SEMI-SUPERVISED DEEP LEARNING REPRESENTATIONS IN EARTH OBSERVATION BASED FOREST MANAGEMENT

Oleg Antropov[1], Matthieu Molinier[1], Rıdvan Salih Kuzu[2], Lloyd Hughes[3], Marc Rußwurm[3], Devis Tuia[3], Corneliu Octavian Dumitru[2], Shaojia Ge[4], Sudipan Saha[5,6], Xiao Xiang Zhu[6]

[1] VTT Technical Research Centre of Finland, P. O. Box 1000, FI-02044 VTT, Finland
[2]Remote Sensing Technology Institute, German Aerospace Center, Weßling, 82234, Germany
[3]Ecole Polytechnique Fédérale de Lausanne (EPFL), Sion, Switzerland
[4]Nanjing University of Science and Technology, 210094 Nanjing, China
[5]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi, India
[6]Data Science in Earth Observation, Technical University of Munich, Munich, 80333, Germany

ABSTRACT

In this study, we examine the potential of several self-supervised deep learning models in predicting forest attributes and detecting forest changes using ESA Sentinel-1 and Sentinel-2 images. The performance of the proposed deep learning models is compared to established conventional machine learning approaches. Studied use-cases include mapping of forest disturbance (windthrown forests, snowload damages) using deep change vector analysis, forest height mapping using UNet+ based models, Momentum contrast and regression modeling. Study areas were represented by several boreal forest sites in Finland. Our results indicate that developed methods allow to achieve superior classification and prediction accuracies compared to traditional methodologies and mimimize the amount of necessary in-situ forestry data.

Index Terms— forest management, deep learning, forest height, boreal zone, satellite image, Sentinel-1, Sentinel-2.

1. INTRODUCTION

Using satellite Earth Observation (EO) data along with in-situ forest measurements is an established approach that allows producing spatially explicit forest variable predictions and forest aerial estimates at various scales [1]. Within model-based forest inference, a model describing the relationship between forest reference and EO sensor-measured signature or its change is used [2]. In wide-area forest mapping, statistical, physics-based, and machine learning (ML) methodologies were used for modeling and prediction purposes, primarily with satellite optical and imaging radar sensors. Deep learning (DL) methodologies are widely adopted for various image classification, and semantic segmentation tasks [3, 4, 5]. To date, several fully convolutional and recurrent neural networks were demonstrated in forest remote sensing [6, 7, 8, 9, 10, 11]. These models often provide improved accuracy in forest classification or predicting forest variables, as well as in forest change mapping. Training of DL models often requires a fully segmented reference label, such as Lidar based forest maps that are costly and not available over wide areas.

On the other hand, a typical scenario in forest management is collecting a sample of forest plots to be further used in forest resource inference. Also, an image-patch level estimate can be available. Within the DL context, such reference data can be considered weak labels [12] and require transfer learning techniques for model fine-tuning.

This paper briefly summarizes our recent advances in developing semi-supervised DL models for forest inventory and change mapping. We consider several use-cases utilizing Sentinel-2 (S2) and Sentinel-1 (S1) images and forest measured plots from Finnish Forest Centre, comparing achieved results with more traditional mapping approaches.

This work was financed by the European Space Agency under contract 4000137253/22/I-DT, Non-supervised representation learning for Sentinels (RepreSent).

2. USE-CASES : DATA AND METHODS

2.1. Forest disturbance mapping using Deep Change Vector Analysis

In the context of forest disturbance caused by natural hazards (such as windstorm and snow-load), training labels are often missing or challenging to collect. Deep Change Vector Analysis (DCVA) [13] provides a viable solution for mapping damaged areas without training a DL model. DCVA is an unsupervised learning technique that utilizes a pre-trained network to extract deep features from bi-temporal images enabling delineation of image-change areas. The main benefit of DCVA is that it does not require prior information and compares deep features extracted from the bi-temporal images on a pixel-by-pixel basis. We examine the potential of DCVA in forest disturbance mapping in two separate use-cases.

First use-case focuses on mapping wind-thrown forest area using Copernicus images. A major windstorm caused severe forest damages in Puolanka area in Finland on June 22nd, 2021 (see Figure 1). We use S2 images acquired before and immediately after the forest windfall for developing the forest damage area mapping method.

