







Forest Disturbance Detection via Self-Supervised and Transfer Learning With Sentinel-1&2 Images

Rıdvan Salih Kuzu , Oleg Antropov , *Member, IEEE*, Matthieu Molinier , *Member, IEEE*, Corneliu Octavian Dumitru , Sudipan Saha , *Member, IEEE*, and Xiao Xiang Zhu , *Fellow, IEEE*

Abstract—In this study, we examine the potential of leveraging self-supervised learning (SSL) and transfer learning methodologies for forest disturbance mapping using Earth Observation (EO) data. Our focus is on natural disturbances caused by windthrow and snowload damages. Particularly, we investigate the potential of knowledge-distillation-based contrastive learning approaches to obtain comprehensive representations of forest structure changes using Copernicus Sentinel-1 and Sentinel-2 satellite imagery. Leveraging pretrained backbone models from knowledge distillation, we employ transfer learning based on deep change vector analysis to delineate forest changes. We demonstrate developed approaches on two use cases, namely, mapping windthrown forest using bitemporal Sentinel-1 and Sentinel-2 images, and mapping forest areas damaged by excessive snowload using bitemporal Sentinel-1 images. Developed self-supervised models were compared in a benchmarking exercise. The best results were provided by pixel-level contrastive learning for Sentinel-1-based snowload damage mapping with an overall accuracy of 84% and an F_1 score of 0.567, and for Sentinel-2-based forest windthrow mapping with an overall accuracy of 76.5% and an F_1 score of 0.692. We expect that developed methodologies can be useful for mapping also other types of forest disturbances using Copernicus Sentinel images or similar EO data. Our findings underscore the potential of SSL and transfer learning for enhancing forest disturbance detection using EO.

Index Terms—Boreal forest, change detection, Sentinel-1, Sentinel-2, windthrown forest, snowload damage, deep learning, self-supervised learning, transfer learning, contrastive learning.

Manuscript received 31 October 2023; revised 15 January 2024; accepted 28 January 2024. Date of publication 1 February 2024; date of current version 20 February 2024. This work was supported in part by the European Space Agency (ESA) through the RepreSent project under Grant 4000137253/22/I-DT, in part by the ESA Network of Resources Initiative with remote sensing data sponsorship under Grant 1B228D, and in part by the Helmholtz Association's Initiative and Networking Fund through Helmholtz AI under Grant ZT-I-PF-5-01 and on the HAICORE@FZJ partition. (*Corresponding author: Rıdvan Salih Kuzu.*)

Rıdvan Salih Kuzu and Corneliu Octavian Dumitru are with the Remote Sensing Technology Institute, German Aerospace Center, 82234 Weßling, Germany (e-mail: ridvan.kuzu@dlr.de; corneliu.dumitru@dlr.de).

Oleg Antropov and Matthieu Molinier are with the VTT Technical Research Centre of Finland Ltd., 02044 Espoo, Finland (e-mail: oleg.antropov@vtt.fi; matthieu.molinier@vtt.fi).

Sudipan Saha is with the Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: sudipan.saha@scai.iitd.ac.in).

Xiao Xiang Zhu is with the Chair of Data Science in Earth Observation, Technical University of Munich, 80333 Munich, Germany (e-mail: xiaoxiang.zhu@tum.de).

The source code of the study: <https://github.com/ridvansalihkuzu/representlib>
Digital Object Identifier 10.1109/JSTARS.2024.3361183

I. INTRODUCTION

THE sustainable management of forests requires timely and precise detection, mapping, and quantification of disturbances that impact forest ecosystems [1], [2], [3], [4]. These disturbances can arise from various sources, including natural events such as wildfires, insect outbreaks, and extreme weather conditions like windthrow and snowload damages [5]. Human activities, such as clear-cutting, selective logging, and land conversion, also contribute to these disruptions. Through careful analysis of these disturbances, forest managers and researchers can conduct timely updates in forest inventories, evaluate the health and resilience of forest ecosystems, pinpoint areas requiring intervention, and develop appropriate conservation strategies.

In this study, we concentrate on two types of natural disturbance resulting in the alteration of forest structure, specifically forest windthrow and forest snowload damages. These disturbances can be harder to detect compared to, e.g., forest clear-cutting, when homogeneous chunks of forest are completely removed, as, depending on the damage severity, the forest still remains in place, but the forest integrity and structure can be strongly disturbed. A key challenge in the assessment of forest disturbances is accurately quantifying the extent of the disturbances. Severe storms cause rapid and extensive damage within short time frames. Timely and accurate delineation of affected areas is needed for quickly guiding intervention activities.

The adoption of earth observation (EO) imagery into forest monitoring enabled timely assessment of the extent and severity of forest disturbances over wide areas [2]. Both synthetic aperture radar (SAR) and optical images were used for mapping forest windthrows, monitoring forest clear-cutting, and selective logging [6], [7], [8], [9]. Use case studies featured Landsat satellites, RADARSAT-2, Advanced Land Observing Satellite Phased-Array-type L-band Synthetic Aperture Radar (ALOS PALSAR), and many other sensors. Since the start of the Copernicus program and the launch of the first Sentinel-1 satellite in 2014 [10] and the first Sentinel-2 satellite in 2015 [11], these image sources were readily adopted by the research community and into operational processing chains as these images are provided at no cost to users.

Traditional approaches for identifying and quantifying forest disturbances have been largely guided by change detection methods. These techniques leverage the temporal dimension of remote sensing data, using radiometric or spectral differences

in multirate EO imagery to uncover disturbances. However, when applying these methods, particularly to bitemporal data, inherent challenges arise. Environmental factors, such as variations in forest transmissivity due to freezing conditions or moisture content changes, manifest themselves through apparent spectral/radiometric changes and can disrupt the results, thereby compromising the accuracy of these techniques.

Another important problem with traditional statistical or machine-learning-based change detection methods is the lack of extensive reference data to guide the change detection or classification models. Besides, these methods often rely on hand-engineered image features (such as image ratios, image differences, and gray-level co-occurrence features), which are frequently nontrivial to formulate, select, and accordingly justify their application. Presently, with the advent of deep learning (DL), actively used for semantic segmentation and change detection in remote sensing imagery (as detailed in Section II-B), such hand-engineered features may no longer be necessary. We expect that DL, in particular transfer learning and self-supervised learning (SSL), will be instrumental in avoiding conventional problems with image feature formulation, selection, and classification or change detection models.

In this study, we adopt a self-supervised strategy incorporating contrastive learning that utilizes knowledge distillation [12], and deep change vector analysis (DCVA) [13], a transfer learning framework specialized for bitemporal change detection. Our method employs the pretrained model obtained from contrastive learning into the DCVA to provide a robust approach for detecting areas of forest disturbances. Thus, this technique permits rapid analysis of forest disturbances without the explicit need for labeled data. To the best of our knowledge, this is the first demonstration of SSL and particular contrastive learning methods in EO-based mapping of forest disturbances caused by natural hazards such as windstorms or excessive snowload.

The rest of this article is organized as follows. Section II provides an overview of the relevant literature. Section III describes study areas, EO, and reference data used in the study. Section IV delineates our methodology. Section V summarizes the experimental framework. Section VI presents and discusses the obtained results. Finally, Section VII concludes this article.

II. LITERATURE OVERVIEW

Here, we first briefly describe prior research in forest disturbance mapping using EO data in the context of mapping forest windthrow- and snowload-damaged areas. Furthermore, we elaborate on DL literature related to EO-based change mapping, particularly recent developments in semi- and self-supervised DL methodologies, and their potential for forest mapping applications.

A. Forest Disturbance Mapping Using EO Data

To date, both optical and SAR satellite images have been used in mapping the extent of forest disturbances in connection with such natural hazards as windstorm and snowfall. Multispectral very high resolution (VHR) optical imagery has been used over boreal [14] and temperate forests [15] to map windthrown forests

with high accuracy. However, such data are usually proprietary. Several forest damage assessments were conducted using freely available high-resolution multispectral data, such as, e.g., Landsat images with 30-m pixel size. Combining Landsat images and forest statistics allowed to achieve up to 65% accuracy in detecting various forest disturbances over American coniferous and mixed deciduous forest in Utah [16], with a wide range of methods suitable for automated forest change detection utilizing Landsat time-series data [17]. Other examples using optical satellite data include assessment of forest areas disturbed by harvesting and strong windthrows in Western Siberia [18], wide area windthrow disturbance over temperate forest zone of European Russia, and the southern boreal forest zone of the United States with up to 75% accuracy [19], as well as pan-European mapping of forest windthrows [20] that included both natural and human-induced forest disturbance with overall accuracy of $92.5 \pm 2.1\%$ and omission error of 32.8%. Overall, subject to image availability over the area of interest, Landsat images [4], Sentinel-2 images [21], [22], or their combination [23] demonstrate a good potential in mapping forest-disturbed areas, the main difficulties being discerning between anthropogenic and natural disturbed forests during forest change detection, relatively modest accuracy with traditional change detection methods, and general image latency due to the cloud-obscured weather and partly absence of illumination during winter at high latitudes.

As a quick reaction to forest disturbances helps guide intervention activities and estimate the volume of economic losses, the potential of SAR imagery (not dependent on cloud situation unlike optical) was also explored. Historically, SAR-based forest disturbance mapping was largely utilizing bitemporal change detection techniques [24], [25], [26], [27], [28], [29], [30], [31] and time-series analysis [7], [32], [33]. These studies, mostly focused on detecting and mapping fire and logging-induced disturbances in forests, show that X -, C -, and L -bands are the most sensitive SAR wavelengths for detecting disturbances in forested landscapes and that sensitivity does not change across environments, but depends more on resolution and SAR imaging mode. In contrast, wind-induced disturbance has been less considered, with only a few studies utilizing SAR imagery reported [6], [34], [35]. Suitable sensors were ALOS PALSAR, RADARSAT-2, TerraSAR-X, and Sentinel-1.

Overall, prior research indicates that forest change detection is mostly performed through techniques relying on representative training data and using hand-engineered EO image features, which is a limitation when reference data are scarcely available. Also, spatial context is not fully taken into account, while primarily spectral or radiometric indices are used.

B. DL for EO-Based Change Detection

DL methods are widely adopted for various image classification and semantic segmentation tasks [36], [37], [38]. To date, several fully convolutional and recurrent neural networks have been demonstrated in forest remote sensing [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50]. These models often provide improved accuracy in the classification

or prediction of forest variables, as well as in forest change mapping. Recently, a modified U-Net model has been developed to detect windthrow damage in Bavarian forests [51]. Similarly, various U-Net models, including Siamese U-Nets, were tested in processing pairs of consecutive images for differentiating change events in temperate forest ecosystems of Ukraine [52] using supervised learning. Furthermore, vanilla U-Net was enhanced with residual and batch normalization layers to segment forest disturbances from windthrows and bark beetles in VHR satellite imagery, highlighting U-Nets' superiority over classical machine learning methods like random forests, support vector machines, and AdaBoost [53].

However, training of DL models often requires a fully segmented reference label, such as, e.g., airborne-laser-scanner-based forest attribute maps that are costly and not available over wide areas. In addition, reference data represented by external sources, expert annotations, or third-party reports cannot be always reliable and/or spatially sporadic or discontinuous. In this situation, earlier demonstrations of forest change detection are not fully relevant [14], [51], [54]. Instead, semisupervised, weakly supervised, transfer learning, or self-supervised methodologies are of key interest for easing the reliance on labeled data constraints, as successful approaches on semantic classification [55] and forest variable prediction [39], [49] demonstrated.

In the context of self-supervised EO-based change detection, one of the pioneering steps was pretraining a convolutional neural network (CNN)-based SSL model with pretext tasks on unlabeled data, followed by fine-tuning the model on a limited annotated dataset via supervised learning, and a change vector analysis to extract changed regions in urban landscapes [56]. Another study [57] utilized a pixelwise contrastive learning approach to improve the precision of flood change detection over benchmark methods that operate on a patch level. This was achieved using a pseudo-Siamese network for pixelwise representation alignment coupled with a loss function based on superpixel features. The threshold operator was used to generate a binary change map by evaluating the cosine similarity of feature vectors in bitemporal images. In a similar manner, Siamese-style contrastive learning for change detection combines image and domain knowledge contrastive losses during training and uses self-knowledge distillation from the teacher network during inference to enhance change detection accuracy [58]. Other noteworthy studies in the domain of EO-imagery-based land-cover change detection using self-supervised and unsupervised learning include using Siamese networks with local and global contrastive losses [59], introducing task-specific Siamese-style contrastive learning with hard sampling and smoothing [60], and extracting features from pretrained CNNs to generate change maps via clustering [61].

