This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3460981

A Deep Semantic Segmentation Approach to Map Forest Tree Dieback in Sentinel-2 Data

Giuseppina Andresini and Annalisa Appice and Donato Malerba, *Senior Member, IEEE*

***Abstract*—Massive tree dieback events triggered by various disturbance agents, such as insect outbreaks, pests, fires and windstorms, have recently compromised the health of forests in numerous countries with a significant impact on ecosystems. The inventory of forest tree dieback plays a key role in understanding the effects of forest disturbance agents and improving forest management strategies. In this article, we illustrate a deep learning approach that trains a U-Net model for the semantic segmentation of Sentinel-2 images of forest areas. The proposed U-Net architecture integrates an attention mechanism to amplify the crucial information and a self-distillation approach to transfer the knowledge within the U-Net architecture. Experimental results demonstrate the significant contribution of both attention and self-distillation to gaining accuracy in two case studies in which we perform the inventory mapping of forest tree dieback caused by insect outbreaks and wildfires, respectively.**

***Index Terms*—Forest tree die-back monitoring, Semantic segmentation, Attention, Self-distillation, Sentinel-2 image processing, Insect outbreak monitoring, Forest wildfires monitoring**

I. INTRODUCTION

Forest and woodland cover around one-third of the total land area on Earth performing vital functions in mitigating climate changes, as well as providing a wide range of ecosystem services driven by their biodiversity [1]. *Forest tree dieback* is an umbrella term that covers disturbance agents of all kinds that negatively affect the health of forests. In particular, forest damages may be caused by both biotic and abiotic agents (or their combination), which cause mortality, or a significant loss of vitality of trees [2].

Forest tree dieback surveillance has a crucial role in natural resource monitoring by enabling foresters to perform informed decision-making for effective environmental conservation and sustainable development. In particular, monitoring forest tree dieback caused by several disturbance agents (e.g., insect outbreak, wildfires) is a fundamental step to improve forest management strategies. Predicting where future disturbances are likely to cause forest tree dieback is certainly a crucial and challenging task in planning preventive mitigation strategies. Besides forest tree dieback prevention, a monitoring system for large-scale inventory of forest tree dieback is also essential to prompt disturbance modelling, particularly in relation to climate or pollution, as well as to trigger quickly post-disturbance forest management [3]. However, inventory strategies currently adopted by foresters are mainly based on terrestrial fieldwork resorting to human assessment of the health status of forest trees. This is laborious and, in the case of large forest zones, it is time-consuming and expensive [4].

On the other hand, recent and rapid developments in remote sensing fields have fostered the observation and imaging of large-scale forest areas at low cost. Although the image availability may be sometimes limited by technical issues (e.g., cloud cover), forest images are commonly available in an inexpensive way,[1] at a range of spatial and temporal scales [5]. Simultaneously with the development of remote sensing technology, image analysis and classification have obtained impressive achievements in environmental monitoring and management. Nowadays deep learning techniques applied to remote sensing data are systematically reducing the amount of forestry fieldwork by increasing forester abilities to speed-up the production of maps of forest inventory results [6].

In particular, U-Net [7] is an increasingly popular deep learning architecture, that is characterised by an encoder-decoder convolutional structure oriented to perform imagery semantic segmentation (i.e., categorization of pixels into a class). This study is boosted by the good performance recently achieved learning a U-Net model in a variety of image types including remote sensing images [8], [9], [10]. Specifically, the presented study looks to further the effort towards training a U-Net model with Sentinel-2 images[2] as input to replicate ground truth forest tree dieback inventory mapping obtained with workfield. The aim of this paper is to explore the incorporation of both an attention mechanism and a self-distillation strategy into a U-Net architecture to improve the accuracy of semantic segmentation upon a standard U-Net architecture.

The *attention* [11] is a popular deep learning mechanism that was originally introduced to improve the accuracy performance of the encoder-decoder model for machine translation. It amplifies crucial information allowing the adaptive selection of the input where the model "sees" the most important information. In this study, we incorporate the channel attention [12] that captures cross-channel interaction in an efficient way. The *self-distillation* [13] is a knowledge distillation strategy, originally formulated for object detection, that realizes a knowledge transfer in the same model - from the deeper layers to the shallow layers. The proposed approach, named TITANIA (self-DisTIlled uneT for Sentinel-2 imAge semaNtIc

G. Andresini, A. Appice and D. Malerba are with the Dipartimento di Informatica, Università degli Studi Aldo Moro di Bari and Consorzio Interuniversitario Nazionale per l'Informatica - CINI, via Orabona, 4 - 70125 Bari - Italy E-mail: giuseppina.andresini@uniba.it, annalisa.appice@uniba.it, donato.malerba@uniba.it

[1]Remote image acquisition may be expensive due to the cost of building, launching, and operating satellite remote-sensing systems. However, satellite image collections, e.g., the Earth Sentinel-2 image collection acquired in the Copernicus program, are freely available.

[2]The Sentinel-2 satellite carries a multi-spectral optical sensor that produces images of the Earth with 10 m, 20 m and 60 m spatial resolution.

segmentAtion), applies a self-distillation strategy to a U-Net model that is trained with attention on Sentinel-2 images of forest scenes.

The effectiveness of TITANIA is shown through an extensive experimentation performed in two case studies regarding the inventory of forest tree dieback patches due to insect (bark beetle) outbreaks and wildfires, respectively. With regard to the task of mapping forest tree dieback caused by bark beetle outbreaks, we note that this study follows our previous work [14], where we presented a machine learning approach to use self-learning for training an XGBoost classification model by leveraging both labelled and unlabelled Sentinel-2 images of forest scenes. As in the case study considered in this work, the XGBoost model was used in [14] to perform the inventory of forest tree dieback caused by bark beetle outbreaks in the unlabelled scenes of the self-learning step. Although the evaluation study illustrated in [14] showed that the self-learning step allowed the XGBoost classification model to achieve valuable accuracy performances to map forest tree dieback in Sentinel-2 images of forest scenes hosting bark beetle outbreaks, the experimentation of this study shows that the deep neural model of TITANIA gains accuracy compared to the machine learning one. Notably, this is coherent with the recent achievements in remote sensing [15] assessing the superiority of deep learning, here enriched with attention mechanism and self-distillation, with respect to machine learning.

The paper is organized as follows. The related work is presented in Section II. Section III describes the method used in this work. Section IV discusses the results obtained. Finally, Section V draws conclusions and illustrates future developments.

II. Related work

Several remote sensing studies leverage machine learning and deep learning algorithms to monitor the health status of forests in Sentinel-2 imagery data, and map different types of forest disturbance events that cause forest tree dieback events [16]. The most used machine learning algorithms are Random Forest [4], [17], [18], [19], [20], [21] and Support Vector Machine [21], [22], [23], [24], while XGBoost has been recently experimented in [14]. These algorithms are often coupled to spectral vegetation indices and specific learning strategies such as cost-based learning [23] or self-learning [14].

