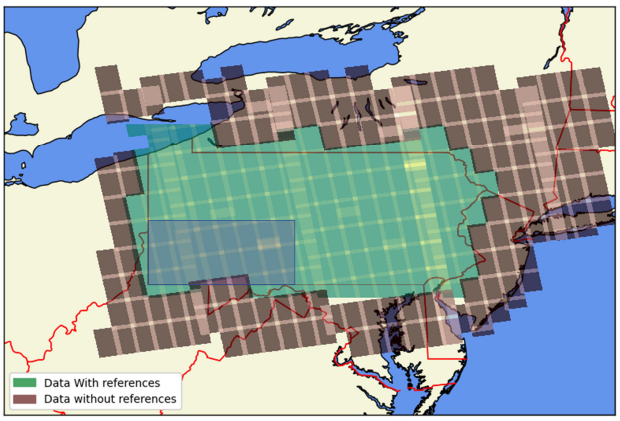

Figure 1 TanDEM-X images acquired in 2011 over Pennsylvania, USA. For the TanDEM-X images indicated in green, high-resolution reference data is available. TanDEM-X images shown in brown are only used in the SSL approach.

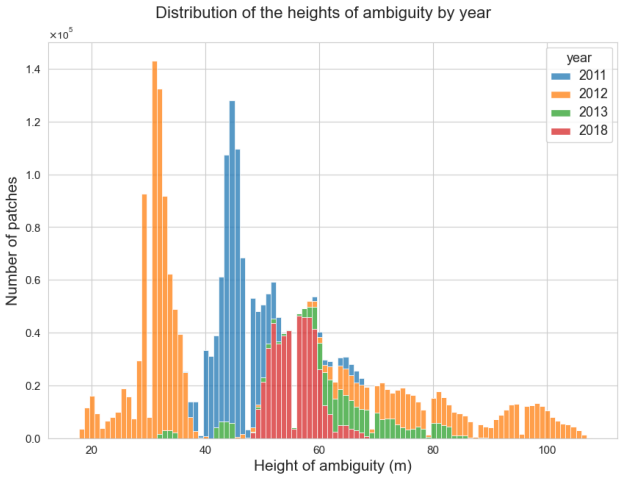

Figure 2 Distribution of the height of ambiguity, for different years, with which TanDEM-X images were acquired over Pennsylvania, USA.

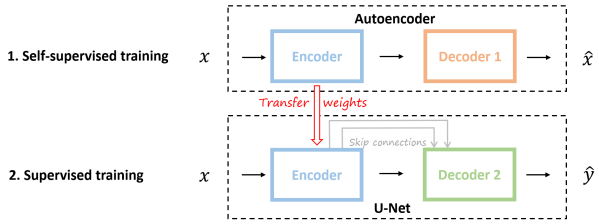

Figure 3 Strategy to combine DL-based approaches for forest mapping with TanDEM-X data.

cell of 6 m × 6 m, the majority class (forest or non-forest) has been set asreference for the map at 6 m resolution and used for further investigations. A test area, representative of the Pennsylvania's landscape, has been selected for testing the models (blue shadowed area in **Figure 1**). These data are under no circumstances used for any learning task in order to have results as unbiased as possible. Moreover, the test images are representative of the whole TanDEM-X acquisition geometries. Please note, the TanDEM-X acquisitions depicted in brown in **Figure 1** are used only in the SSL investigations, since no reference data are available.

As in previous works [8], 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 h_{amb} and the local incidence angle are selected as inputs, too. **Figure 2** shows the distribution of the h_{amb} over the test region for different years. TanDEM-X acquisitions are available for almost all considered h_{amb} values. Please note that, 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.

3.2 Deep learning-based approaches

The main strategyof our work is depicted in (**Figure 3**), where SSL is used to improve the final classification with 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 downstream tasks. For the SL part, based on previous works, a U-Net-shaped CNN is considered in our approach, since it showed the best performance for forest classification using TanDEM-X data [7]. After a first self-supervised training, which does not require the use of external reference data, we transfer weights from the autoencoder to the U-Net for the segmentation task.

Autoencoders are DL architectures capable of reducing the dimensionality of the input data and reconstruct it from the lowdimensional space called latent space. 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 [13], only very few studies have applied them for InSAR data. In the current study, we evaluate a standard and a masked autoencoder - denoted as "identity" and "inpainting" below. While both aim to reconstruct the original input image, the masked autoencoder has to tackle the additional challenge that part of the input is artifically 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. **Figure 4** shows the chosen design, where the encoding part is the same as the U-Net and the decoder part is only made of transposed convolutions without skip connections from the encoder.