In another study [62], an autoencoder network was utilized, interpreting a high reconstruction loss in unsupervised learning as an indicator of forest disturbance, and a threshold was used for detecting disturbances using Tukey's method. Similarly, multiresolution deep feature maps derived by convolutional autoencoders can be used for detecting deforestation [63]. Other reported studies on the use of unsupervised approaches confirm that they can produce results comparable to or exceeding the

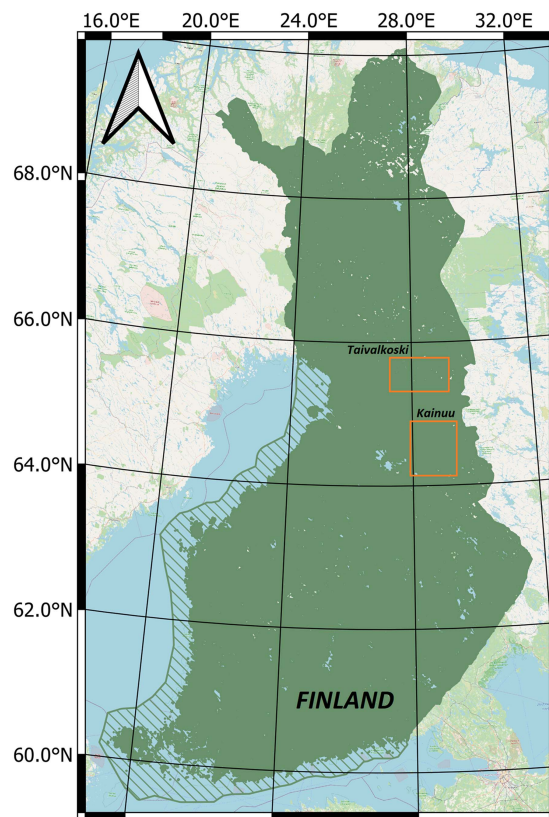


Fig. 1. Location of the Taivalkoski study site for forest windthrow mapping, and the Kainuu study site for snowload damage mapping in Finland.

supervised ones, e.g., [64] in forest fire damaged area mapping, or improve over traditional techniques when contrastive learning is added in the framework of semisupervised approaches, e.g., in forest height mapping [65].

As evident from the literature above, there has been a discernible trend in mapping of forests and forest disturbances from traditional machine learning methods toward DL, particularly focusing on U-Net variations. Nevertheless, U-Nets, or similar supervised models, typically require substantial amounts of labeled reference data that are either sparse or uncertain. In this situation, SSL and transfer learning represent an attractive way to overcome these limitations by leveraging unlabeled samples [12]. Inspired by recent advancements in land-cover change detection, which emphasize the need to reduce reliance on extensive training data and annotations, and building upon the promising outcomes of our initial efforts in forest disturbance mapping using DCVA [66], our study aims to explore the potential of various knowledge distillation techniques with transfer learning. Incorporating problem-agnostic SSL model weights enables swift response and analysis of postdisturbance scenarios, regardless of the forest disturbance type, reducing the need for customization.

III. DATA

Here, we describe our study areas, EO, and reference data available for both use cases: 1) mapping forest windthrow using

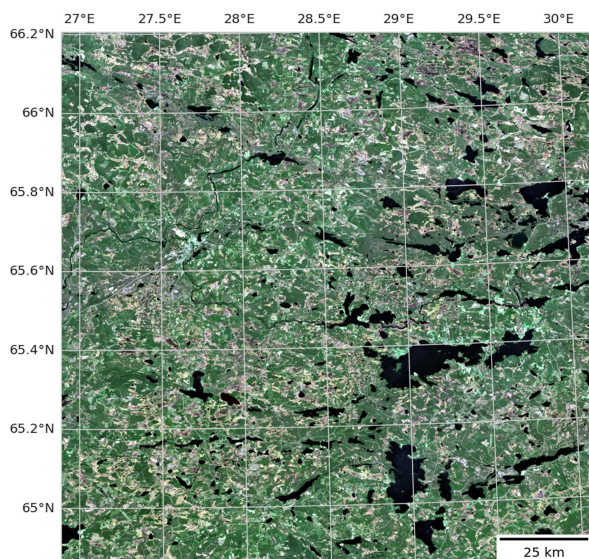


Fig. 2. Forest windthrow site: Sentinel-2 image (natural color composite) acquired on June 28, 2021 shortly after the windstorm.

Sentinel-1 and Sentinel-2 images and 2) mapping forest areas damaged due to excessive snowload using Sentinel-1 images. The study sites are shown schematically in Fig. 1.

A. Windstorm Damage Mapping

1) *Study Site and Damage Event Description*: The windstorm damage site was located in the vicinity of Taivalkoski municipality and Tyrjäjärvi lake, in the North Ostrobothnia province in northern Finland. The forests in the study site typically consist of a mix of coniferous and deciduous trees. Coniferous trees, such as pine, spruce, and fir, are prevalent in the region. On June 21, 2021, a major storm occurred in the area leading to major forest disturbances and forest windthrow. In this use case, we evaluated the potential of both Sentinel-2 and Sentinel-1 for accurately delineating the areas of the damaged forest using bi-temporal change detection scenarios.

2) *EO and Ground Reference Data*: European Space Agency (ESA) Sentinel-2 images were acquired both before the windstorm event, on June 4, 2021, and June 28, 2021, shortly after the forest windthrow. The postdisturbance Sentinel-2 image is shown in Fig. 2. The Level 2A surface reflectance product systematically generated by the ESA and distributed in tiles of $110 \times 110 \text{ km}^2$ was used in the data processing. The multispectral instrument on board Sentinel-2 satellites has 13 spectral bands with 10-m (four bands), 20-m (six bands), and 60-m (three bands) spatial resolutions. In our further analysis, we used RGB and NIR channels that were found most useful for monitoring boreal forests in prior studies [67] and were also available in pretrained backbone models [66], [68].

ESA Sentinel-1 images used in mapping forest windthrow were acquired on June 1, 2021, and July 7, 2021. The original dual-polarization Sentinel-1A images available as ground range detected (GRD) products were downloaded from the ESA Open Hub. GRD products were radiometrically terrain-flattened and

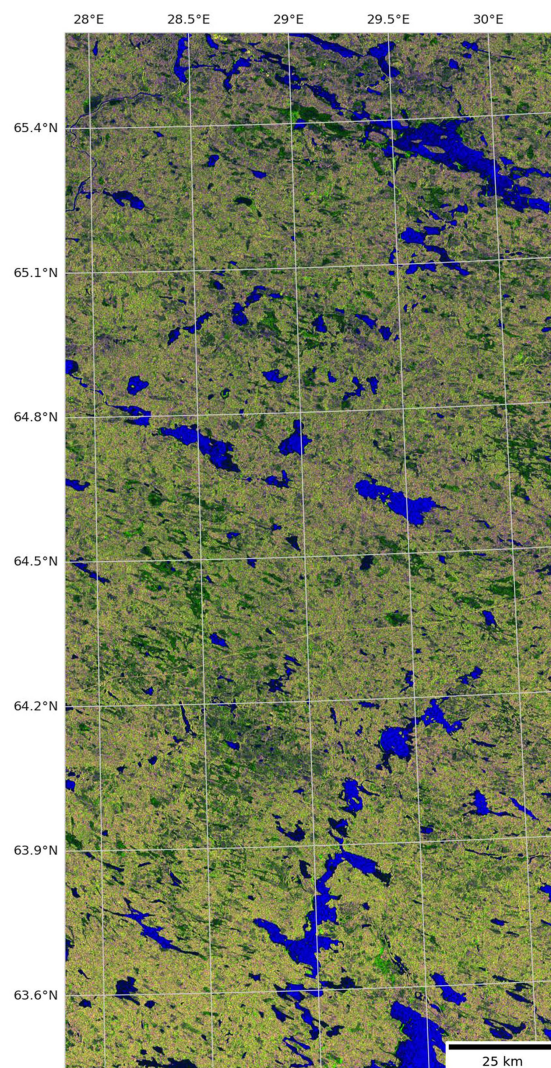


Fig. 3. Forest snowload damage site: Sentinel-1 image (VV in red, VH in green, and VV/VH in blue) after the snowload damages.

orthorectified with VTT in-house software using a local digital elevation model available from the National Land Survey of Finland [69]. Final preprocessed images were in gamma-naught format [70].

Information from the Finnish Forest Centre based on forest owner reporting was used as ground reference data for damaged stands. Reporting of forest owners was done on a forest compartment level, with reference data that can be considered as classic *weak labels*. Intact stands were derived using visual interpretation of optical images acquired after the event, as well as the latest update of multisource National Forest Inventory [71] systematically produced by the Natural Resource Institute of Finland.

B. Snowload Damage Mapping

1) *Study Site and Damage Event Description*: The study site was located in the Kainuu province of Finland and is also shown in Fig. 3. The study site represents a typical mixed

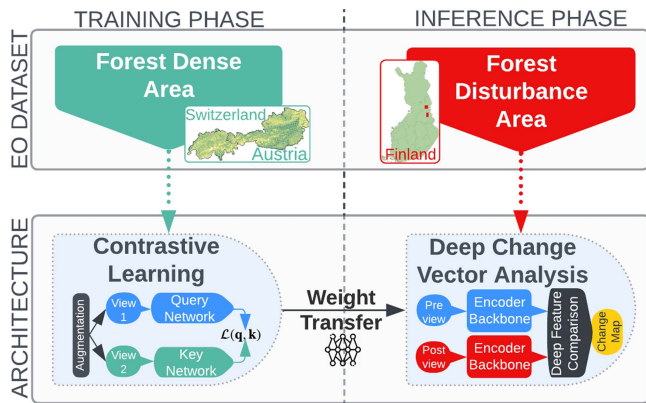


Fig. 4. General overview of contrastive learning (left). The specific contrastive learning methods based on knowledge distillation applied in this study (right).

boreal forestland. During the winter 2017–2018 season, severe damages were inflicted to the forests located within the study area by excessive snowload. The damage varied by extent and severity, with certain tree trunks either snapped or bent in an arc. Across broader areas, the damage manifested as broken branches and bent or crooked trees [7].

2) *EO and Ground Reference Data*: ESA Sentinel-1 images used in the study were acquired on November 12, 2017 (before any reported disturbances) and on March 24, 2018 (after all damages were inflicted) were used. Sentinel-1 images were preprocessed in the same fashion as described in Section III-A. No suitable Sentinel-2 data were available because of the illuminating conditions in the winter season.

Reference data were represented by sanitary cutting reports available from the Finnish Forest Centre. Reporting here is also performed on the forest compartment level. Importantly, some of the cut areas were not actually damaged but cut as a partly damaged compartment, representing another kind of *weak labels*.

IV. METHODOLOGY

Here, we outline the overall methodology in Section IV-A, describe in detail contrastive learning methodologies utilized within our training framework in Section IV-B, and elaborate on the DCVA approach implemented during the inference process in Section IV-C.

A. Overview of the Proposed Approach

Our approach, illustrated in Fig. 4, begins by training a contrastive learning model on the *Forest Dense Area* dataset, described in detail further in Section IV-A1. We employ a ResNet-18 [72] encoder as a backbone model for this purpose. Subsequently, we transfer the pretrained model weights obtained through contrastive learning to the DCVA. This transfer enables DCVA to utilize these weights for deriving deep feature vectors for both the *pre-event* (before the disturbance) and *post-event* (after the disturbance) views. By comparing these features, we generate a change magnitude vector. Applying a suitable threshold allows us to create a binary change map, clearly delineating disturbance and nondisturbance areas within our study area.