On the other hand, deep learning strategies have recently seen a massive rise in popularity in several remote sensing image analysis problems [15] by including problems of forest health monitoring. [25] evaluates the performance of a U-Net model trained to detect the forest damage induced by bark beetle hotspots in Sentinel-2 images. [26] investigates the performance of a U-Net model trained for forest change detection in airborne images. [27] tests a combination of a Deep Neural Network, a Support Vector Machine and a Random Forest. [28] explores the performance of a Convolution Neural Network trained in UAV images of forest areas damaged by bark beetle hotspots. [29] presents an adjusted Mask R-CNN to map standing dead trees in color-infrared aerial images of forest areas. [9] integrates the attention mechanism in a standard U-Net model and uses the defined architecture for detecting deforestation within 3-band and 4-band Sentinel-2 images. [30] tests multiple deep neural models, including U-Net models and U-Net models with attention, for wildfire detection in Sentinel-2 images.

Despite several recent remote sensing studies have contributed to achieve good performance by using both U-Net and U-Net with attention architectures in several problems of forest health monitoring, to the best of our knowledge, no previous study has explored achievements of self-distilling a U-Net model to perform the inventory mapping of forest tree dieback in Sentinel-2 images. In fact, the self-distillation strategy is mostly applied to object detection. [31] uses self-distillation to boost the accuracy of various deep neural networks trained and evaluated for object detection in RGB images. A few recent studies have started the investigation of self-distillation within semantic segmentation in the medical domain. For example, [32] explores the achievements of self-distillation with semantic segmentation of 3D medical images.

III. Proposed method

The schema of TITANIA is shown in Figure 1. Let us consider $\mathcal{D} = \{(\mathbf{X}, \mathbf{Y})\}$ a collection of Sentinel-2 images of forest scenes recorded so that every mask $\mathbf{Y}$ is one-to-one associated with an image $\mathbf{X}$. In particular, $\mathbf{Y} \in \{0,1\}^{W \times H}$ is a binary mask with 1 = "damaged" and 0 = "healthy", while $\mathbf{X} \in \mathbb{R}^{W \times H \times C}$ is a tensor of spectral data. W and H are the width and height of the imagery scene, while C is the number of spectral channels.

The architecture of TITANIA consists of two main components:

- A backbone model (or teacher model) that is implemented with a U-Net architecture integrating an attention mechanism.
- Student models using the intermediate information of the decoder part of the U-Net architecture.

Both the teacher and students contribute to the final semantic segmentation performed by TITANIA.

U-Net is a deep learning architecture consisting of two components: the encoder (or downsampling path) and the decoder (or upsampling path). In TITANIA, we used an encoder part, where each downsampling layer consists of blocks composed of both a Batch Normalization layer (BN) and a 2D Convolutional layer, followed by a Max-pooling layer for downsampling. The height and width of the tensor halve after each step, while the channel size remains unchanged. In the decoder part, each upsampling layer consists of one transposed Convolution layer and blocks of Batch Normalization (BN) and 2D Convolutional layers. The decoder part of the U-Net is symmetric to the encoder part.

During the upsampling path, the recreated spatial information may lack accuracy [33]. This issue is commonly addressed in the traditional U-net architecture using skip connections to integrate spatial information from the downsampling path to the upsampling path. Following the suggestion of [33], we use a channel attention within the skip connections, to highlight

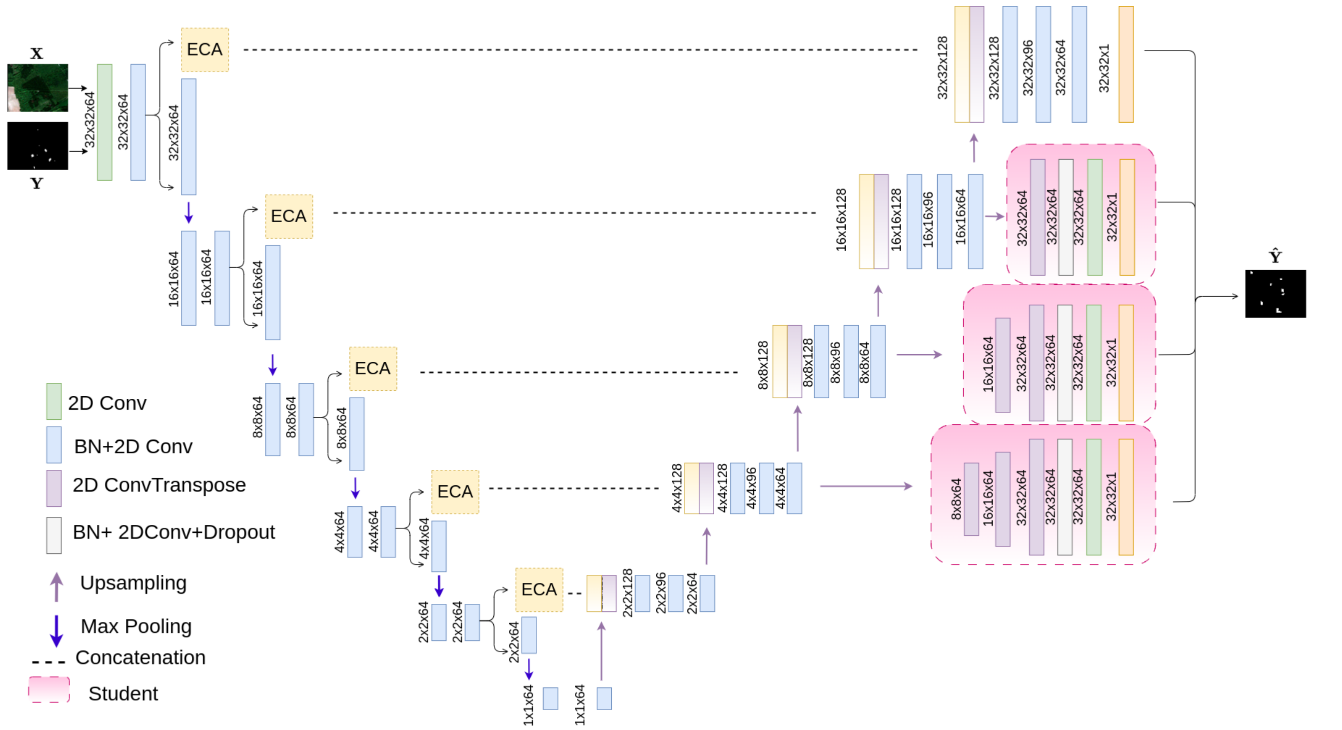


Figure 1: Schema of TITANIA. Abbreviations: Conv = Convolution; BN = Batch Normalization; ECA=Efficient Channel Attention

important information and mitigate the impact of redundant data. For this purpose, we adopt the Efficient Channel Attention mechanism (ECA), introduced in [12]. Notably, this channel attention mechanism is shown to achieve a good trade-off between accuracy and efficiency [12].