The proposed approach aims to determine the impact of a self-supervised pre-training on the downstream task so as

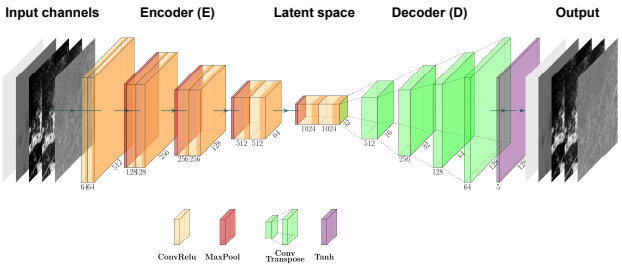

Figure 4 Autoencoder architecture derived from a U-Net-shape CNN. The encoding part is common to both CNNs, while the decoder part is only made of transposed convolutions.

to find a compromise between the final performance and the amount of reference data required to reach it. For this purpose, different scenarios have been defined. The best case scenario, where only the U-Net is used, consists of training a model with as many reference data as possible. The employed TanDEM-X data were acquired between 2011 and 2012, chosen to closely align with thereference data, and is representative of the global variability of the h_{amb}. This scenario is defined as FL100, meaning fully supervised learning with 100% of the data inside the green area defined in **Figure 1**. Different competing scenarios are created based on: a) the pretext task used in the SSL part (identity (Id) or inpaiting (In)); b) the type of training after transferring the weights from SSL to SL. 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 input data in the SL part. The input acquisitions (1.5%, 8%, and 22%) are randomly selected from the ones used for the FL100 case but keeping the constraint of being representative of the different h_{amb}.

To transfer knowledge as relevant as possible to the U-Net, the autoencoder needs to learn from the complete range of possible geometries of acquisition of TanDEM-X considering both h_{amb} and orbit directions. We use acquisitions from inside and in the vicinity of Pennsylvania, USA, from 2011, 2012, 2013 and 2018 for the SSL training and validation (data inside the brown area defined in **Figure 1**). Acquisitions with available references are also employed to optimize the utilization of data for the SSL, aiming to maximize the dataset's extent. The validation data are selected on the basis of geographical coordinates and are representative of h_{amb} variability and all types of terrain. For the purpose of testing on the whole variability of h_{amb} we build different sets of test acquisitions of similar sizes in which we distinguish $h_{\text{amb}} < 40$ m (low h_{amb}), $h_{\text{amb}} \in [40-50]$ m (medium h_{amb}) and $h_{\text{amb}} > 65$ m (high h_{amb}).

3.3 Results

The obtained results for the different scenarios and for the test dataset are summarized in **Table 1**. Instead of overall accuracy, the F1-score is used to assess the performance, since the proportion of forest is unbalanced among the considered patches. The different scenarios are compared to the ideal case FL100. In general all scenarios obtained very high performance. 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.

An example of reference Lidar forest maps and the corresponding confusion matrices for TanDEM-X of two different areas of 1024x1024 pixels are presented in **Figure 5** for visual inspection purposes. The results are color-coded with respect to the terms of the confusion matrix: True Positives (TP, pixels correctly classified as forest), False Negatives (FN, pixels wrongly classified as non-forest), False Positives (FP, pixels wrongly classified as forest) and True Negatives (TN, pixels correctly classified as non-forest). The FL100 ideal scenario shows a good agreement with the reference ground truth. Regarding the different SSL scenarios under investigation, and using only 1.5% of the total input data with available references, the best performance is obtained for the scenario using inpainting as pretext task in the SSL training and training both the encoder and decoder part of the network during the SL part. In the other scenarios (FL and SSL In D) a higher number of FP and FN pixels is observed.

In **Figure 6** a zoom in of a TanDEM-X forest map, acquired with $h_{\text{amb}} = 44\ m$ in 2011 over the defined test region in **Figure 1**, is compared with the corresponding Lidar reference data. Narrow paths and small holes in the forested areas are well detected. Also forest areas are correctly delimited. This forest map has been obtained using inpainting in the SSL part. With respect to the SL part, encoder and decoder have been trained and only 1.5% of the available TanDEM-X images overlapping the Lidar reference data have been used.

Table 1 Obtained F_1-score for different testing scenarios. FL100 refers to the ideal scenario, using 100% of the TanDEM-X data where reference data are available. For all other competing scenarios, only a percentage of the total input data are used, as indicated. Pretext tasks in SSL training: identity (Id) and inpainting (In). Use of transferred weight in the SL training: only trainable decoder (D) and encoder + decoder trainable (D + E).