To conduct a comparative evaluation of the effectiveness of contrastive learning versus supervised learning, we additionally utilize the SEN12MS [68] pretrained weights of ResNet-18. In this scenario, we bypass the SSL training phase and directly introduce the supervised benchmark weights into DCVA in the inference phase. This also allows us to produce a change map using transfer learning.

1) *Training*: In this subsection, we outline our training process, emphasizing key steps such as dataset preparation, feature extraction, maintaining the encoder consistency, and encoder optimization for downstream tasks.

- 1) *Data acquisition*: We have curated separate training data referred to as the *Forest Dense Area* dataset as illustrated in Fig. 4, assembled from Sentinel-1 and Sentinel-2 patches accumulated over Austria and Switzerland, comprising 3417 and 1722 patches of size 256×256 , respectively, from the year 2018. Employing this dataset as input for contrastive learning allows for preventing potential bias as features are learned from a dataset completely independent of our study area.
- 2) *Augmentation*: In this phase, image patches are first duplicated to form two distinct views—*view 1* and *view 2*. Furthermore, a series of transformations is applied upon them, including random cropping, color distortion, and Gaussian blur with the purpose of dataset augmentation, to increase its diversity and enforce the model to learn invariant and robust features.
- 3) *Feature learning*: After augmentation, each view is independently channeled into two encoders, with each encoder featuring a ResNet-18 base encoder, a projection head, and optionally followed by a predictor or propagator. The outputs from these encoders undergo L_2 normalization, and a contrastive loss function is implemented based on the learning methodologies described in Section IV-B. This strategy encourages the model to minimize the distance between representations of identical augmented image patches (positive pairs) while simultaneously maximizing the distance between different image patches (negative pairs) within the embedding space.
- 4) *Consistency Maintenance Among Encoders*: To maintain consistency between the two encoders, a momentum update rule is employed, as illustrated in Fig. 5. This involves updating the weights of the key network with a moving average of the weights of the query network. This moving average is typically retained with a momentum value, aiding in stabilizing the learning and mitigating risks of mode collapse as observed in methods like *BYOL* [73], *Pix-Contrast*, or *PixPro* [74]. Alternatively, consistency can be preserved by keeping identical weights across the two encoders throughout the training process, as performed in Simple Siamese Representation Learning (*SimSiam*) [75].
- 5) *Post-Training Adjustments*: At the end of the training phase, all additional modules aside from the base encoder are removed. This refined base encoder is then deployed for downstream tasks such as forest disturbance delineation. As an independent feature extractor, the encoder can either function directly or undergo further fine-tuning

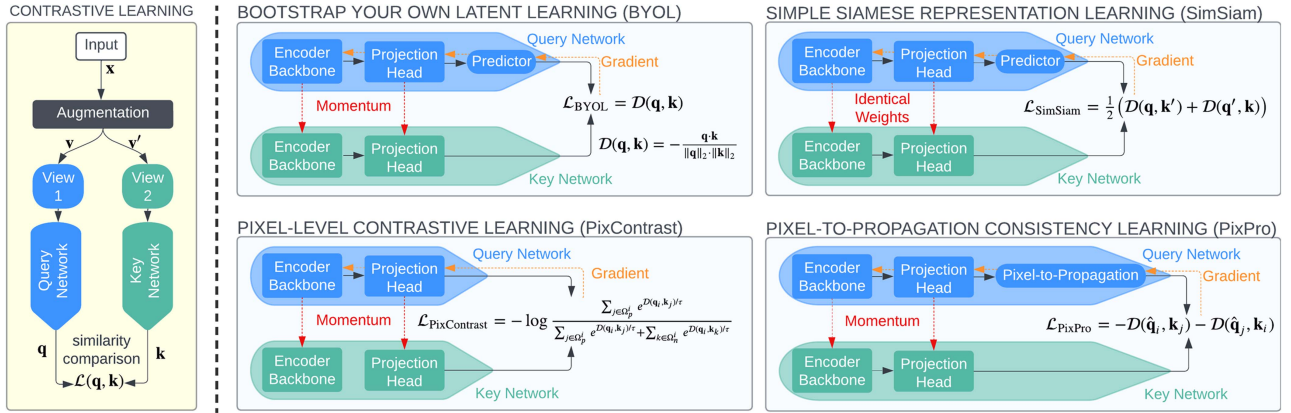


Fig. 5. General overview of contrastive learning (left). The specific contrastive learning methods based on knowledge distillation applied in this study (right).

for different tasks. Further insights on the application of this feature extractor in our DCVA methodology are elaborated in the next section.

2) *Inference*: In this subsection, we explore the inference phase of our process, where the trained encoder, tailored during the training phase, is utilized to execute downstream tasks. This includes the application of the DCVA approach, using the feature representations learned by the encoder to effectively identify and interpret forest disturbances, as detailed below.

- 1) *Data preprocessing*: At this stage, pre-event and post-event images are subjected to radiometric normalization to counteract the distortions caused by the atmosphere and other physical occurrences [76]. Moreover, the images are further processed to correct geometric inconsistencies due to varying satellite sensor viewing angles and any misalignments arising from terrain impact. Following these adjustments, images can undergo orthorectification, pansharpening, and coregistration [77], [78].
- 2) *Deep feature extraction and comparison*: Here, bitemporal deep features are derived by feeding the preprocessed images individually into a pretrained model and then extracting features from certain model layers. Initial layers of the model capture basic visual concepts, while deeper layers apprehend more complex concepts by integrating lower level features. A key challenge here is to balance the choice of layers from which features are extracted [13]. Finally, the deep feature hypervector is constructed by pooling features from multiple layers of the model, thereby creating a multiscale representation, as elaborated in Section IV-C.
- 3) *Binary change detection*: This stage seeks to differentiate between unchanged and altered pixels on the premise that unchanged pixels produce similar deep features, whereas altered pixels generate dissimilar deep features. To ensure a thorough comparison, a deep magnitude is computed for each pixel. Pixels are then categorized into two groups using a decision boundary or cutoff value to produce two sets for altered and unchanged pixels. By the conclusion of this phase, we possess the change map that illustrates regions of forest disturbance and nondisturbance, as represented in Fig. 4.

TABLE I
COMPARISON OF KEY CHARACTERISTICS IN THE CONTRASTIVE LEARNING METHODS UTILIZED

	Feature Level	Consistence Maintenance	Contrastive Mechanism
BYOL	Instance	Momentum based	Cosine distance based, nonsymmetrical
SimSiam	Instance	Identical weights	Cosine distance based, symmetrical
PixContrast	Pixel	Momentum based	Cross entropy based, nonsymmetrical
PixPro	Pixel	Momentum based	Cosine distance based, nonsymmetrical

B. Contrastive Learning Approaches Employed in Training

In our empirical studies, we assessed four distinct contrastive learning methods, namely Bootstrap Your Own Latent (BYOL) [73], SimSiam [75], Pixel-level Contrastive Learning (PixContrast), and Pixel-to-Propagation Consistency Learning (PixPro) [74]. The selection was intended to examine the potential effects of instance-level approaches (e.g., BYOL and SimSiam) compared to pixel-level ones (e.g., PixPro and PixContrast) on disturbance mapping. This aim stems from recent contrastive learning research indicating that pixel-level methods can enhance feature discrimination and understanding [79], proving useful for tasks like object detection and semantic segmentation [80]. Besides, it can provide robust learning against diverse contexts and inconsistencies, crucial for accuracy in difficult conditions. Furthermore, it seems particularly effective for complex aerial and satellite imagery, addressing challenges like large footprints and variable acquisition conditions [81]. An additional objective of this examination was to ascertain potential improvements that might be gained by using model weights that have been extracted from these contrastive learning methods, as opposed to those derived from supervised learning.

Table I provides an overview of the main characteristics of the contrastive learning methods used in this study, with subsequent detailed explanations for each method given as follows.

1) *Bootstrap Your Own Latent*: In BYOL [73], an image and its augmented view are considered a positive pair, and it optimizes the similarity between the representations of this pair without the need for negative pairs. Let us denote an image as x and two augmented versions of this image as v and v' . Also, let us denote the query (online) network as Q with parameters θ and the key (target) network as K with parameters ξ . The online

network Q generates a query from v and the target network K generates a key from v' as

$$\mathbf{q} = Q(v; \theta), \quad \mathbf{k} = K(v'; \xi). \quad (1)$$

During training, the parameters ξ of the key network are updated using a moving average (momentum) of the parameters θ of the query network, as also illustrated in Fig. 5. The objective of BYOL is to minimize the mean squared difference between the normalized query and key vectors. The loss function of BYOL is, therefore, given by

$$\mathcal{L}_{\text{BYOL}}(\theta, \xi) = \mathcal{D}(\mathbf{q}, \mathbf{k}), \quad \text{where } \mathcal{D}(\mathbf{q}, \mathbf{k}) = -\frac{\mathbf{q} \cdot \mathbf{k}}{\|\mathbf{q}\|_2 \cdot \|\mathbf{k}\|_2}. \quad (2)$$

This loss function encourages the online network to learn representations that maximize the similarity between the query and key vectors, thus enabling the model to learn robust and invariant features from the input data.

2) *Simple Siamese Representation Learning*: SimSiam [75] is designed to overcome the necessity of negative samples by utilizing two identical networks that maximize the similarity between two differently augmented views of the same image. This is achieved by encoding the images into feature vectors and then transforming them into query (prediction) and key (target) vectors, which are then made similar to each other. Given an image x , we create two augmented views, denoted as v and v' . These views are then fed into two identical networks, f_θ (the backbone encoder and the projection head), and h_θ (the predictor), with parameters θ . These networks generate the following query and key vectors:

$$\mathbf{q} = h_\theta(f_\theta(v)), \quad \mathbf{k} = f_\theta(v), \quad \mathbf{q}' = h_\theta(f_\theta(v')), \quad \mathbf{k}' = f_\theta(v'). \quad (3)$$

In this configuration, the backbone encoder and the projection head of the key network are identical to those of query network but with their parameters θ detached from the computational graph to prevent gradients from flowing back, as illustrated in Fig. 5. Thus, SimSiam is aimed at maximizing the similarity of the query and key vectors, which is achieved by minimizing the following contrastive loss:

$$\mathcal{L}_{\text{SimSiam}}(\theta) = \frac{1}{2} (\mathcal{D}(\mathbf{q}, \mathbf{k}') + \mathcal{D}(\mathbf{q}', \mathbf{k})). \quad (4)$$

In other words, the goal of this function is to minimize the distance between the prediction and target vectors, thereby encouraging the encoder to derive significant representations.

3) *Pixel-Level Contrastive Learning*: PixContrast [74] postulates that pixel-level discrimination can yield powerful visual representations that are applicable to a wide range of downstream tasks. Similar to conventional contrastive learning approaches, PixContrast initiates the process by generating two augmented views from the same image. These views are resized to a fixed resolution and processed through a query network Q and a key network K to compute image features. These networks comprise a backbone network and a projection head network. The latter includes two sequential 1×1 convolution layers that produce feature maps of a specific spatial resolution and use the learned backbone representations for feature transfer.

PixContrast, in contrast to conventional approaches, generates a feature map rather than a single image feature vector for each view, thereby enabling tasks at the pixel level. In this approach, each pixel in the feature map is first mapped back to the original image space, and the distances between all pixel pairs from the two feature maps are computed and normalized, thus forming positive and negative pairs based on a set threshold. Consequently, the contrastive loss function for PixContrast is articulated as

$$\mathcal{L}_{\text{PixContrast}} = -\log \frac{\sum_{j \in \Omega_p^i} e^{\mathcal{D}(\mathbf{q}_i, \mathbf{k}_j)/\tau}}{\sum_{j \in \Omega_p^i} e^{\mathcal{D}(\mathbf{q}_i, \mathbf{k}_j)/\tau} + \sum_{k \in \Omega_n^i} e^{\mathcal{D}(\mathbf{q}_i, \mathbf{k}_k)/\tau}}. \quad (5)$$

Here, \mathbf{q}_i and \mathbf{k}_i denote the pixel feature vectors in the two views, and Ω_p^i and Ω_n^i signify the sets of pixels identified as positive and negative, respectively, in relation to pixel i from the two views. The parameter τ is a scalar temperature hyperparameter. This loss is then averaged over all pixels on the first view that overlaps with the second view. A comparable calculation is performed for a pixel j on the second view, and the result is averaged.