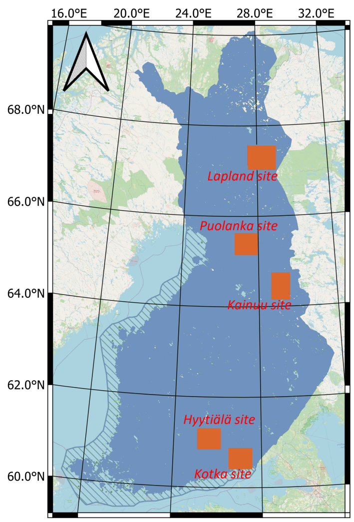

Fig. 1. Forest study sites in Finland

In another use-case study, a series of 12 S1 images are collected to delineate damaged forest areas caused by heavy snowload during winter 2017-2018 in the Kainuu province of Finland [14]. Noteworthy, this type of damage is continuous and can get severe within this time frame. Additionally, other types of changes can be visible in SAR images over boreal forests in connection with weather and seasonal conditions.

2.2. Forest variable prediction using self-supervision and regression methods

In this use case, we predict forest canopy height at the Hyytiälä site using S1 and S2 imagery. We use self-supervised pretraining with fine-tuning on a selection of in-situ forestry parameter samples to learn a regression model for this task.

First we apply Momentum Contrast (MoCo) [15] to train modality-specific (S1 and S2) encoder networks. These networks are trained with a cross-modality contrastive loss on the Sen12MS dataset [16]. The pretrained encoders subsequently form the backbone of our regression network. The output representations from these encoders are merged into a 1024-dimensional vector, which is then regressed via a single linear projection layer. For training this layer in a supervised fashion, we utilize a dataset composed of 340 forest plots with confirmed and validated canopy heights.

Furthermore, we tested a Model Agnostic Meta Learning (MAML) approach [17]. MAML is a few-shot transfer learning approach that extracts common knowledge from a collection of different-but-related tasks. Each task is represented by a small set of annotated samples. Training over thousands of iterations with different batches of source tasks yields a meta-model with parameters that are optimized to previously unseen tasks with few training samples. Here, we fine-tune the meta-model on S2 image patches and minimize the mean squared error to the reference tree height target variable.

The training dataset uses $40m \times 40m$ (4px), $120m \times 120m$ (12px) and $240m \times 240m$ (24px) image patches extracted from both S1 and S2 imagery, with 2 and 13 bands, respectively. In-situ measured forest plots are used as training data.

2.3. Forest mapping using UNet+ and transfer learning

A UNet+ model pretrained over sparse taiga forest (Lapland site) using fully segmented reference labels was used in forest height prediction in southern Finland where only a limited number of forest plots are available for calibrating the model. An improved semi-supervised deep learning UNet+ model allows to produce a spatially explicit pixel-level forest inventory. A Squeeze-and-Excitation attention module is used to recalibrate the multi-source features using channel self-attention [8]. The UNet+ model is firstly pretrained using accessible Lidar based forest data over Lapland and further fine-tuned using a small sample of forest plots from the target study site (Kotka site in Figure 1). The EO datasets are represented by 14-channel tensor with S2 image bands, S1 yearly composite bands, as well as ALOS-2 PALSAR-2 and TanDEM-X image features. The model training uses image patches with size 256 px $\times$ 256 px, all input images preprocessed and resampled to 10m $\times$ 10m spacing.

3. RESULTS AND DISCUSSION

Developed methodologies have consistently produced better results than conventional ML methods or required smaller amount of training data. Further we briefly discuss results of each studied use-case and developed methodologies.

3.1. Forest disturbance mapping using DCVA

Examples of produced forest damage maps due to snow-load damage and windstorm are shown in Figure 2. In snow-load damage mapping, using the available set of validation stand-level labels, up to $77.5.\%$ accuracy has been achieved in the initial experiments using the DCVA approach. Importantly, tested DCVA approach allows change detection between bi-temporal S2 images. Baseline methodologies using multitemporal stand-level features report accuracy levels in the range of 71-90 % depending on the method (logistic regression, support vector machines or improved kNN) [14]. While DCVA results already appear competitive, further work will concentrate on extending the analysis to adapt DCVA to deal with S1 image time series, as well as to integrate elements of contrastive learning into methodology.

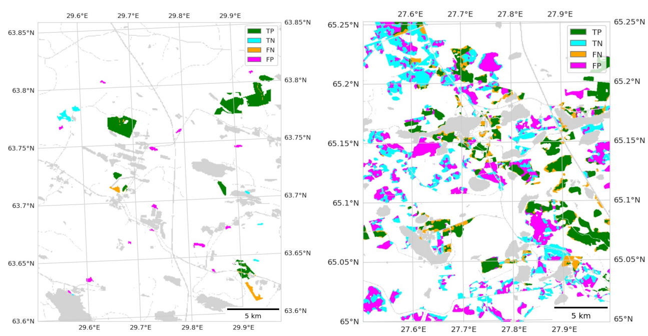

Fig. 2. Forest damage maps produced using DCVA: snow damage mapping using S1 images (left); windstorm damage mapping using S2 images (right).