As shown in Figure 2, ECA takes tensor data with dimension $W' \times H' \times C'$ as input. Notably, the input of the ECA module is the output of a Convolutional layer. First ECA applies the Global Average Pooling (GAP) to reduce the input tensor to an output tensor with dimension $1 \times 1 \times C'$. This reduction is performed by computing the channel-wise average on the input tensor. Subsequently, it applies a 1D Convolutional layer with a kernel size equal to k. This layer keeps a tensor with dimension $1 \times 1 \times C'$ and learns local cross-channel dependencies between neighbouring channels. The Sigmoid function σ is applied to the output of the 1D Convolutional layer. This produces a tensor with dimension $1 \times 1 \times C'$ that contains the attention values. Finally, ECA computes the element-wise product between the input data tensor and the attention tensor by producing the output attention-aware data tensor with dimension $W' \times H' \times C'$.

Self-distillation is a variant of knowledge distillation. Traditional knowledge distillation implements a two-stage training method where a pre-trained (complex) teacher model is used to train a (simple) student model. Differently, self-distillation combines teacher and students into a single neural network by learning both in a one-stage training method. In TITANIA, we consider the U-Net as the teacher model and Fully Convolutional Networks as student models. Students start from the output of upsampling layers.

Let us consider $\hat{y}_T$ the predictions computed with the

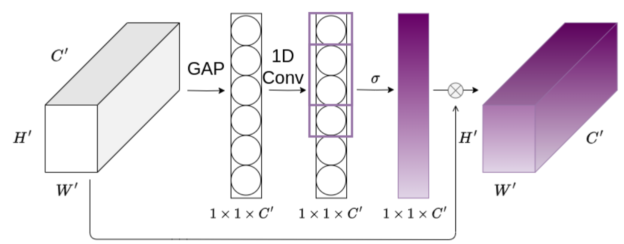


Figure 2: Schema of ECA. Abbreviations: 1D Conv = 1D Convolutional layer; GAP = Global Average Pooling

teacher and $\hat{y}_{S_i}$ the predictions produced by the student S_i with $i = 1, \ldots, N$. The distillation loss minimised by TITANIA is defined as:

$$\mathbf{L_D} = (1 - \alpha)\mathbf{L_T}(y, \hat{y}) + \alpha \mathbf{L_{KL}}(\hat{y_T}, \hat{y}_\mathbf{S}), \quad (1)$$

with $\alpha \in [0, 1]$ that balances the two terms $\mathbf{L_T}$ and $\mathbf{L_{KL}}$. The term $\mathbf{L_T}$ refers to the Tversky loss that is commonly used to address the issue of data imbalance [34]. In this study, $\mathbf{L_T}$ is defined as the average of the Tversky loss computed on the predictions of both the teacher and students, that is:

$$\mathbf{L_T}(\hat{y}, \hat{y}) = \frac{1}{N+1}\left(\sum_{i=1}^{N} \mathbf{L_T}(y, \hat{y}_{S_i}) + \mathbf{L_T}(y, \hat{y}_T)\right). \quad (2)$$

The term $\mathbf{L_{KL}}(\hat{y}_T, \hat{y}_\mathbf{S})$ refers to the Kullback-Leibler (KL) divergence computed to distil the knowledge between predictions of the teacher and the predictions of each student. In

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3460981

particular, $\mathbf{L_{KL}}$ is defined as follows:

$$\mathbf{L_{KL}}(\hat{y}_T, \hat{y}_\mathbf{S}) = \frac{1}{N}\sum_{i=1}^{N} KL(\frac{\hat{y}_T}{\tau}, \frac{\hat{y}_{S_i}}{\tau}), \tag{3}$$

where τ is a user-defined parameter, called temperature, introduced to generate soft labels of both $\hat{y}_T$ and $\hat{y}_{S_i}$. The higher the value of τ, the softer the probability distribution computed over the classes [35].

Finally, TITANIA returns the segmentation map of an image $\mathbf{X}$ by using the soft rule on the teacher and students. For each imagery pixel, we collect the probabilities predicted by the teacher and students for both class labels and select the class label that achieves the highest probability.

IV. Evaluation study

A. Case study and data description

We evaluated the performance of TITANIA in two case studies, namely BarkBeetle and WildFires, which involve collections of non-overlapping forest scenes damaged by bark beetle and wildfires, respectively. In both case studies, we processed collections of Sentinel-2, cloud-free images of selected forest scenes downloaded from ESA Copernicus open access hub. Sentinel-2 spectral bands with different spatial resolutions from 10 meters were all re-sampled at the same pixel size of 10 meters resolution.

1) BarkBeetle case study: In the BarkBeetle case study, we considered a collection of 87 squared, non-overlapping, forest scenes spanned in the Northeast of France. These scenes cover spruce forestry areas, which were infested by bark beetles in October 2018. The study scenes are all covered by the map delineating the boundaries of bark beetle hotspots causing forest tree dieback in the Northeast of France in October 2018.[3] As no large windthrows occurred in the years before 2018 in the study scenes, the mapped outbreaks were very likely caused by the hot summer droughts in 2018 [36].

The scenes of this study cover 2128785 pixels at 10 square meters resolution. The size of the study scenes varies from 38×44 to 268×256 pixels at 10 square meters resolution, while the percentage of damaged territory per scene varies from 0.0027% to 0.13% of the scene surface. The total percentage of damaged territory of the entire study area is 0.022%. We conducted the evaluation study by considering 71 scenes (covering 1639971 pixels at 10 square meters resolution) as training set and 16 scenes (covering 488814 pixels) as testing set.

We considered six distinct collections of Sentinel-2 images acquired in the study territory. These collections, namely BarkBeetleMay, BarkBeetleJun, BarkBeetleJul, BarkBeetleAug, BarkBeetleSept and BarkBeetleOct, were acquired monthly from May 2018 to October 2018, respectively.[4] Specifically, each imagery collection contains the Sentinel-2 images of the study scenes acquired in the selected month in 2018. For example, BarkBeetleOct collects Sentinel-2 images of study scenes acquired in October 2018 – the time at which the ground truth maps of the bark beetle outbreak were produced. Figure 3 shows the RGB of Sentinel-2 images collected in BarkBeetleMay, BarkBeetleJun, BarkBeetleJul, BarkBeetleAug, BarkBeetleSept and BarkBeetleOct for one of the testing scenes of the BarkBeetle case study.

2) WildFires case study: In the WildFires case study, we considered a collection of 13 squared, non-overlapping, scenes of forest areas spanned across Spain with different terrain types. These scenes hosted wildfire events between June 2017 and August 2018. The ground truth masks of the burned areas were available based on vector data of Copernicus EMS annotations.

The scenes of this study cover 16196048 pixels at 10 square meters resolution. The size of study scenes varies from 564×531 to 1537×1873 pixels at 10 square meters resolution, while the percentage of burned areas per scene varies from 0.0022% to 0.46%. The total percentage of burned territory in the entire study area is 0.094%. We conducted the evaluation study by considering 10 scenes (covering 11376495 pixels) as training set and 3 scenes (covering 4819553 pixels) as testing set.