4) *Pixel-to-Propagation Consistency Learning*: PixPro [74] combines spatial sensitivity and smoothness, in contrast to PixContrast, which concentrates solely on spatial sensitivity, with the goal of improving transfer performance in tasks that require dense prediction. Spatial sensitivity facilitates the differentiation of nearby pixels, essential for precise predictions in boundary regions where labels vary. Meanwhile, smoothness promotes similarity among adjacent pixels, assisting predictions within same-label areas. In PixPro, the pretext task consists of two key components: The first is a pixel propagation module that filters a pixel's features by propagating similar pixel features, thus offering a denoising/smoothing effect and more coherent pixel-level predictions. The second part is an asymmetric query and key architectures; one branch produces a standard feature map, while the other includes the pixel propagation module, as seen in Fig. 5. The task aims for feature consistency across both branches, ignoring negative pairs. For each pixel feature \mathbf{q}_i , the pixel propagation module calculates its smoothed transform $\hat{\mathbf{q}}_i$ by propagating features from all pixels \mathbf{q}_j within the same image Ω as follows:

$$\hat{\mathbf{q}}_i = \sum_{j \in \Omega} \max(\mathcal{D}(\mathbf{q}_i, \mathbf{q}_j), 0)^\gamma \cdot \mathcal{G}(\mathbf{q}_j) \quad (6)$$

where γ serves as an exponent to fine-tune the sharpness of the similarity function and \mathcal{G} represents a transform function, which can be implemented using linear layers with batch normalization and ReLU activation in between consecutive layers. Consequently, the loss function is expressed as

$$\mathcal{L}_{\text{PixPro}} = -\mathcal{D}(\hat{\mathbf{q}}_i, \mathbf{k}_j) - \mathcal{D}(\hat{\mathbf{q}}_j, \mathbf{k}_i) \quad (7)$$

where i and j stand for corresponding pixel pairs across two augmented views, and $\hat{\mathbf{q}}_i$ and \mathbf{k}_i represent pixel feature vectors yielded by the query Q and key K networks, respectively. Finally, the loss is computed by averaging across all the positive pairs in every image.

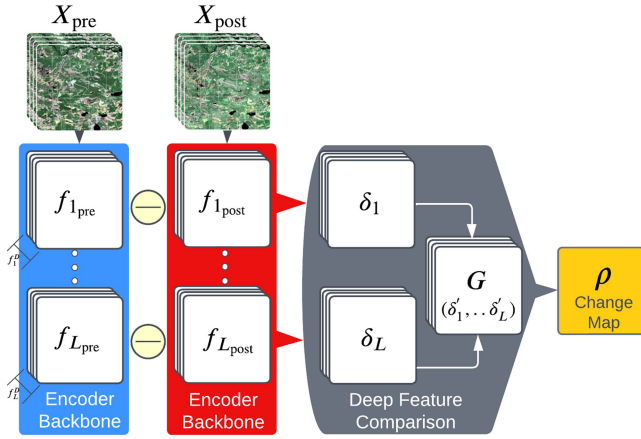


Fig. 6. Visual flow for change map generation from bitemporal observations.

C. DCVA Employed in Inference

In conducting inference for the identification of forest disturbances, we utilize DCVA, a method devised for recognizing alterations in imagery captured before and after a specific event [13]. Essentially, DCVA utilizes a backbone encoder to extract deep features from the images without training the encoder (an instance of transfer learning) and then contrasts these features to ascertain changes, segregating pixels into either changed or unchanged categories.

Aligned with this methodology, our contrastive learning scheme, elucidated in Section IV-B and depicted in Fig. 5, provides a pretrained encoder backbone (denoted as f) to DCVA. Consequently, after preprocessing, bitemporal images X_{pre} and X_{post} are channeled through f . Features from designated layers L are harvested, denoted as $f_{l_{pre}}$ and $f_{l_{post}}$ for each layer $l \in L$. The difference vector for each layer is computed as

$$\delta_l = f_{l_{post}} - f_{l_{pre}}. \quad (8)$$

A deep change vector G is synthesized by concatenating selected features from each layer

$$G = (\delta'_1, \dots, \delta'_l, \dots, \delta'_L) \quad (9)$$

where δ'_l is a subset of δ_l chosen based on variance indicating sensitivity to change information. Thus, the deep magnitude ρ for each pixel (r, c) is computed as

$$\rho(r, c) = \sqrt{\sum_{d=1}^D (g_d)^2} \quad (10)$$

where D represents the total dimension of vector G and g_d is the d th component of G . Finally, a threshold \mathcal{T} is employed on ρ to classify pixels into changed (ω_c) and unchanged (ω_{nc}) classes

$$\rho(r, c) \in \begin{cases} \omega_{nc}, & \text{if } \rho(r, c) < \mathcal{T} \\ \omega_c, & \text{otherwise.} \end{cases} \quad (11)$$

Fig. 6 also summarizes the steps for generating the disturbance (change) map for forests via transfer learning using the pretrained weights of the encoder obtained after contrastive learning.

V. EXPERIMENTAL FRAMEWORK

This part of the document outlines the methods used to experiment with various learning techniques and measurements to confirm the effectiveness of our approach.

A. Model Configuration and Pretraining

In this research, the DCVA employs ResNet-18 [72] for its baseline encoder architecture. In addition, the initial weights of our baseline model were derived from supervised pretraining on the SEN12MS dataset [68], encompassing 541 986 patches each of dimensions 256×256 .

Given that ResNet-18 comprises four principal convolutional blocks (f_l , where $1 \leq l \leq 4$) prior to the fully connected layers, our approach focused on deriving the change vector G from one of these blocks. Consequently, our experimental design was divided into two key objectives.

- 1) *Layer suitability for disturbance types*: The task involved pinpointing the ideal f_l to extract the change vector G . Given that the preliminary layers of a CNN predominantly identify basic attributes like lines and curves, and the subsequent layers apprehend complex patterns and textures, a crucial aspect was to gauge the relevance of using these layers concerning distinct forest disturbance scenarios.
- 2) *Comparative analysis of SSL encoders*: Following the determination of the optimal f_l , a comparison was made between the baseline DCVA encoder and those integrating pretrained SSL models, with the objective of identifying the most efficient SSL methodology for the creation of disturbance change maps.

B. Network Initialization and Optimization

In this study, SSL frameworks were constructed using ResNet-18 [72] as the encoder backbone. The AdamW optimizer [82] was utilized, with its coefficients β_1 and β_2 set to 0.9 and 0.98, respectively, for calculating running averages. The initial learning rate was defined within the range [0.0005 – 0.01]. When a plateau was detected in the training process, the learning rate was halved. Simultaneously, the weight decay parameter λ was fixed at $5e - 6$. Training was carried out in batches of 32, with a cap of 15 epochs. Hyperparameter tuning was performed using a Bayesian optimization methodology.

C. Software Specifications

The research experiments were conducted using the PYTORCH framework, version 1.10 [83]. Computational processing was facilitated by Python libraries such as NUMPY, SCIPY, PANDAS, and SCIKIT-LEARN. For Bayesian hyperparameter tuning, OPTUNA was employed. To guarantee consistent replication of the machine learning workflow, all software dependencies were encapsulated within a DOCKER container, the details of which are available in the project's source code repository.

D. Evaluation Metrics

To evaluate the efficacy of our approach in mapping forest disturbances, we employed the criteria outlined as follows. In this context, disturbed areas (D) are considered the positive class, while intact areas (I) are deemed the negative class. The accuracy for each class is calculated as follows:

$$\text{Acc}_D = \frac{T_P}{T_P + F_N} \quad (\text{also known as Sensitivity}) \quad (12)$$

$$\text{Acc}_I = \frac{T_N}{T_N + F_P} \quad (\text{also known as Specificity}). \quad (13)$$

Here, T_P and T_N represent the correctly classified disturbed and intact areas, respectively. F_N denotes disturbed areas mistakenly labeled as intact, while F_P signifies intact areas incorrectly classified as disturbed. The overall accuracy is then expressed as

$$\text{Acc}_O = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}. \quad (14)$$

Furthermore, the F_1 score is computed as

$$F_1 = 2 \times \frac{P \times R}{P + R}, \quad P = \frac{T_P}{T_P + F_P}, \quad R = \frac{T_P}{T_P + F_N} \quad (15)$$

where P and R stand for *precision* and *recall*, respectively. For the area under the curve (AUC), the formula is

$$\text{AUC} = \frac{1}{2} \sum_{i=1}^{n-1} (\text{Acc}_{D_i} + \text{Acc}_{D_{i+1}}) \times ((1 - \text{Acc}_{I_{i+1}}) - (1 - \text{Acc}_{I_i})) \quad (16)$$

where n denotes the number of data points utilized in constructing the receiver operating characteristic (ROC) curve, with the threshold \mathcal{T} being increased in equal increments.

We also utilized Cohen's kappa κ statistic as a metric to gauge the level of agreement beyond what might occur randomly [84]. Given a $k \times k$ confusion matrix with elements represented as f_{ij} , we compute

$$r_i = \sum_{j=1}^k f_{ij} \quad \forall i, \quad \text{and} \quad c_j = \sum_{i=1}^k f_{ij} \quad \forall j \quad (17)$$

$$P_e = \frac{1}{N^2} \sum_{i=1}^k r_i c_i \quad (18)$$

$$P_o = \frac{1}{N} \sum_{j=1}^k f_{jj} \quad (19)$$

where the variables r_i and c_j represent the totals of rows and columns for classes i and j , respectively. P_e indicates the expected agreement that would occur by chance, while P_o depicts the observed agreement (effectively the overall accuracy). Thus, the κ metric is given by

$$\kappa = \frac{P_o - P_e}{1 - P_e}. \quad (20)$$

TABLE II

LAYERWISE PERFORMANCE OF THE BASELINE ENCODER IN EXTRACTING THE SNOWLOAD DAMAGE MAP USING SENTINEL-1

	Precision	Sensitivity (Acc _D)	Specificity (Acc _I)	Accuracy (Acc _O)	Cohen's Kappa (κ)	F_1	AUC
Layer-1 (f_1)	0.223	0.733	0.708	0.711	0.219	0.407	0.800
Layer-2 (f_2)	0.25	0.745	0.744	0.744	0.261	0.478	0.804
Layer-3 (f_3)	0.277	0.729	0.782	0.777	0.297	0.484	0.838
Layer-4 (f_4)	0.279	0.759	0.776	0.774	0.303	0.483	0.844

Best performing ones are shown bold.

TABLE III

LAYERWISE PERFORMANCE OF THE BASELINE ENCODER IN EXTRACTING THE WINDTHROW DAMAGE MAP USING SENTINEL-1

	Precision	Sensitivity (Acc _D)	Specificity (Acc _I)	Accuracy (Acc _O)	Cohen's Kappa (κ)	F_1	AUC
Layer-1 (f_1)	0.511	0.672	0.693	0.686	0.337	0.566	0.769
Layer-2 (f_2)	0.496	0.688	0.667	0.674	0.322	0.566	0.770
Layer-3 (f_3)	0.513	0.698	0.684	0.688	0.349	0.586	0.784
Layer-4 (f_4)	0.579	0.726	0.748	0.741	0.444	0.647	0.833

Best performing ones are shown bold.

κ falls in the range $[0, 1]$, providing insights into the level of agreement: poor ($0.0 \leq \kappa < 0.2$), fair ($0.2 \leq \kappa < 0.4$), moderate ($0.4 \leq \kappa < 0.6$), good ($0.6 \leq \kappa < 0.8$), or very good ($0.8 \leq \kappa \leq 1.0$).

VI. RESULTS AND DISCUSSION

Here, we start the presentation of our results with an ablation analysis of the proposed methodology, as detailed in Section VI-A. A comparative evaluation between the baseline and the experimented SSL schemes is presented in Section VI-B. A qualitative assessment is conducted in Section VI-C. Findings drawn from our results are compared to earlier reported scholarly contributions in Section VI-D.