In mapping windtrown forests, our initial results show 74.6% accuracy for identifying the damage through zero-shot learning using DCVA. The observations derived from Figure 3 suggest that snow damage assessment necessitates the utilization of more complex features compared to windstorm damage assessment. This inference is drawn from the finding that the higher layers of the Resnet-18 model, which capture intricate and abstract representations, prove to be more significant in the context of snow damage assessment. In contrast, windstorm damage assessment relies on the lower layers of the architecture, which likely capture more basic and fundamental features. These findings highlight the varying nature and requirements of assessing damage caused by different phenomena, emphasizing the need for tailored approaches and feature extraction techniques in remote sensing applications. Further research plans include incorporating S1 time series into the analysis and also aiming at combined use of S1 and S2 images for damage area assessment.

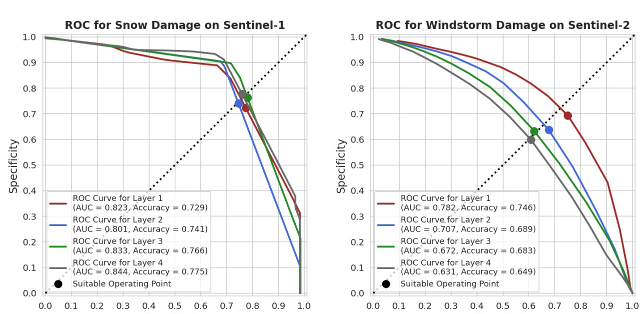

Fig. 3. Accuracy performance of DCVA using various layers of ResNet-18: for snow damage detection (left) and for windstorm damage detection (right).

3.2. Forest variable prediction using self-supervision and regression methods

The prediction accuracy of the MoCo-based network with combined S1 & S2 images (RMSE of 4.01 m and a R^2 of 0.65) was improved compared to baseline kNN regression and additionally tested MAML model when evaluated on an independent test set of forest plots. This was also the case when only S2 data were used in the model fine-tuning. Such performance can be attributed to the strong representations MoCo learns during the pretraining phase, which enhances the model's capacity to generalize to new tasks. Furthermore, adding Sentinel-1 had a limited impact on model performance and the main regression signal is coming from Sentinel-2. Achieved prediction accuracies were in line or better than in other reported studies in boreal forests using S1 and S2 (or Landsat) images and in-situ forest plot data [6]. Scatterplots illustrating performance for various models and image patch sizes are shown in Figure 4, and an example of the predicted forest height maps is shown in Figure 5.

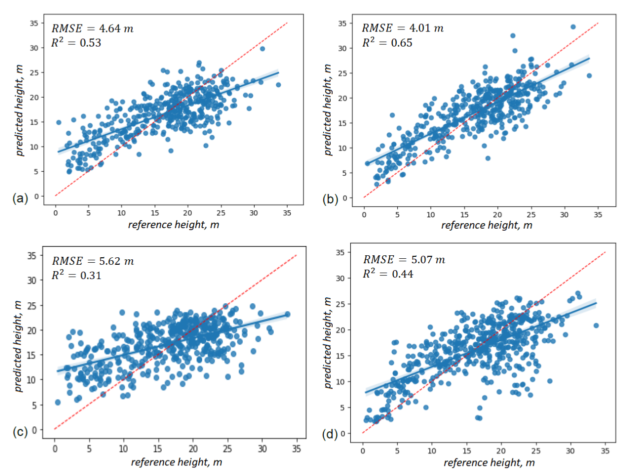

Fig. 4. Forest height prediction performance using self-supervision and baseline models: (a) and (b) show fine-tuned S1 and S2 MoCo model with 4px and 12 px image patches; (c) The 24x patch MAML model; (d) kNN regression method

3.3. Forest mapping using UNet+ and tranfer learning

In model pretraining with ALS data, achieved prediction accuracies were high with RMSE of 3.7m on pixel-level outperforming traditional forest mapping approaches, such as k-nearest neighbours, random forests or multiple linear regression (RMSE of 3.9-4.7m). Comparison of methodologies was done in scenario when traditional ML methods were trained with forest plots and EO data over the target site. Gain in accuracy was in range of 2-6 % units. Moreover, use of pretrained UNet+ model and further fine-tuning the model with plots demonstrated better stability when number of plots was