We considered two Sentinel-2 image collections, namely WildFiresPre and WildFiresPost, respectively. WildFiresPre contains Sentinel-2 images of study scenes collected one month before the activation date of each forest wildfire event occurred in every study scene. WildFiresPost contains Sentinel-2 images collected one month after the activation date of each forest wildfire event occurred in every study scene. The two datasets are publicly available,[5] while a detailed description of the data is provided by [37]. Figure 4 shows the RGB of Sentinel-2 images collected in WildFiresPre and WildFiresPost, respectively, for one of the testing scenes of the WildFires case study.

B. Implementation details

We developed TITANIA[6] in Python 3 using the high-level neural network API Keras 2.11 integrated into TensorFlow.

The teacher backbone was based on a U-Net architecture optimized for satellite images [7] and extended here with the attention mechanism. In particular, each downsampling layer of the encoder part of the adopted U-Net architecture consists of 3 blocks containing both a BN layer and a 2D Convolutional layer per block. The decoder part implements each upsampling layer with one transposed Convolutional layer and three blocks

[3]The bark beetle outbreak ground truth map can be accessed via https://macarte.ign.fr/carte/3bd52aa2b6422a3a58b5086576f91080/Foyers+de+scolytes+dans+les+pessi\%C3\%A8res+et+les+sapini\%C3\%A8res+du+Nord-Est+de+la+France,+automne+2018-printemps+2019

[4] The Sentinel-2 imagery collections of the BarkBeetle case study were produced within the EU research project SWIFTT (https://swiftt.eu/), to allow the research partners of the project to develop machine learning which can provide forest managers with models for monitoring forest tree dieback due to bark beetle infestation.

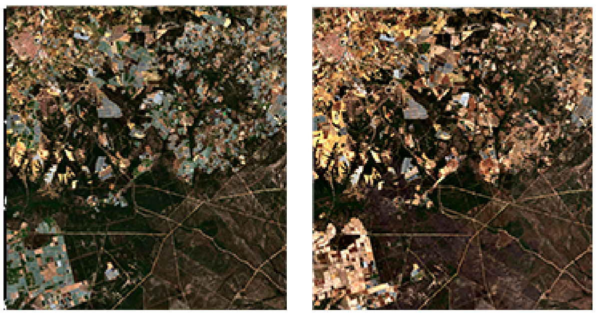
[5]We used the images of the directory pink in https://doi.org/10.5281/zenodo.6597139

[6]https://github.com/gsndr/TITANIA

[7]https://github.com/karolzak/keras-unet

BarkBeetleMay BarkBeetleJun BarkBeetleJul BarkBeetleAug BarkBeetleSept BarkBeetleOct

Figure 3: RGB of Sentinel-2 images acquired for one of the testing scenes of the BarkBeetle case study in May 2018 (BarkBeetleMay), June 2018 (BarkBeetleJun), July 2018 (BarkBeetleJul), August 2018 (BarkBeetleAug), September 2018 (BarkBeetleSept) and October 2018 (BarkBeetleOct), respectively.



WildFiresPre WildFiresPost

Figure 4: RGB of Sentinel-2 images acquired for one of the testing scenes of WildFires before the activation date of the forest wildfire (WildFiresPre) and after the activation date (WildFiresPost), respectively.

containing a BN layer and a 2D Convolutional layer per block. For each Convolutional layer, the kernel size was set equal to 3×3. In addition, a 2×2 Max-pooling operation was used with stride equal to 2 for downsampling. According to the experimental setting reported in [12], the kernel size of the ECA module was set equal to 3.

Student architectures were implemented as Fully Convolutional Networks starting from upsampling blocks of the teacher backbone. Each student includes (1) a variable number of transposed Convolution layers, (2) a block with a BN layer, a 2D Convolutional layer and the Dropout layer, (3) a 2D Convolutional Layer and (4) the semantic segmentation layer. The transposed Convolution layers are used to reconstruct the output of the starting upsampling block to the size $32 \times 32 \times 64$. Each transposed Convolution layer doubles the width and height of the previous layer. So, the number of transposed Convolution layers in each student depends on the size of the starting upsampling block. Specifically, TITANIA integrated 3 students that were placed in the 3rd, 4th and 5th upsampling blocks (see Figure 1). In this study, experiments were performed to evaluate the effectiveness of this decision (see Section IV-D1). The dropout layer was used to perform data regularisation and prevent overfitting.

In all layers, with the exception of the semantic segmentation layers of both teacher and students, the Rectified Linear Unit function (ReLU) was used as the activation function. The Sigmoid activation function was used in the semantic segmentation layers.

Table I: Hyper-parameter search space

Hyper-parameter	Values
Mini-batch size	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
Augmentation	{True, False}
Learning rate	[0.0001, 0.01]
Dropout rate	[0, 1]
α	[0, 1]
τ	[1, 30]

The training stage of TITANIA was performed with imagery tiles extracted on scene tiles with size 32×32. These tiles were extracted from the training Sentinel-2 images by using tiler library.[8]. In addition, an imagery tile augmentation strategy was taken into account, to improve the performance of the model. The augmentation strategy was implemented by using the Albumentations library [9]. Specifically, we quintupled the number of training imagery tiles by creating new tiles with traditional computer vision transformation operators (i.e., Horizontal Flip, Vertical Flip, Random Rotate, Transpose and Grid Distortion).

The hyper-parameters of the final TITANIA model were optimized using the tree-structured Parzen estimator algorithm, implemented in the Hyperopt library. The optimization was performed using a random stratified split of 20% of the entire training imagery tile set as a validation set. The hyper-parameter configuration, that achieved the highest F1 score on the validation set, was automatically selected. The hyper-parameter search space explored in this study is reported in Table I. The TITANIA model was trained with mini-batches by back-propagation, with a maximum number of epochs set equal to 150. An early-stopping approach was used to retain the best semantic segmentation model. The Adam update rule was adopted for the gradient-based optimization.

C. Accuracy metrics

To evaluate the accuracy performance of scene maps produced through semantic segmentation we measured both F1 and IoU scores. The F1 score is commonly used in the evaluation of the accuracy of models trained in both classification

[8]https://github.com/the-lay/tiler
[9]https://albumentations.ai/

and semantic segmentation problems. This is the harmonic mean of Precision and Recall. The Precision=$\frac{TP}{TP+FP}$ is the fraction of pixels correctly classified in a specific class (TP) among pixels of the considered class ($TP+FP$). The Recall=$\frac{TP}{TP+FN}$ is the fraction of pixels correctly classified in a specific class (TP) among pixels classified in the considered class ($TP+FN$). In this study, we computed the F1 score for the two opposite classes of both case studies: “healthy” (h) and “damaged” (d). The IoU (Intersection over Union) score is the ratio of the intersected area to the combined area of prediction and ground truth, that is, IoU=$\frac{TP}{TP+FP+FN}$. This is commonly used in the evaluation of the accuracy of models trained in both semantic segmentation and object detection problems. All metrics are reported in percentages. The higher the values of F1 and IoU, the more accurate the evaluated model. All these metrics were computed on the semantic segmentation maps produced for the testing scenes using the models trained on the training scenes.