A. Ablation Study for the Baseline Approach

In this section, we analyze the impact of the backbone attributes and type of imaging sensor on the forest disturbance mapping over our study areas.

Table II presents the layerwise performance of the baseline encoder for detecting snowload damages using Sentinel-1. It is evident that as we progress from f_1 to f_4 , there is a noticeable improvement in the majority of accuracy metrics. Notably, Layer-4 (f_4) exhibits the best disturbance detection performance according to most of the criteria.

Table III details the layerwise performance of the baseline encoder in discerning windthrow damages using Sentinel-1. Similar to the snowload case, there is a clear trend of improved performance from the initial layers to the final layer, with f_4 standing out as the most effective. On the other hand, Table IV highlights the baseline encoder's layerwise performance for windthrow damages using Sentinel-2 data. Unlike prior cases, the third layer (f_3) outperforms across metrics, potentially due to the transition from Sentinel-1 to Sentinel-2.

Nevertheless, when only considering windthrow mapping, Sentinel-1 exhibits superior performance, achieving an F_1 score of 0.647 as opposed to 0.609 for Sentinel-2 as highlighted at the

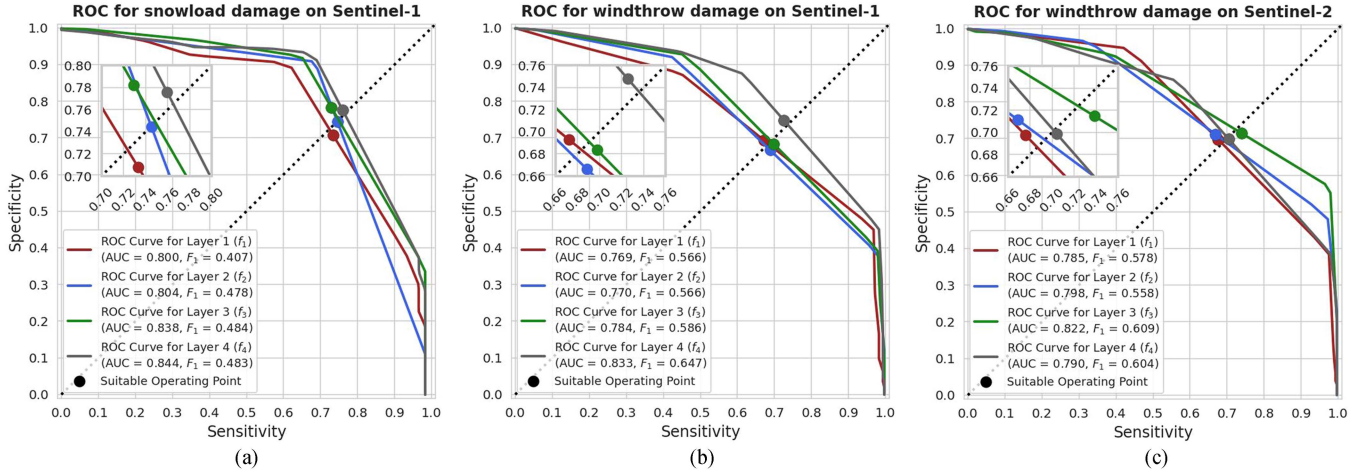


Fig. 7. ROCs showing layerwise performance of the baseline encoder for different forest disturbance maps and sensors. (a) Snowload damage map using Sentinel-1. (b) Windthrow damage map using Sentinel-1. (c) Windthrow damage map using Sentinel-2.

TABLE IV
LAYERWISE PERFORMANCE OF THE BASELINE ENCODER IN EXTRACTING THE WINDTHROW DAMAGE MAP USING SENTINEL-2

	Precision	Sensitivity (Acc _D)	Specificity (Acc _I)	Accuracy (Acc _O)	Cohen's Kappa (κ)	F_1	AUC
Layer-1 (f_1)	0.516	0.676	0.698	0.691	0.346	0.578	0.785
Layer-2 (f_2)	0.526	0.669	0.712	0.698	0.355	0.558	0.798
Layer-3 (f_3)	0.553	0.738	0.715	0.722	0.417	0.609	0.822
Layer-4 (f_4)	0.527	0.704	0.699	0.700	0.370	0.604	0.790

Best performing ones are shown bold.

suitable operating points where Acc_D (sensitivity) and Acc_I (specificity) meet on the ROC curves, as illustrated in Fig. 7.

Upon examining the baseline DCVA encoder's ability to detect windthrow and snowload disturbances using Sentinel-1, as presented in Tables II and III, our method demonstrates a higher performance for windthrow mapping. This is underscored by the notable κ and F_1 scores, primarily due to the better precision of the baseline in pinpointing windthrow.

B. Benchmark Comparisons

Our proposed method leverages the transfer of pretrained model weights into the DCVA encoder to produce forest disturbance maps. To compare the detection accuracy of SSL-based backbone encoders versus the baseline approach, we additionally conducted a benchmarking testing using a range of SSL methodologies described in Section IV-B, including PixPro, PixContrast, SimSiam, and BYOL.

In the context of Sentinel-1-based snowload and windthrow mapping, as summarized in Tables V and VI, DCVA encoders employing pixel-level SSL techniques demonstrate greater detection accuracy compared to those utilizing instance-level SSL techniques. Specifically, PixContrast demonstrates superior performance in snowload disturbance mapping, while PixPro appears more suitable in delineating forest areas damaged by windstorm. The corresponding ROC curves shown in Fig. 8(a) and (b), underscore their potential in discerning forest disturbances,

TABLE V
COMPARING THE BASELINE WITH SSL-BASED DCVA ENCODERS IN EXTRACTING THE SNOWLOAD DAMAGE MAP USING SENTINEL-1

	Precision	Sensitivity (Acc _D)	Specificity (Acc _I)	Accuracy (Acc _O)	Cohen's Kappa (κ)	F_1	AUC
Baseline DCVA	0.279	0.759	0.776	0.774	0.303	0.483	0.844
PixPro	0.356	0.791	0.837	0.832	0.407	0.543	0.868
PixContrast	0.370	0.802	0.844	0.840	0.426	0.567	0.875
SimSiam	0.284	0.821	0.763	0.769	0.318	0.505	0.851
BYOL	0.273	0.792	0.759	0.763	0.300	0.477	0.846

Best performing ones are shown bold.

TABLE VI
COMPARING THE BASELINE WITH SSL-BASED DCVA ENCODERS IN EXTRACTING THE WINDTHROW DAMAGE MAP USING SENTINEL-1

	Precision	Sensitivity (Acc _D)	Specificity (Acc _I)	Accuracy (Acc _O)	Cohen's Kappa (κ)	F_1	AUC
Baseline DCVA	0.579	0.726	0.748	0.741	0.444	0.647	0.833
PixPro	0.605	0.763	0.762	0.763	0.492	0.689	0.835
PixContrast	0.578	0.741	0.742	0.742	0.450	0.650	0.842
SimSiam	0.547	0.732	0.711	0.718	0.407	0.612	0.824
BYOL	0.575	0.740	0.738	0.739	0.445	0.662	0.819

Best performing ones are shown bold.

TABLE VII
COMPARING THE BASELINE WITH SSL-BASED DCVA ENCODERS IN EXTRACTING THE WINDTHROW DAMAGE MAP USING SENTINEL-2

	Precision	Sensitivity (Acc _D)	Specificity (Acc _I)	Accuracy (Acc _O)	Cohen's Kappa (κ)	F_1	AUC
Baseline DCVA	0.553	0.738	0.715	0.722	0.417	0.609	0.822
PixPro	0.606	0.742	0.770	0.761	0.483	0.679	0.860
PixContrast	0.608	0.768	0.763	0.765	0.497	0.692	0.840
SimSiam	0.546	0.686	0.728	0.714	0.388	0.582	0.813
BYOL	0.571	0.727	0.739	0.735	0.435	0.635	0.843

Best performing ones are shown bold.

clearly surpassing the classification accuracy of baseline DCVA encoder.

Delving into a comparative study between Sentinel-1 and Sentinel-2 for windthrow damage mapping, pixel-level SSL algorithms consistently outpace their counterparts, as visualized in Fig. 8(b) and (c). Nevertheless, while Sentinel-1 surpasses

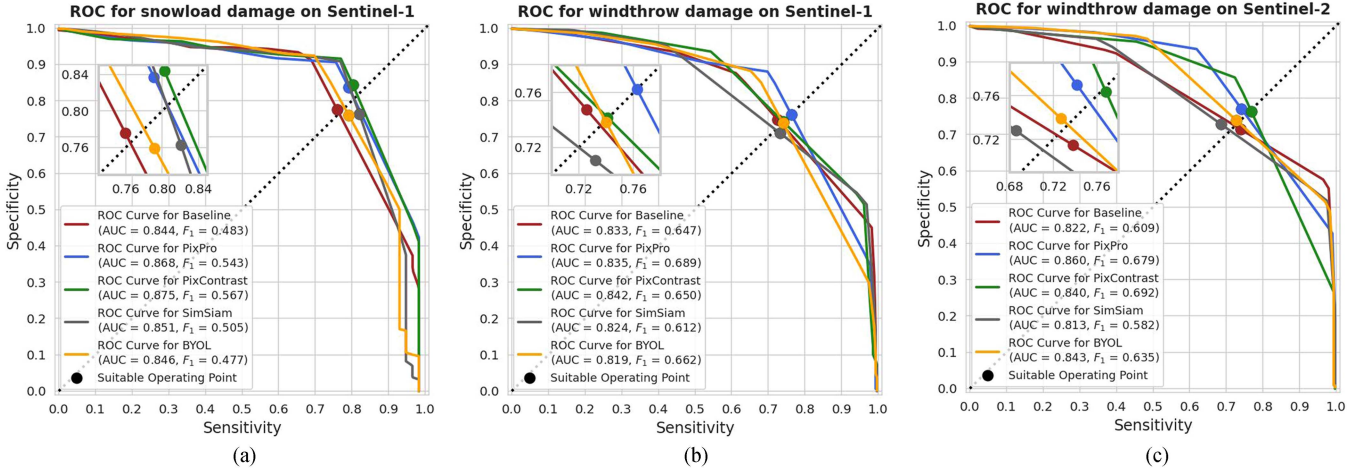


Fig. 8. ROCs in comparing the baseline with SSL-based DCVA encoders for different forest disturbance maps and sensors. (a) Snowload damage map using Sentinel-1. (b) Windthrow damage map using Sentinel-1. (c) Windthrow damage map using Sentinel-2.

Sentinel-2 in the baseline DCVA with accuracy scores of 0.741 and 0.722, respectively, SSL integration in DCVA realizes a more pronounced enhancement with Sentinel-2, evidenced by an accuracy increment of 0.042, in contrast to a 0.022 enhancement in Sentinel-1, as compared in Table VI and VII.

It is worth mentioning that the observed disparities in performance for disturbance detection when using Sentinel-1 and Sentinel-2 data arise from different imaging nature of these sensors and the consequent modifications in the model architecture. Sentinel-1, utilizing SAR, is particularly effective at identifying structural changes in forests due to its sensitivity to volumetric structure and moisture content. On the other hand, Sentinel-2, through multispectral imaging, excels at discerning changes in vegetation health by capturing comprehensive spectral data from the Earth’s surface. The requirement to tailor the input layers of our DCVA model to suit the different channel counts of these sensors also influences the model weights, resulting in varied interpretations and reflections of forest disturbances in the data. Recognizing these sensor-specific strengths is crucial for the precise analysis of forest disturbances, highlighting the importance of using SAR and multispectral data interchangeably in environmental monitoring.

In summary, using SSLs for DCVA generally provides an improvement in terms of detection accuracy over the baseline DCVA even within a limited training duration (restricted to a maximum of 15 epochs in our assessments). Given the inherent nature of DCVA to discern pixel-level alterations, it is evident that pixel-level contrastive learning frameworks yield more significant enhancements than instance-level contrastive learning strategies, in line with recent findings in the literature of contrastive learning [79], [80], [81].