F1-score	FL100	FL	SSL Id D	SSL Id E+D	SSL In D	SSL In E+D
SL with 1.5% of data						
Low h_{amb}	**0.920**	0.894	0.868	0.886	0.898	**0.906**
Mid h_{amb}	**0.910**	0.857	0.843	0.857	0.859	**0.878**
High h_{amb}	**0.919**	0.859	0.852	0.852	0.863	**0.878**
SL with 8% of data						
Low h_{amb}	**0.920**	0.910	0.894	0.903	0.904	**0.913**
Mid h_{amb}	**0.910**	0.876	0.855	0.867	0.875	**0.884**
High h_{amb}	**0.919**	0.907	0.886	0.890	0.902	**0.911**
SL with 22% of data						
Low h_{amb}	**0.920**	**0.916**	0.892	0.909	0.908	0.914
Mid h_{amb}	**0.910**	0.898	0.875	0.890	0.885	**0.901**
High h_{amb}	**0.919**	**0.911**	0.889	0.905	0.899	0.908

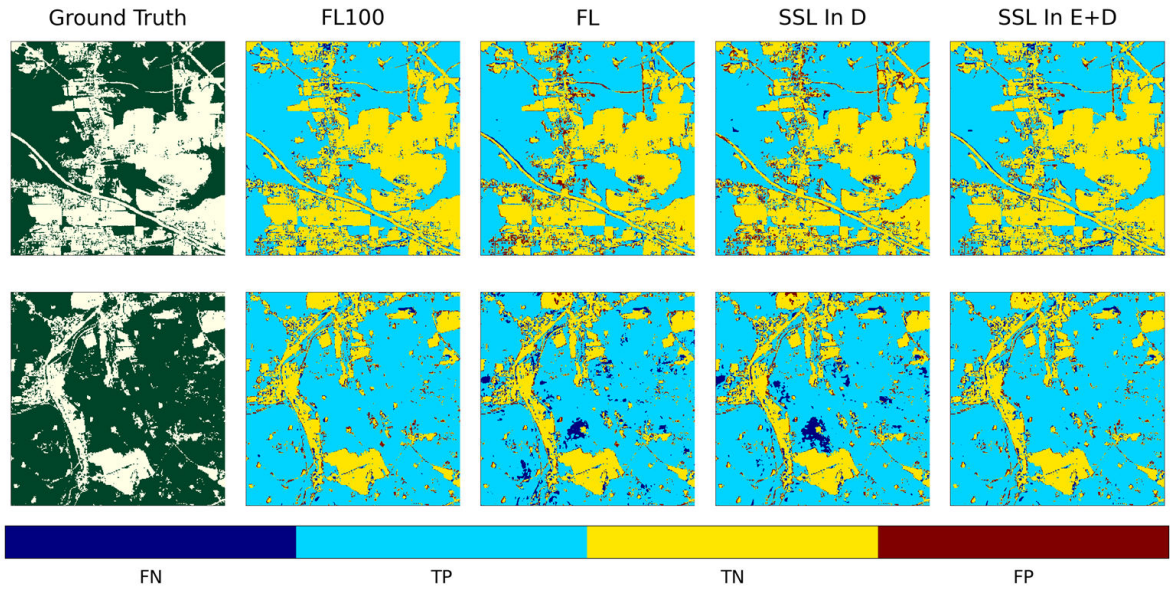

Figure 5 Qualitative comparison of results for 2 different areas of 1024x1024 pixels. On the ground truth plots, green areas correspond to forests and white areas to non-forested zones.