D. Results and discussion

The experimental study is organized as follows. We start performing a quantitative analysis of the accuracy performance of TITANIA. This analysis includes:

1) A sensitivity study to assess the effect of both position and number of students on the proposed architecture (Section IV-D1).
2) An ablation study to analyse the effect of both attention and distillation on the proposed architecture (Section IV-D2).
3) A comparative study to compare the performance of self-distillation to that of traditional knowledge distillation (Section IV-D3).
4) A temporal study of the performance of segmentation models trained and tested on a collection of Sentinel-2 images acquired in the study scenes at different timestamps (Section IV-D4).

In BarkBeetle, experiments (1)-(4) were performed with dataset BarkBeetleOct, while experiment (5) was performed by considering datasets: BarkBeetleMay,BarkBeetleJun, BarkBeetleJul, BarkBeetleAug, BarkBeetleSept and BarkBeetleOct, respectively. In WildFires, experiments (1)-(4) were performed with dataset WildFiresPost, while experiment (5) was performed with datasets: WildFiresPre and WildFiresPost, respectively.

Specifically, in both case studies, experiments (1)-(4) aim to explore the ability of the proposed method to perform the inventory of forest tree dieback caused by either bark beetle outbreaks or wildfires. The inventory task aims to prompt outbreak or wildfire modelling, particularly in relation to climate and pollution, as well as to trigger post-outbreak forest management. In particular, with regard to bark beetle outbreaks, from the autumn until the following spring, the dieback process commonly decelerates due to the dormancy of bark beetles. So, according to some communications with foresters reported by [4], the beginning of the autumn, i.e., October in this study, may be considered the most suitable period for proactive measures, i.e., for looking for areas of infested trees and removing them from the forest before next spring. In addition, large-scale inventory results may support locating undiscovered previous breeding sites and prompting the application of field monitoring in the vicinity of the detected hotspots, to identify recent, early-stage infestations in the neighbour area [38], [14]. On the other hand, experiment (5) allows us to perform a preliminary investigation of the performance of the proposed method in predicting future bark beetle outbreaks or forest fires based on data enclosed in Sentinel-2 images.

Subsequently, we perform a qualitative analysis of performances by analysing some semantic segmentation maps produced by TITANIA in the two considered case studies (Section IV-D5). We perform this map analysis referred to testing scenes within the different collections of Sentinel-2 images considered in the temporal exploration of the performance done in both case studies.

1) Exploring position and number of shallow students: To explore the joint effect of both the position and the number of students on the performance of TITANIA, we considered the following schemes:

- Deep Student Schema: students are placed in the deeper layers (the nearest to the output) of the decoder of the U-Net architecture. This schema was tested with the number of students varying among 1, 2, 3, 4 and 5.
- Shallow Student Schema: students are placed in the first layers of the decoder of the U-Net architecture. This schema was tested with the number of students varying among 1, 2, 3 and 4. As the decoder of the U-Net of TITANIA comprises five layers, the Shallow Student Schema with 5 students is the same as the Deep Student Schema with 5 students.
- Uniform Student Schema: students are distributed in the decoder of the U-Net architecture following a uniform distribution. This schema was tested with the number of students equal to 2 or 3.

Table II reports F1(h), F1(d) and IoU measured in experiments performed on both BarkBeetleOct and WildFiresPost by using TITANIA with the three student schemes described above. In BarkBeetleOct, the tested configurations of Shallow Student Schema achieve similar performances independently of the number of students. The same outcome is observed, in both datasets, for the tested configurations of Uniform Student Schema. Instead, in both datasets, Deep Student Schema configurations show some variability in the performance especially in F1(d) an IoU. However, in both datasets, the intermediate configuration of Deep Student Schema, i.e., the configuration with three students placed in the last three layers of the decoder of the U-Net architecture, achieves the highest performance of this comparative study in all the three metrics. This can be explained by considering that the Deep Student Schema places the shallow students in the deeper layers (the nearest to the output) of the decoder of the U-Net architecture. In this way, shallow models can handle a deeper representation of input data for self-distillation. Specifically, the data representation inputted to the shallow models passes across initial layers of the decoder before being processed through self-distillation. Our feeling is that

Table II: Accuracy performance of TITANIA measured on the testing Sentinel-2 images of BarkBeetleOct and WildFiresPost for models trained on the training set of BarkBeetleOct and WildFiresPost by TITANIA considering the Deep Student Schema, Shallow Student Schema and Uniform Student Schema to define the position and the number of students. (*) denotes the default configuration of TITANIA. The best results are in bold.

Schema	Student position	BarkBeetleOct			WildFiresPost		
		F1(h)	F1(d)	IoU	F1(h)	F1(d)	IoU
Deep Student Schema	{5}	99.47	66.59	49.92	97.19	86.74	76.59
	{5,4}	99.54	67.18	50.58	97.91	89.50	80.98
	{5,4,3}*	**99.51**	**68.11**	**51.65**	**98.39**	**91.45**	**84.24**
	{5,4,3,2}	99.48	66.99	50.36	97.42	87.12	77.18
	{5,4,3,2,1}	99.50	66.60	49.92	97.92	88.67	79.65
Shallow Student Schema	{1}	99.44	64.15	47.21	95.67	81.70	69.06
	{1,2}	99.55	64.69	47.81	97.26	84.07	72.52
	{1,2,3}	99.44	64.20	47.28	96.10	82.40	70.07
	{1,2,3,4}	99.48	64.29	47.37	97.34	87.38	77.59
Uniform Student Schema	{1,3,5}	99.52	66.83	50.19	97.49	88.29	79.04
	{2,4}	99.49	66.53	49.85	97.78	89.44	80.90

Table III: Accuracy performance of TITANIA and its variants: U-Net, Attention U-Net, Self-distilled U-Net. The best results are in bold

Configuration	BarkBeetleOct			WildFiresPost		
	F1(h)	F1(d)	IoU	F1(h)	F1(d)	IoU
U-Net	99.52	61.79	44.71	95.81	82.36	70.00
Attention U-Net	99.48	65.93	49.18	96.91	86.16	75.69
Self-distilled U-Net	99.50	67.98	51.49	98.22	91.13	83.71
TITANIA	**99.52**	**68.11**	**51.65**	**98.39**	**91.45**	**84.24**

this can contribute to better learn complex relations among spectral data by boosting a gain in the capability of separating the two classes. This configuration is considered the default configuration of TITANIA (as reported in the implementation details), and it is used in the remaining experiments of this study.