C. Qualitative Evaluation of Detected Changes

Figs. 9 and 10 demonstrate the generated forest damage maps along with Sentinel-1 and Sentinel-2 image signatures before and after the forest disturbance. Both figures show selected patches of satellite images from several areas prior to disturbance

and after the disturbance, as well as generated disturbance maps along with their accuracy assessment versus ground reference data. Visual analysis and interpretation of Sentinel-2 images are comparatively easier due to the distinct changes in spectral signatures. Fully destroyed forests are captured with high precision. The performance in partially disturbed forest areas is also commendable. The relatively modest accuracy figures can be attributed to the extensive removal of forested areas, leading to overreporting, where the entire forest compartment is cleared despite only partial damage to specific stands. This is a normal and understandable forest management practice. However, it is noteworthy that the proportion of false positives is relatively low, indicating the suitability of the developed approach for operational applications with a minimal number of false alarms.

The visual assessment of snowload-induced damages using image patches cropped from Sentinel-1 images proves to be quite complex. In these cases, the forest remains intact, but its structural integrity is significantly altered. Moreover, seasonal changes are clearly visible between Sentinel-1 images of November and March. While general land cover distinctions are relatively apparent, distinguishing damaged areas within forests visually presents a considerable challenge. The complexity of visually analyzing changes in Sentinel-1 images underscores the necessity of developing and adapting approaches that go beyond simple image ratios and other basic hand-engineered features. Such approaches should take both radiometric and spatial context into account, given the intricate nature of the alterations observed.

D. Comparison to Prior Studies

Our results compare favorably to earlier studies focusing on bitemporal EO-imagery-based mapping of windstorm disturbances [18], [19]. The detailed comparison can be difficult as reporting practices in papers are different; no-disturbance class is sometimes missing or dominates overall accuracy if its proportion is too big. Then, the analysis should be focused on the probability of successful detection of change class, that

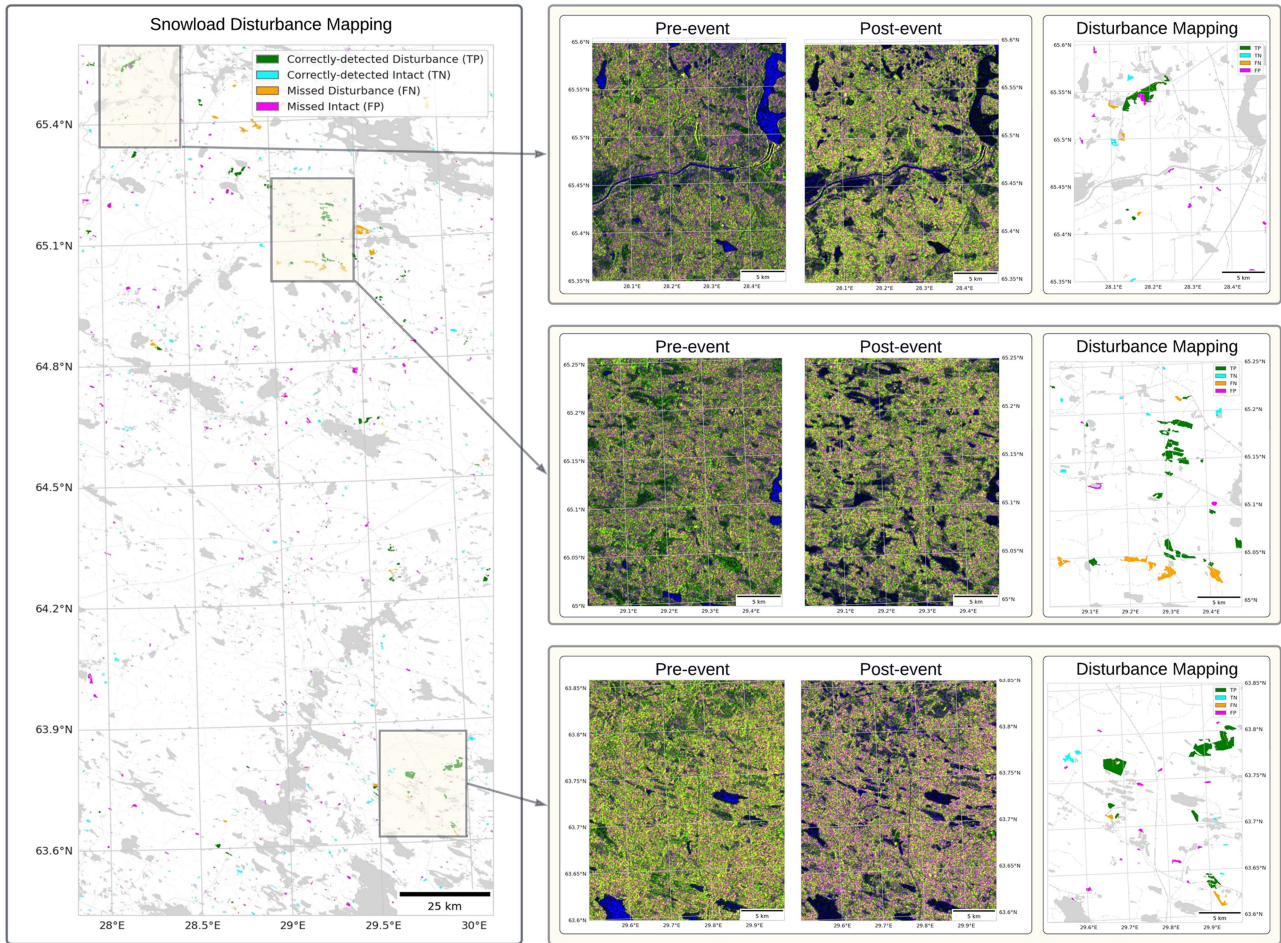


Fig. 9. Sentinel-1 image signatures (VV—red, VH—green, and VV/VH—blue) over the Kainuu study area before and after a snowload damage, with the corresponding disturbance mapping generated via DCVA.

is, evaluation of how well the detection method works, and if the user's accuracy (related to commission error) is high. Omission errors can be relatively high [20]. Typical reported accuracy for forest windstorm mapping is normally on the level of 60–80% with both optical and SAR data similar in resolution to Sentinel-2 and Sentinel-1 sensors and is often aimed to be a preliminary estimate to guide forest intervention activity [8] or to perform a retrospective study using images acquired long after the disturbance [22]. However, when images collected long after the windstorm are used to assess forest-damaged areas, it is possible that forest intervention activities, such as felling damaged trees and removing fallen logs, have already taken place. In such cases, detection algorithms might capture signs of human intervention. Human-induced disturbances such as clear-cutting are easier to spot, and differentiating between anthropogenic and natural disturbances can be problematic. On the other hand, accuracy figures can be influenced by the time lag between the occurrence of the windthrow event and the satellite image acquisition, as vegetation regrowth can affect the detectability of the disturbance.

Importantly, our study concentrates on bitemporal change detection. In line with several other studies, we expect that further improvement can be gained by leveraging EO image

time series, particularly before the disturbance in connection with snowload accumulation [7], and after the disturbance for both windstorm-induced and snowload damages [9]. This improvement, however, can occur at the cost of timeliness in damage assessment and reporting, as accumulating images requires time.

E. Strengths and Limitations of the Proposed Methodology

To begin with, the domain-agnostic design of the proposed method offers broad applicability for mapping various types of forest disturbances and detecting other land changes. Then, the methodology is proven to be efficient and swift, particularly crucial in urgent situations where ground measurements are unavailable or the observed disturbance is uncommon. Finally, it achieves remarkable accuracy in disturbance mapping, establishing itself as a reliable alternative to conventional methods.

Despite its strengths, our approach faces some bottlenecks. A key limitation is the dependence on the quality of pretrained models, which significantly influences its success. In scenarios lacking readily available SSL models, initiating training from scratch can be resource intensive. In addition, the methodology currently does not account for the multitemporality of gradual

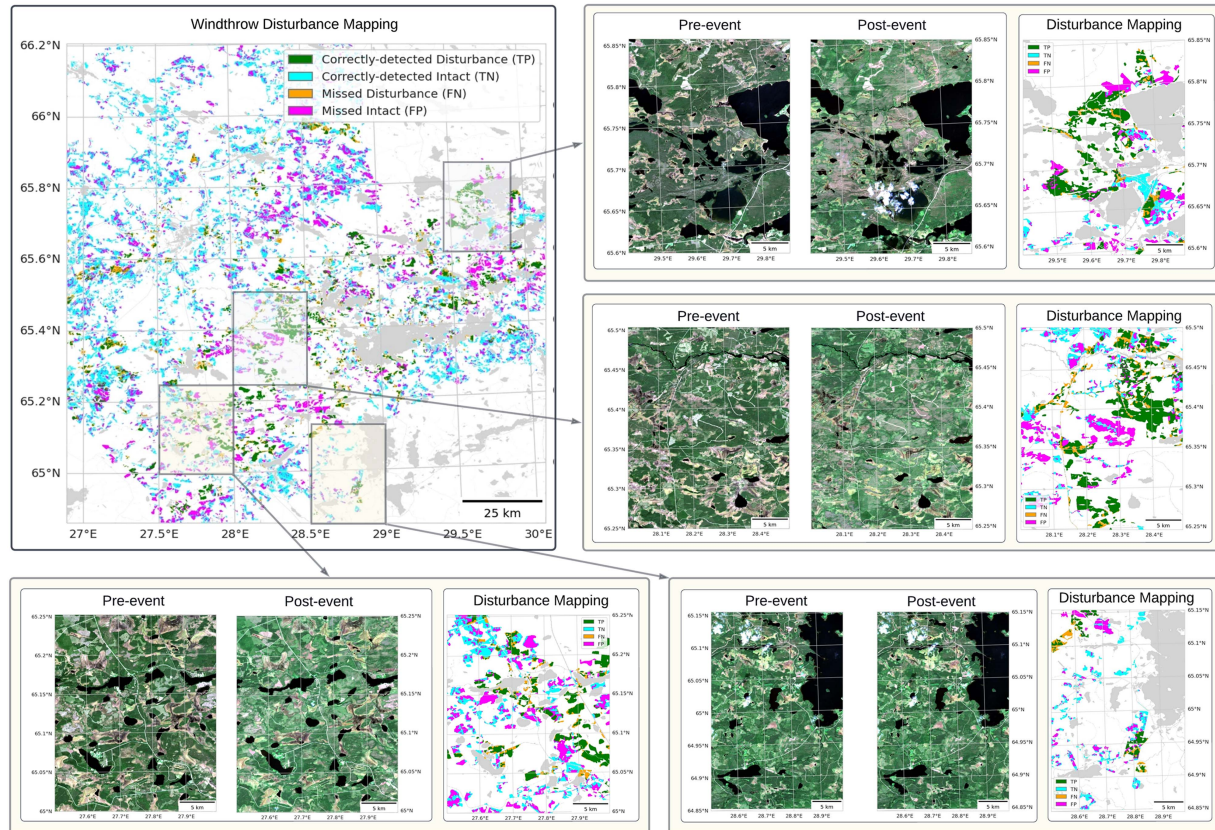


Fig. 10. Sentinel-2 image signatures (natural color coded) over the Taivalkoski study area before and after windthrow damage, with the corresponding disturbance mapping generated via DCVA.

changes [85], focusing instead on bitemporal inputs. Future efforts aim to address this gap. Another critical challenge is the method’s sensitivity to all surface changes, without distinguishing specific disturbance types, unless supplemented by expert domain knowledge.

F. Adaptability and Generalization of the Proposed Method

The methodologies adopted in this study, SSL with DCVA applied to EO data for detecting forest disturbances, have broad applicability to other study areas and similar applications as detailed below.

- 1) *Geographical adaptability*: The SSL and DCVA are not inherently restricted by geographic boundaries. They can be applied to any location where sufficient EO data are available, making them suitable for various types of forests, agricultural lands, urban areas, and other landscapes across the globe. This versatility is also evident from the EO literature, where applications of SSL [12] and those of DCVA [13] have been independently documented for change detection, as detailed in Section II-B.
- 2) *Flexibility with disturbance types*: While our study primarily focuses on natural disturbances such as windthrow and snowload, the presented methodology is potentially suitable for detecting other types of disturbances. Notably, we have run successful preliminary tests on also

on clear-cutting, a typical human-induced disturbance. However, these experiments were not included in this work since our main emphasis is on natural disturbances, and will be reported separately in future communications. In essence, the demonstrated capability to handle different disturbance types (both natural and human-induced) extends the applicability of our methodology to a wide range of environmental monitoring tasks.