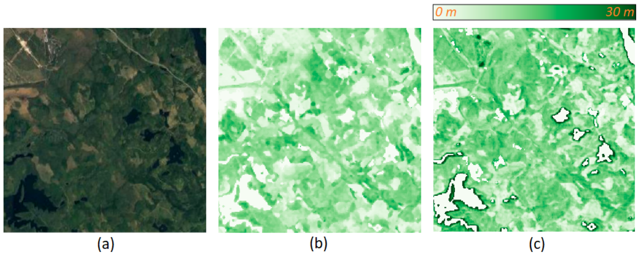

Fig. 5. MoCo based height prediction usig S1 & S2 images, $5km \times 5km$ area: (a) S2 image natural color composite, b) Lidar forest height, c) produced forest height map.

reduced to tens (instead of hundreds), or specific forest strata (tall trees or small trees) were completely missing. In such scenarios, kNN or random forests typically failed to provide meaningful predictions, while fine-tuned UNet+ models perfomed much better, provided that initial model pretraining was done using representative forest inventory data.

4. CONCLUSIONS

Our results indicate that semi-supervised DL methodologies enable the effective use of weak labels in EO image based semantic segmentation and regression tasks for several key forest management applications.

5. REFERENCES

[1] R.E. McRoberts and E.O. Tomppo, "Remote sensing support for national forest inventories," *Remote Sens. Env.*, vol. 110, no. 4, pp. 412–419, 2007.

[2] GFOI, *Integrating remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative*, Group on Earth Observations, Geneva, Switzerland, 2014.

[3] C. Persello, J. D. Wegner, R. Hänsch, D. Tuia, P. Ghamisi, M. Koeva, and G. Camps-Valls, "Deep Learning and Earth Observation to Support the Sustainable Development Goals: Current approaches, open challenges, and future opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 172–200, 2022.

[4] X. X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu, and R. Bamler, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, 2021.

[5] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," *IEEE Geosci. Remote Sens. Magazine*, vol. 5:4, pp. 8–36, 2017.

[6] H. Astola, L. Seitsonen, E. Halme, M. Molinier, and A. Lönnqvist, "Deep neural networks with transfer learning for forest variable estimation using Sentinel-2 imagery in boreal forest," *Remote Sens.*, vol. 13, 2021.

[7] S. Illarionova, A. Trekin, V. Ignatiev, and I. Oseledets, "Tree species mapping on sentinel-2 satellite imagery with weakly supervised classification and object-wise sampling," *Forests*, vol. 12, no. 10, 2021.

[8] S. Ge, H. Gu, W. Su, J. Praks, and O. Antropov, "Improved semisupervised UNet deep learning model for forest height mapping with satellite SAR and optical data," *IEEE JSTARS*, vol. 15, pp. 5776–5787, 2022.

[9] S. Ge, W. Su, H. Gu, Y. Rauste, J. Praks, and O. Antropov, "Improved LSTM model for boreal forest height mapping using Sentinel-1 time series," *Remote Sens.*, vol. 14, no. 21, 2022.

[10] M. Mõttus, P. Pham, E. Halme, M. Molinier, H. Cu, and J. Laaksonen, "Taiga: A novel dataset for multitask learning of continuous and categorical forest variables from hyperspectral imagery," *IEEE TGRS*, vol. 60, pp. 1–11, 2022.

[11] C. Bolyn, P. Lejeune, A. Michez, and N. Latte, "Mapping tree species proportions from satellite imagery using spectral–spatial deep learning," *Remote Sens. Env.*, vol. 280, 2022.

[12] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, "Weakly supervised deep learning for segmentation of remote sensing imagery," *Remote Sens.*, vol. 12, no. 2, 2020.

[13] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised Deep Change Vector Analysis for Multiple-Change Detection in VHR Images," *IEEE TGRS*, vol. 57, no. 6, pp. 3677–3693, 2019.

[14] E. Tomppo, O. Antropov, and J. Praks, "Boreal forest snow damage mapping using multi-temporal Sentinel-1 data," *Remote Sens.*, vol. 11, no. 4, 2019.

[15] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE CVPR*, 2020, pp. 9729–9738.

[16] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Sen12ms – a curated dataset of georeferenced multispectral sentinel-1/2 imagery for deep learning and data fusion," *ISPRS Ann.*, vol. IV-2/W7, pp. 153–160, 2019.

[17] M. Rußwurm, S. Wang, M. Körner, and D. Lobell, "Meta-learning for few-shot land cover classification," in *Proc. IEEE CVPR Workshops.*, 2020, pp. 788–796.