2) Ablation architecture analysis: TITANIA combines U-Net, channel attention and self-distillation. To explore the effect of each component on the performance of TITANIA, we perform an ablation study that includes the following baselines of TITANIA:

- U-Net that trains the standard U-Net architecture.
- Attention U-Net that trains the U-Net architecture with channel attention.
- Self-distilled U-Net that trains the standard U-Net self-distilled with three students placed in the deeper layers of the decoder according to the Deep Student Schema.

Table III reports F1(h), F1(d) and IoU measured in the experiments performed on both BarkBeetleOct and WildFiresPost with TITANIA and its baselines: U-Net, Attention U-Net and Self-distilled U-Net. In both datasets, both the channel attention and the self-distillation strategy improve the performance of the standard U-Net architecture along the three metrics. In general, the gain of performance achieved with the self-distillation strategy is greater than the gain achieved with the channel attention. This confirms the intuition that integrating a self-distillation approach into the U-Net architecture can help us to improve the performance of the final semantic segmentation model accounting for the ensemble of predictions yielded from both the teacher and the shallow student. This result is coherent with multiple studies performed in various fields [39], [40] showing that ensemble learning commonly outperforms single model learning. In addition, the channel attention integrated into TITANIA allows us to gain accuracy thanks to its ability to capture the local cross-channel relationships between every channel and its neighbour channels [12]. In conclusion, the ablation study shows that the joint use of channel attention and self-distillation strategy allows us to achieve better semantic segmentation performance in both datasets.

3) Self-distillation versus Knowledge distillation: To analyse the effectiveness of adopting self-distillation in semantic segmentation problems, we compare the performance of the self-distillation strategy currently realized in TITANIA to that of the traditional knowledge distillation strategy [35]. In this study, we consider the competitor KD that performed the knowledge distillation strategy with the Attention U-Net network as teacher and a tiny Fully Convolutional network as student. In particular, the student of KD shares the same architecture as the students of TITANIA. According to the theory reported in [35], first the teacher of KD was pre-trained and, then the student was trained jointly with the teacher by optimizing the KD loss. The student was, finally, used to produce the segmentation maps of testing Sentinel-2 images.

Figure 5 shows F1(h), F1(h) and IoU measured in experiments performed on both BarkBeetleOct and WildFiresPost with TITANIA and its competitor KD. In both datasets, the self-distillation strategy outperforms the traditional knowledge distillation strategy in all the collected metrics. Notably, this is coherent with the conclusions drawn in [31] for classification models, which suggested that training two separate models can lead to a degradation of the performance of the classifier compared to the use of the self-distillation strategy. This degradation of accuracy may depend on a gap between the teacher's model and the student's model due to architectural and capability differences, which may lead to a degradation of performance in terms of accuracy. In fact, the knowledge

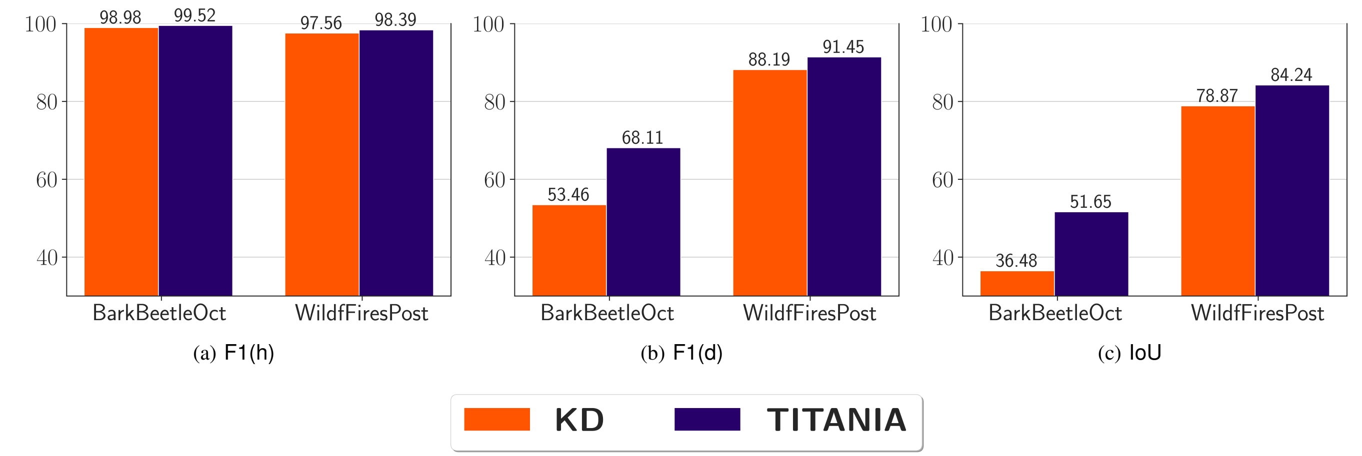


Figure 5: F1(h), F1(d) and IoU of KD and TITANIA

Table IV: Accuracy performance of semantic segmentation produced with Attention U-Net and TITANIA in the Sentinel-2 imagery collections of study scenes acquired over time. For each timestamped imagery collection, the best results are in bold.

Imagery set	Attention U-Net			TITANIA		
	F1(h)	F1(d)	IoU	F1(h)	F1(d)	IoU
BarkBeetleMay	98.19	27.58	15.99	**98.61**	**32.10**	**19.12**
BarkBeetleJun	96.84	20.09	11.17	**98.54**	**32.32**	**19.28**
BarkBeetleJul	98.09	26.47	15.26	**99.07**	**33.37**	**20.03**
BarkBeetleAug	98.92	39.04	24.25	**99.23**	**45.20**	**29.20**
BarkBeetleSept	99.43	61.39	44.29	**99.48**	**64.21**	**47.28**
BarkBeetleOct	99.48	65.93	49.18	**99.52**	**68.11**	**51.65**
WildFiresPre	87.99	50.36	33.66	**89.81**	**55.17**	**38.09**
WildFiresPost	96.91	86.16	75.69	**98.39**	**91.45**	**84.24**

distillation uses a smaller and less complex student model than the teacher model for the final decision. On the other hand, in the self-distillation strategy tested in this study, both the teacher and the shallow student models are trained together and involved in the same semantic segmentation task.

4) Temporal study: We perform a temporal study, to explore the performance of the semantic segmentation outcome produced with the Sentinel-2 image collections acquired for study scenes in different time periods before the production of the ground truth map of the disturbance events causing the forest tree dieback. In this analysis, we focus on the effect of self-distillation. So, we compare the performance of TITANIA and its baseline Attention U-Net, that takes away the self-training strategy for training the semantic segmentation model. Both methods were trained and tested on the imagery collections: BarkBeetleMay, BarkBeetleJun, BarkBeetleJul, BarkBeetleAug, BarkBeetleSept and BarkBeetleOct of the BarkBeetle case study, and on the imagery collections: WildFiresPre and WildFiresPost of the WildFires case study. This analysis allows us to explore the ability of training semantic segmentation models which are able to predict where forest disturbance events are likely to cause future tree dieback.