- 3) *Applicability to various EO data and sensors*: Even though our approach is novel in terms of combining SSL and DCVA into a forest disturbance mapping pipeline, and our analysis primarily relied on Sentinel-1 and Sentinel-2 data, independent literature on SSL [12] and DCVA [13] demonstrates that both apply to multiple use cases and sensor types, including hyperspectral imagery and high-resolution airborne imagery.

In summary, the wide applicability of these methods, along with their proven effectiveness in various scenarios, highlights their potential as valuable tools in EO data analysis and change detection across diverse domains and settings.

VII. CONCLUSION AND OUTLOOK

In this study, we demonstrated the potential of SSL and transfer learning methods in the framework of DCVA for assessing forest disturbances using EO images. Studied cases

included forest disturbances induced by a strong windstorm and heavy snowload. Various SSL backbone encoders were tested, including PixPro, PixContrast, SimSiam, and BYOL. In the context of DCVA, we found that pixel-level contrastive learning yielded better differentiation of disturbed and nondisturbed forest areas. The most notable outcomes were observed when integrating pixel-level contrastive learning with DCVA. This approach yielded accuracies of 0.840 and 0.765 and F_1 scores of 0.567 and 0.692 for snowload and windthrow mappings using Sentinel-1 and Sentinel-2 images, respectively. Our results demonstrated the efficiency of the developed approach and indicated its superiority over previously reported methods in forest windthrow and snowload disturbance mapping. Our further research will concentrate on incorporating and testing other possible EO data sources, as well as adapting methodologies to time-series processing, especially in the context of SAR-based mapping of forest disturbances.

ACKNOWLEDGMENT

The authors would like to thank Prof. Erkki Tomppo of the University of Helsinki for his help in reference data curation in the context of the snowload damage study. The forest reference data used in this study were based on the data provided by the Finnish Forest Centre.

REFERENCES

- [1] R. Birdsey and Y. Pan, "Trends in management of the world's forests and impacts on carbon stocks," *Forest Ecol. Manage.*, vol. 355, pp. 83–90, 2015.
- [2] N. G. McDowell et al., "Global satellite monitoring of climate-induced vegetation disturbances," *Trends Plant Sci.*, vol. 20, no. 2, pp. 114–123, 2015.
- [3] A. S. Mori, K. P. Lertzman, and L. Gustafsson, "Biodiversity and ecosystem services in forest ecosystems: A research agenda for applied forest ecology," *J. Appl. Ecol.*, vol. 54, no. 1, pp. 12–27, 2017.
- [4] A. Heredia-Telles, P. M. López-Serrano, M. Molinier, and C. Wehenkel, "Evaluation of forest cover loss in properties in the Sierra Madre Occidental, State of Durango, Mexico, certified by the Forest Stewardship Council," *Trees Forests People*, vol. 4, 2023, Art. no. 100454.
- [5] R. Seidl et al., "Forest disturbances under climate change," *Nature Climate Change*, vol. 7, no. 6, pp. 395–402, 2017.
- [6] M. A. Tanase, C. Aponte, S. Mermoz, A. Bouvet, T. Le Toan, and M. Heurich, "Detection of windthrows and insect outbreaks by L-band SAR: A case study in the Bavarian Forest National Park," *Remote Sens. Environ.*, vol. 209, pp. 700–711, 2018.
- [7] 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, Art. no. 384.
- [8] M. Rüetschi, D. Small, and L. T. Waser, "Rapid detection of windthrows using Sentinel-1 C-band SAR data," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 115.
- [9] E. Tomppo, G. Ronoud, O. Antropov, H. Hytönen, and J. Praks, "Detection of forest windstorm damages with multitemporal SAR data—A case study: Finland," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 383.
- [10] R. Torres et al., "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, 2012.
- [11] M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.
- [12] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 4, pp. 213–247, Dec. 2022.
- [13] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.
- [14] D. E. Kislov and K. A. Korznikov, "Automatic windthrow detection using very-high-resolution satellite imagery and deep learning," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1145.
- [15] K. Einzmann, M. Immitzer, S. Böck, O. Bauer, A. Schmitt, and C. Atzberger, "Windthrow detection in European forests with very high-resolution optical data," *Forests*, vol. 8, no. 1, 2017, Art. no. 21.
- [16] T. A. Schroeder et al., "Improving estimates of forest disturbance by combining observations from landsat time series with U.S. Forest Service Forest Inventory and Analysis data," *Remote Sens. Environ.*, vol. 154, pp. 61–73, 2014.
- [17] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, "A comprehensive change detection method for updating the National Land Cover Database to circa2011," *Remote Sens. Environ.*, vol. 132, pp. 159–175, 2013.
- [18] E. A. Dyukarev, N. N. Pologova, E. A. Golovatskaya, and A. G. Dyukarev, "Forest cover disturbances in the South Taiga of West Siberia," *Environ. Res. Lett.*, vol. 6, Sep. 2011, Art. no. 035203.
- [19] M. Baumann, M. Ozdogan, P. T. Wolter, A. Krylov, N. Vladimirova, and V. C. Radeloff, "Landsat remote sensing of forest windfall disturbance," *Remote Sens. Environ.*, vol. 143, pp. 171–179, 2014.
- [20] C. Senf and R. Seidl, "Mapping the forest disturbance regimes of Europe," *Nature Sustainability*, vol. 4, no. 1, pp. 63–70, 2021.
- [21] M. Löw and T. Koukal, "Phenology modelling and forest disturbance mapping with Sentinel-2 time series in Austria," *Remote Sens.*, vol. 12, no. 24, 2020, Art. no. 4191.
- [22] V. Olmo, E. Tordoni, F. Petruzzellis, G. Bacaro, and A. Altobelli, "Use of Sentinel-2 satellite data for windthrows monitoring and delimiting: The case of "Vaia" storm in Friuli Venezia Giulia Region (North-Eastern Italy)," *Remote Sens.*, vol. 13, no. 8, 2021, Art. no. 1530.
- [23] X. Tang, K. H. Bratley, K. Cho, E. L. Bullock, P. Olofsson, and C. E. Woodcock, "Near real-time monitoring of tropical forest disturbance by fusion of landsat, Sentinel-2, and Sentinel-1 data," *Remote Sens. Environ.*, vol. 294, 2023, Art. no. 113626.
- [24] A. Thiele, M. Boldt, and S. Hinz, "Automated detection of storm damage in forest areas by analyzing TerraSAR-X data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 1672–1675.
- [25] R. A. S. Rosa, D. Fernandes, J. B. Nogueira, and C. Wimmer, "Automatic change detection in multitemporal X- and P-band SAR images using Gram-Schmidt process," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 2797–2800.
- [26] M. Santoro, A. Pantze, J. E. S. Fransson, J. Dahlgren, and A. Persson, "Nation-wide clear-cut mapping in Sweden using ALOS PALSAR strip images," *Remote Sens.*, vol. 4, no. 6, pp. 1693–1715, 2012.
- [27] E. J. M. Rignot and J. J. van Zyl, "Change detection techniques for ERS-1 SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 896–906, Jul. 1993.
- [28] M. Tanase, R. Kennedy, and C. Aponte, "Radar burn ratio for fire severity estimation at canopy level: An example for temperate forests," *Remote Sens. Environ.*, vol. 170, pp. 14–31, 2015.
- [29] A. Pantze, M. Santoro, and J. E. S. Fransson, "Change detection of boreal forest using bi-temporal ALOS PALSAR backscatter data," *Remote Sens. Environ.*, vol. 155, pp. 120–128, 2014.
- [30] N. Joshi et al., "Mapping dynamics of deforestation and forest degradation in tropical forests using radar satellite data," *Environ. Res. Lett.*, vol. 10, no. 3, 2015, Art. no. 034014.
- [31] E. C. De Grandi, E. Mitchard, I. H. Woodhouse, and G. D. De Grandi, "Spatial wavelet statistics of SAR backscatter for characterizing degraded forest: A case study from Cameroon," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3572–3584, Jul. 2015.
- [32] M. G. Hethcoat, J. M. Carreiras, D. P. Edwards, R. G. Bryant, and S. Quegan, "Detecting tropical selective logging with C-band SAR data may require a time series approach," *Remote Sens. Environ.*, vol. 259, 2021, Art. no. 112411.
- [33] O. Antropov, Y. Rauste, J. Praks, F. M. Seifert, and T. Häme, "Mapping forest disturbance due to selective logging in the Congo Basin with RADARSAT-2 time series," *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 740.
- [34] J. E. Fransson, A. Pantze, L. E. Eriksson, M. J. Soja, and M. Santoro, "Mapping of wind-thrown forests using satellite SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 1242–1245.
- [35] L. E. B. Eriksson, J. E. S. Fransson, M. J. Soja, and M. Santoro, "Backscatter signatures of wind-thrown forest in satellite SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 6435–6438.
- [36] C. Persello et al., "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, Jun. 2022.