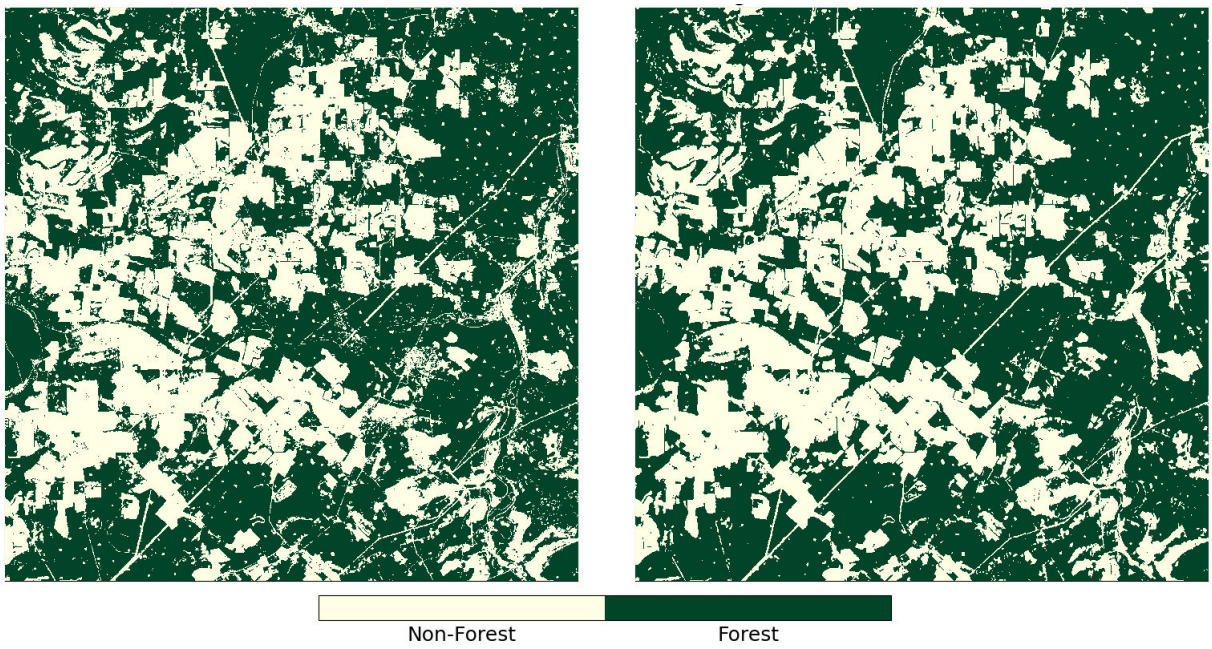

Figure 6 Comparison between the reference Lidar forest map (left) and the predicted TanDEM-X forest map obtained through SSL (right).

4 Conclusions and outlook

By extending existing research and considering a training strategy accounting for the global variability of TanDEM-X acquisitions with respect to the acquisition geometry, we have successfully demonstrated the effectiveness of convolutional networks for mapping forests using TanDEM-X bistatic InSAR data at a resolution of 6 m.

To address the challenge of limited referenced data at this resolution, we proposed and evaluated self-supervised pretraining approaches with the identity reconstruction and the inpainting as pretext tasks. Specifically using inpainting and sufficient data representing all TanDEM-X acquisition geometries, showed considerable benefits, such as better performance and stability during training than the other competitive scenarios.

The successful implementation of the self-supervised pretraining strategy is very promising, particularly in regions like the Amazon rain forest, where reference labelled data is scarce and challenging to obtain. This approach opens new possibilities for accurate forest mapping with TanDEM-X bistatic images leading to improved environmental monitoring and conservation efforts over such areas.

5 Literature

[1] Hansen, M.C.: Potapov, P.: Moore, R.: Hancher, M.: Turubanova, S.: Tyukavina, A.: Thau, D.: Stehman, S.: Goetz, S.: Loveland, T.: Kommareddy, A.: Egorov, A.: Chini, L.: Justice, C.O.: Townshend, J.: High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342. 850–853, 2013.

[2] Gong, P. et al.: Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017, Science Bulletin, vol. 64, no. 6, 2019, pp. 370–373.

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

[4] Masanobu, S.: Takuya, I.: Takeshi, M.: Manabu, W.: Tomohiro, S.: Rajesh, T.: Richard, L.: New global forest/non-forest maps from ALOS PALSAR data (2007–2010). Remote Sensing of Environment, 2014, vol. 155, pp. 13–31.

[5] Martone, M.: Rizzoli, P.: Wecklich, W.: González, C.: Bueso-Bello, J.L.: Valdo, P.: Schulze, D.: Zink, M.: Krieger, G.: Moreira, A.: The global forest/non-forest map from TanDEM-X interferometric SAR data. Remote Sensing of Environment, Feb. 2018, vol. 205, pp. 352–373.

[6] Rizzoli, P.: Martone, M.: Gonzalez, C.: Wecklich, C.: Borla Tridon, D.: Bräutigam, B.: Bachmann, M.: Schulze, D.: Fritz, T.: Huber, M.: Wessel, B.: Krieger, G.: Zink, M.: Moreira, A.: Generation and performance assessment of the global TanDEM-X digital elevation model, ISPRS Journal of Photogrammetry and Remote Sensing, Oct. 2017, vol. 132, pp. 119–139.

[7] Mazza, A.: Sica, F.: Rizzoli, P.: Scarpa, G.: TanDEM-X Forest Mapping using Convolutional Neural Networks. Remote Sens. 2019, 11(24), 2980.

[8] Bueso-Bello, J.L.: Carcereri, D.: Martone, M.: González, C.: Posovszky, P.: Rizzoli, P.: Deep Learning for Mapping Tropical Forests with TanDEM-X Bistatic InSAR Data. Remote Sens. 2022, 14(16), 3981.

[9] Martone, M.: Sica, F.: González, C.: Bueso-Bello, J.L.: Valdo, P.: Rizzoli, P.: High-Resolution Forest Mapping from TanDEM-X Interferometric Data Exploiting Nonlocal Filtering. Remote Sensing, 2018, 10, 1477.

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

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

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

[13] Wang, Y.: Albrecht, C.M.: Ait Ali Braham, N.: Mou, L.: Zhu, X.X.: Self-supervised Learning in Remote Sensing: A Review. IEEE Geoscience and Remote Sensing Magazine (GRSM), 2022, 10 (4), pp. 213–247.