Table IV reports F1(h), F1(d) and IoU collected in the temporal study performed in the two case studies. As expected the performance of both methods improves as the Sentinel-2 images used for training and testing are acquired closer to the period in which the study tree dieback truth map was observed. In any case, TITANIA systematically outperforms its baseline Attention U-Net regardless of the period in which the Sentinel-2 images were acquired. These results show that despite the prediction of the early stages of the forest tree decay is a difficult task to address using satellite imagery only, the use of the self-distillation approach of TITANIA proves to be more robust than the Attention U-Net even in the early detection task. Our feeling is that this result is achieved thanks to the ability of self-distillation to help TITANIA models find flat minima, by allowing better generalization performances [41].

By focusing on the case study regarding the detection of the forest tree dieback caused by bark beetle infestation hotspots, in a recently published paper [14], we have illustrated our previous work describing a machine learning approach for bark beetle-induced tree dieback in Sentinel-2 images. This approach, named SILVIA, integrates classification, spectral vegetation indices and self-learning. As SILVIA was tested in the same case study in the Northeast of France evaluated in this work, we consider it as a related method for this case study. The experimentation illustrated in [14] was performed using a preliminary version of the Sentinel-2 dataset of the study scenes, which was created for October 2018 by selecting 10 bands only (and excluding Red Edge 2, Narrow NIR and SWIR-CIRRUS). To achieve a fair comparison, we ran SILVIA on the complete imagery collections BarkBeetleMay, BarkBeetleJun, BarkBeetleJul, BarkBeetleAug, BarkBeetleSept and BarkBeetleOct. Figure 6 shows the IoU measured in these imagery collections with TITANIA and the related method SILVIA. Notably, these results show that TITANIA systematically outperforms SILVIA in all the months of the study. We point out that the self-learning step of SILVIA was performed with the XGBoost classifier, while TITANIA implemented a deep learning-based approach enriched with self-distillation and attention. In general, deep learning algorithms

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3460981

have been proven to be more effective than traditional machine learning strategies thanks to their ability to learn non-linear dependencies between features.

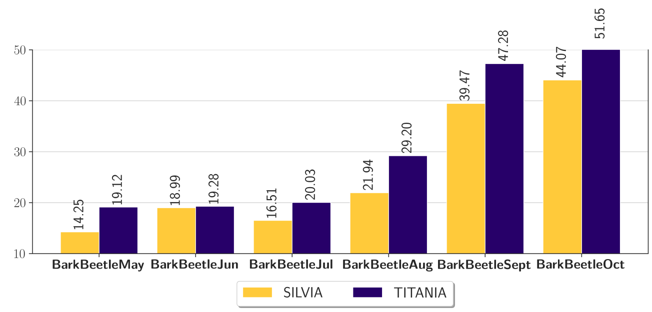


Figure 6: IoU measured with SILVIA and TITANIA in the Sentinel-2 imagery collections of the BarkBeetle study scenes acquired from May 2018 to October 2018.

5) Semantic segmentation map analysis: To complete this investigation we analyse some semantic segmentation maps produced in both the considered case studies. Specifically, we analyse segmentation maps produced in the ablation study (Section IV-D2) with models trained by TITANIA and its three baselines: U-Net, Attention U-Net and Self-distilled U-Net, respectively. In addition, focusing the attention on the effect of self-distillation, we analyse the semantic segmentation maps produced with both TITANIA and Attention U-Net in the temporal study (Section IV-D4) of both case studies.

With regard to the BarkBeetle case study, Figure 7 reports the semantic segmentation maps produced for the testing sample scene shown in Figure 3 with models trained by U-Net, Attention U-Net, Self-distilled U-Net and TITANIA, respectively. These maps were produced in the ablation experiments conducted on BarkBeetleOct. The maps show that all the methods of the ablation study are able to identify the three large damaged patches which appear in the ground truth map of the bark beetle outbreaks of this scene. In addition, all the methods alert the same false damaged patch. On the other hand, TITANIA is able to discover four small damaged patches that are neglected by both U-Net and Attention U-Net. Notably, three of these small damaged patches are also neglected by Self-distilled U-Net. Hence, this visual analysis shows that the proposed combination of U-Net, attention and self-distillation can actually contribute to detecting some tree dieback patches caused by bark beetle infestation, which are unseen in the output of the ablation configurations.

Continuing the analysis of the segmentation maps produced in the BarkBeetle case study, Figure 8 shows the semantic segmentation maps produced by both TITANIA and Attention U-Net in the temporal study for the testing sample scene reported in Figure 3. Specifically, these maps were produced in the experiments conducted on the Sentinel-2 image collections: BarkBeetleMay, BarkBeetleJun, BarkBeetleJul, BarkBeetleAug, BarkBeetleSept and BarkBeetleOct, respectively. The segmentation maps show that considering the ground truth maps as bark beetle outbreaks will appear in October 2018, the number of false alarms (pixels wrongly labelled as damaged by the infestation) is higher the further the processed images are acquired from October 2018. However, the false alarm patches are less extensive with TITANIA than with Attention U-Net. Finally, in all the study months the patches alerted as possibly infested are close to the areas where the bark beetle outbreaks will be visible in October 2018.

With regard to the WildFires case study, Figure 10 reports the semantic segmentation maps produced for the testing sample scene shown in Figure 4. These maps were produced with models trained by U-Net, Attention U-Net, Self-distilled U-Net and TITANIA, respectively in the ablation experiment conducted on the Sentinel-2 imagery collection WildFires-Post. The use of the self-distillation strategy allows us to remove the presence of false salt-and-pepper alerts which appear in the semantic segmentation maps produced with both U-Net and Attention U-Net. On the other hand, TITANIA delimit the wildfire areas slightly better than Self-distilled U-Net.

Finally, Figure 10 reports the semantic segmentation maps produced for the study testing scene shown in Figure 4 by Attention U-Net and TITANIA in the temporal study performed with WildFiresPre and WildFiresPost, respectively. TITANIA takes advantage of the self-distillation strategy and delimits the damaged area in the image acquired one month after the wildfire activation better than Attention U-Net. In addition, TITANIA removes false salt-and-pepper alarms in the map produced in the experiment performed on WildFiresPost. On the other hand, although TITANIA fails to remove false salt-and-pepper alarms in the Sentinel-2 image acquired before the wildfire activation, it is still able to outperform Attention U-Net by increasing the number of true alarms in the experiment performed on WildFiresPre.

V. CONCLUSION

In this study we have explored the effect of performing a self-distillation strategy and leveraging a channel attention mechanism in a U-Net architecture trained to perform the inventory of forest tree dieback in Sentinel-2 images. The effectiveness of the proposed method TITANIA was evaluated in two case studies which refer to forest scenes damaged by bark beetle and wildfires, respectively. In particular, the effectiveness of TITANIA was quantified in terms of semantic segmentation accuracy. The obtained results have shown that TITANIA is able to outperform ablation approaches that discard the self-distillation strategy and/or the attention mechanism. Furthermore, the experimental results have clearly highlighted that the self-distillation strategy outperforms the traditional knowledge distillation strategy. Finally, self-distillation allows us to gain accuracy compared to the baseline segmentation models also in the early detection task.