- [37] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.
- [38] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5:4, pp. 8–36, Dec. 2017.
- [39] 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, Art. no. 2392.
- [40] 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, Art. no. 1413.
- [41] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5776–5787, 2022.
- [42] 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, Art. no. 5560.
- [43] C. Bolyn, P. Lejeune, A. Michez, and N. Latte, "Mapping tree species proportions from satellite imagery using spectral–spatial deep learning," *Remote Sens. Environ.*, vol. 280, 2022, Art. no. 113205.
- [44] 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, Art. no. 207.
- [45] S. Illarionova, D. Shadrin, V. Ignatiev, S. Shayakhmetov, A. Trekin, and I. Oseledets, "Estimation of the canopy height model from multispectral satellite imagery with convolutional neural networks," *IEEE Access*, vol. 10, pp. 34116–34132, 2022.
- [46] N. Lang, K. Schindler, and J. D. Wegner, "Country-wide high-resolution vegetation height mapping with Sentinel-2," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111347.
- [47] M. Möttus, 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 Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521711.
- [48] S. Björk, S. N. Anfinsen, E. Næsset, T. Gobakken, and E. Zahabu, "On the potential of sequential and nonsequential regression models for Sentinel-1-based biomass prediction in Tanzanian Miombo forests," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4612–4639, 2022.
- [49] S. Ge, O. Antropov, T. Häme, R. E. McRoberts, and J. Miettinen, "Deep learning model transfer in forest mapping using multi-source satellite SAR and optical images," *Remote Sens.*, vol. 15, no. 21, 2023, Art. no. 5152.
- [50] M. Zhang, W. Li, X. Zhao, H. Liu, R. Tao, and Q. Du, "Morphological transformation and spatial-logical aggregation for tree species classification using hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501212.
- [51] Z. M. Hamdi, M. Brandmeier, and C. Straub, "Forest damage assessment using deep learning on high resolution remote sensing data," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 1976.
- [52] K. Isaenkov, M. Yushchuk, V. Khrantsov, and O. Seliverstov, "Deep learning for regular change detection in Ukrainian forest ecosystem with Sentinel-2," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 364–376, 2020.
- [53] D. E. Kislov, K. A. Korznikov, J. Altman, A. S. Vozmishcheva, and P. V. Krestov, "Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images," *Remote Sens. Ecol. Conservation*, vol. 7, no. 3, pp. 355–368, 2021.
- [54] A. Safonova, S. Tabik, D. Alcaraz-Segura, A. Rubtsov, Y. Maglinets, and F. Herrera, "Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 643.
- [55] J. Li et al., "A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images," *Remote Sens. Environ.*, vol. 280, 2022, Art. no. 113197.
- [56] M. Leenstra, D. Marcos, F. Bovolo, and D. Tuia, "Self-supervised pre-training enhances change detection in Sentinel-2 imagery," in *Proc. Int. Conf. Pattern Recognit.*, 2021, pp. 578–590.
- [57] Y. Chen and L. Bruzzone, "A self-supervised approach to pixel-level change detection in bi-temporal RS images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4413911.
- [58] L. Yan, J. Yang, and J. Wang, "Domain knowledge-guided self-supervised change detection for remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4167–4179, 2023.
- [59] C. Wang, S. Du, W. Sun, and D. Fan, "Self-supervised learning for high-resolution remote sensing images change detection with variational information bottleneck," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5849–5866, 2023.
- [60] Z. Cai, Z. Jiang, and Y. Yuan, "Task-related self-supervised learning for remote sensing image change detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 1535–1539.
- [61] J. Paul, B. U. Shankar, B. Bhattacharyya, and A. K. Datta, "Unsupervised change detection in remote sensing images using CNN based transfer learning," in *Proc. 5th Int. Conf. Comput. Data Sci.*, Nashik, India, 2021, pp. 463–474.
- [62] G. Zhou, M. Liu, and X. Liu, "An autoencoder-based model for forest disturbance detection using Landsat time series data," *Int. J. Digit. Earth*, vol. 14, no. 9, pp. 1087–1102, 2021.
- [63] L. Bergamasco, S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised change detection using convolutional-autoencoder multiresolution features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4408119.
- [64] T. Di Martino, B. Le Saux, R. Guinvarc'h, L. Thirion-Lefevre, and E. Colin, "Detection of forest fires through deep unsupervised learning modeling of Sentinel-1 time series," *ISPRS Int. J. Geo- Inf.*, vol. 12, no. 8, 2023, Art. no. 332.
- [65] S. Ge, H. Gu, W. Su, A. Lönnqvist, and O. Antropov, "A novel semisupervised contrastive regression framework for forest inventory mapping with multisensor satellite data," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 2502705.
- [66] O. Antropov et al., "Semi-supervised deep learning representations in earth observation based forest management," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 650–653.
- [67] J. Miettinen et al., "Demonstration of large area forest volume and primary production estimation approach based on Sentinel-2 imagery and process based ecosystem modelling," *Int. J. Remote Sens.*, vol. 42, no. 24, pp. 9467–9489, 2021.
- [68] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS—A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," in *Proc. Photogrammetric Image Anal. Conf.*, 2019, pp. 153–160.
- [69] Y. Rauste, A. Lonnqvist, M. Molinier, J.-B. Henry, and T. Hame, "Orthorectification and terrain correction of polarimetric SAR data applied in the ALOS/Palsar context," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 1618–1621.
- [70] D. Small, "Flattening gamma: Radiometric terrain correction for SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 3081–3093, Aug. 2011.
- [71] E. Tomppo and M. Halme, "Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: A genetic algorithm approach," *Remote Sens. Environ.*, vol. 92, no. 1, pp. 1–20, 2004.
- [72] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [73] J.-B. Grill et al., "Bootstrap your own latent—A new approach to self-supervised learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 21271–21284.
- [74] Z. Xie, Y. Lin, Z. Zhang, Y. Cao, S. Lin, and H. Hu, "Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 16684–16693.
- [75] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15750–15758.
- [76] F. Pacifici, N. Longbotham, and W. J. Emery, "The importance of physical quantities for the analysis of multitemporal and multiangular optical very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6241–6256, Oct. 2014.
- [77] F. Bovolo, L. Bruzzone, L. Capobianco, A. Garzelli, S. Marchesi, and F. Nencini, "Analysis of the effects of pansharpening in change detection on VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 53–57, Jan. 2010.
- [78] Y. Han, F. Bovolo, and L. Bruzzone, "An approach to fine coregistration between very high resolution multispectral images based on registration noise distribution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6650–6662, 2015.
- [79] M. Chen, Z. Zheng, Y. Yang, and T.-S. Chua, "PiPa: Pixel-and patch-wise self-supervised learning for domain adaptive semantic segmentation," in *Proc. 31st ACM Int. Conf. Multimedia*, 2023, pp. 1905–1914.

- [80] Z. Pang, Y. Nakashima, M. Otani, and H. Nagahara, "Revisiting pixel-level contrastive pre-training on scene images," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2024, pp. 1784–1793.
- [81] M. Tang, K. Georgiou, H. Qi, C. Champion, and M. Bosch, "Semantic segmentation in aerial imagery using multi-level contrastive learning with local consistency," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2023, pp. 3798–3807.
- [82] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. 7th Int. Conf. Learn. Representations*, 2019.
- [83] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.
- [84] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.
- [85] M. Molinier, J. Miettinen, D. Ienco, S. Qiu, and Z. Zhu, "Optical satellite image time series analysis for environment applications: From classical methods to deep learning and beyond," in *Change Detection and Image Time Series Analysis 2: Supervised Methods*. Hoboken, NJ, USA: Wiley, 2021, pp. 109–154.



Rıdvan Salih Kuzu received the bachelor's degree in electrical and electronics engineering and the master's degree in system and control engineering from Boğaziçi University, İstanbul, Türkiye, in 2010 and 2017, respectively, and the Ph.D. degree in applied electronics from Roma Tre University, Rome, Italy, in 2021.

Since 2010, he has held various positions at both private companies and public research institutions, including roles as an R&D engineer, Consultant, Manager, and Marie-Curie Research Fellow. Since 2021,

he has been a Senior Researcher with the Remote Sensing Technology Institute, German Aerospace Center, Weßling, Germany. He is also an AI Consultant with the Helmholtz AI Cooperation Unit. He is involved in several earth observation projects as a Senior Researcher and Principal Investigator. His research interests include representation learning, hyperspectral image processing, self-supervised and few-shot learning, explainable AI, and multimodal foundation models.

Dr. Kuzu received the 15th European Biometrics Research Award for his doctoral thesis by the European Association for Biometrics in 2021, the Best Demo Award at the 9th GTTI Thematic Meeting on Multimedia Signal Processing in 2019, the Best Paper Award at the International Conference on Pattern Recognition Applications and Methods in 2022, and the Hyperview Challenge Award at the International Conference on Image Processing in 2022.



Oleg Antropov (Member, IEEE) received the M.Sc. degree in computer science and the Candidate of Sciences degree in radio physics from Dnipropetrovsk National University, Dnipro, Ukraine, in 2006 and 2010, respectively, and the Ph.D. degree in space technology from Aalto University, Espoo, Finland, in 2014.

Since 2010, he has been a Research Scientist and Senior Scientist with the Remote Sensing Group, VTT Technical Research Centre of Finland Ltd., Espoo. He was a Postdoctoral Associate with Aalto

University, during 2014–2018, and a Research Consultant with ICEYE, Espoo, in 2018. His current research interests include the development of physics-based and deep learning modeling approaches to effectively apply multisource synthetic aperture radar and optical data for various environmental applications, with a central emphasis on forest inventories, soil moisture retrieval, and crop mapping.



Matthieu Molinier (Member, IEEE) received the Engineering degree in multiple subjects including physics (including quantum, crystals, wave propagation etc.), electronics, automation, photonics, computer science, programming, and signal and image processing from École Nationale Supérieure de Physique de Strasbourg, Illkirch-Graffenstaden, France, and the M.Sc. degree in image processing from Université Louis Pasteur, Strasbourg, France, both in 2004.

Since then, he has been with VTT Technical Research Centre of Finland Ltd., Espoo, Finland, as a Research Scientist and then Senior Scientist in earth observation, with a main focus on machine learning and spatiotemporal analysis in optical satellite images. He is involved in several projects of the European Space Agency and the European Commission Programs, as a Work Package Manager. He is the author of 70 articles in scientific journals and conferences. His current research interests include dense satellite image time-series analysis, and weakly supervised deep learning for multispectral and hyperspectral images, with applications to environment monitoring.

Mr. Molinier is an Active Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.



Corneliu Octavian Dumitru received the B.S. and M.S. degrees in artificial intelligence and pattern recognition from the Faculty of Electronics, Telecommunications and Information Technology 2001 and 2002, respectively, and the Ph.D. degree in engineering in 2001, all from the Politehnica University of Bucharest (UPB), Bucharest, Romania, and the Ph.D. degree in telecommunications from Pierre and Marie Curie University, Paris, France, in 2010.

He was a Research Assistant between 2001 and 2007 and a Lecturer between 2007 and 2011 with UPB, where he delivered lectures and seminars and supervised laboratory works in the fields of information and estimation theory, communication theory, and signal processing. From 2007 and 2009, he was a Researcher in industrial projects with Telecom SudParis, Evry, France (formerly INT) delivering algorithms for the audio visual and film industry and for the mobile service provider SFR (Vodafone group in France). Since 2010, he has been a Scientist with the Remote Sensing Technology Institute, German Aerospace Center, Weßling, Germany. Since 2002, he has been supervising/cosupervising bachelor, master, and Ph.D. theses in the field of artificial intelligence, machine learning, data mining, knowledge discovery in databases, semantics, compression, benchmarking dataset creation, impact of global changes, speech recognition, speaker verification, and watermarking. He is currently involved in several space projects as project management in the frame of the European Space Agency and the European Commission Programs for information extraction, explainable artificial intelligence, machine learning and knowledge discovery, semantics, ontologies, and knowledge graphs using remote sensing imagery. His research interests include stochastic process information, model-based sequence recognition and understanding, basics of man-machine communication, information management, semantics, change detection in image time series, data mining, and image retrieval in extended databases. His research applications in earth observation are focused on the analysis of urban areas, changes in polar areas, monitoring of natural disasters, monitoring of the coastal area, and security.

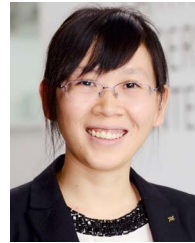


Sudipan Saha (Member, IEEE) received the M.Tech. degree in electrical engineering from the Indian Institute of Technology Bombay, Mumbai, India, in 2014, and the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, and Fondazione Bruno Kessler, Trento, in 2020.

He was a Postdoctoral Researcher with the Technical University of Munich (TUM), Munich, Germany, and an Engineer with TSMC Limited, Hsinchu, Taiwan, from 2015 to 2016. In 2019, he was a Guest

Researcher with TUM. He is currently an Assistant Professor with the Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi, India. His research interests include multitemporal remote sensing image analysis, domain adaptation, time-series analysis, image segmentation, deep learning, image processing, and pattern recognition.

Dr. Saha was the recipient of the Fondazione Bruno Kessler Best Student Award in 2020. He is a Reviewer for several international journals and served as a Guest Editor for *Remote Sensing* special issue on “Advanced Artificial Intelligence for Remote Sensing: Methodology and Application.”



Xiao Xiang Zhu (Fellow, IEEE) received the M.Sc., Dr.-Ing., and “Habilitation” degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is currently the Chair Professor for Data Science in Earth Observation with TUM. She was the Founding Head of the Department “EO Data Science,” Remote Sensing Technology Institute, German Aerospace Center, Weßling, Germany. Since 2020, she has been the Principal Investigator and Director

of the international future AI lab “AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond,” Munich. Since 2020, she has been a Director of the Munich Data Science Institute, TUM. From 2019 to 2022, she has been a co-coordinator of the Munich Data Science Research School and the Head of Helmholtz Artificial Intelligence—Research Field “Aeronautics, Space and Transport.” She was a Guest Scientist or Visiting Professor with the Italian National Research Council—Institute for Electromagnetic Sensing of the Environment, Naples, Italy; Fudan University, Shanghai, China; the University of Tokyo, Tokyo, Japan; and University of California, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is currently a Visiting AI Professor with European Space Agency’s Phi-lab, Frascati, Italy. Her main research interests include remote sensing and earth observation, signal processing, machine learning, and data science, with their applications in tackling societal grand challenges, e.g., global urbanization, United Nations’ sustainable development goals, and climate change.

Dr. Zhu is a Member of Young Academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities, the German National Academy of Sciences Leopoldina, and the Bavarian Academy of Sciences and Humanities. She is on the Scientific Advisory Board of several research organizations, including the German Research Center for Geosciences (2020–2023) and the Potsdam Institute for Climate Impact Research. She is an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *Pattern Recognition* and an Area Editor for special issues of *IEEE Signal Processing Magazine*.