Possible future research directions could be devoted to equip TITANIA with data fusion mechanisms, to process simultaneous observation data acquired for the same forest scenes from different sources (e.g., Sentinel-1 and Sentinel-2 images, meteorological and pollution parameters, terrestrial data). This multi-source approach may allow us to further improve the performance of TITANIA particularly in the early detection task. In fact, as recently reported in [42], terrestrial monitoring

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3460981

(a) U-Net

(b) Attention U-Net

(c) Self-distilled U-Net

(d) TITANIA

Figure 7: Predicted inventory of the bark beetle-induced forest dieback in the sample scene of the testing set shown in Figure 4. The maps were produced with U-Net (7a), Attention U-Net (7b), Self-distilled U-Net (7c) and TITANIA (7d) experimented using the imagery collection BarkBeetleOct. The white pixels are the ones predicted in the "damaged" class. The yellow polygons delimit the boundary of the ground truth hotspots of the bark beetle infestation in this sample scene.

may significant help in detecting early crown discoloration caused by bark beetle infestation. On the other hand, [43] has recently discussed the crucial role of climate information to prevent and promptly fight fires.

A further research direction refers to the exploration of the transferability of the semantic segmentation model trained for a specific disturbance in a specific geographic area and time period to different geographic areas or time periods. In addition, we intend to explore possible transfer learning techniques to update the semantic segmentation model trained to map forest tree dieback caused by a specific disturbance agent to a different disturbance agent. With this regard, we also plan to explore the performance of the proposed method in presence of multiple disturbance agents.

Acknowledgments

Giuseppina Andresini is supported by the project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI, under the NRRP MUR program funded by the NextGenerationEU. Annalisa Appice and Donato Malerba acknowledge support from the SWIFTT project, funded by the European Union under Grant Agreement 101082732. The authors wish to thank the remote sensing company WildSense for preparing the ground truth maps of the bark beetle case study and Hanna Yailymova and Natalya Gordiiko of the Space Research Institute of the National Academy of Sciences of Ukraine for downloading the collections of Sentinel-2 images processed in the bark beetle case study. The authors wish to thank anonymous reviewers for the helpful suggestions they provided to improve the paper.

CRediT Authorship Contribution Statement

Giuseppina Andresini: Conceptualization, Methodology, Software, Data curation, Investigation, Validation, Visualization, Writing - original draft, Writing - review & editing. **Annalisa Appice**: Conceptualization, Methodology, Software, Investigation, Validation, Supervision, Writing - original draft, Writing - review & editing. **Donato Malerba**: Conceptualization, Investigation, Writing - review & editing.

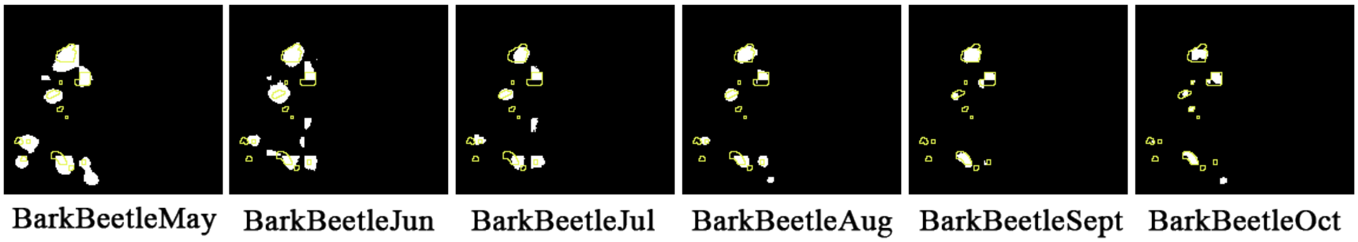
References

[1] A. Ali, "Linking forest ecosystem processes, functions and services under integrative social–ecological research agenda: current knowledge and perspectives," *Science of The Total Environment*, vol. 892, p. 164768, 2023.

[2] A. Requardt, A. Schuck, and M. Kohl, "Means of combating forest dieback - EU support for maintaining forest health and vitality," *iForest - Biogeosciences and Forestry*, no. 1, pp. 38–42, 2009.

[3] M. Kovac, P. Gasparini, M. Notarangelo, M. Rizzo, I. Cañellas, L. F. de Uña, and I. Alberdi, "Towards a set of national forest inventory indicators to be used for assessing the conservation status of the habitats directive forest habitat types," *Journal for Nature Conservation*, vol. 53, p. 125747, 2020.

(a) U-Net



(b) TITANIA

Figure 8: Predicted inventory maps of the bark beetle-induced tree dieback in the sample scene on the testing set shown in Figure 3. Maps were produced by Attention U-Net and TITANIA in the temporal study conducted using the imagery data collections: BarkBeetleMay, BarkBeetleJun, BarkBeetleJul, BarkBeetleAug, BarkBeetleSept and BarkBeetleOct, respectively.

[4] V. Bárta, P. Lukeš, and L. Homolová, "Early detection of bark beetle infestation in norway spruce forests of central europe using sentinel-2," *International Journal of Applied Earth Observation and Geoinformation*, vol. 100, p. 102335, 2021.

[5] A. M. Lechner, G. M. Foody, and D. S. Boyd, "Applications in remote sensing to forest ecology and management," *One Earth*, vol. 2, no. 5, pp. 405–412, 2020.

[6] F. E. Fassnacht, J. C. White, M. A. Wulder, and E. Næsset, "Remote sensing in forestry: current challenges, considerations and directions," *Forestry: An International Journal of Forest Research*, pp. 1–27, 2023.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[8] S. Ge, H. Gu, W. Su, J. Praks, and O. Antropov, "Improved semisupervised unet deep learning model for forest height mapping with satellite sar and optical data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 5776–5787, 2022.

[9] D. John and C. Zhang, "An attention-based U-Net for detecting deforestation within satellite sensor imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, p. 102685, 2022.

[10] A. U. Waldeland, Øivind Due Trier, and A.-B. Salberg, "Forest mapping and monitoring in Africa using Sentinel-2 data and deep learning," *International Journal of Applied Earth Observation and Geoinformation*, vol. 111, p. 102840, 2022.

[11] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *International Conference on Learning Representations, ICLR 2015*, 2015, pp. 1–15.

[12] W. Qilong, W. Banggu, Z. Pengfei, L. Peihua, Z. Wangmeng, and H. Qinghua, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *onference on Computer Vision and Pattern Recognition, CVPR 2020*. IEEE, 2020, pp. 11 531–11 539.

[13] L. Zhang, C. Bao, and K. Ma, "Self-distillation: Towards efficient and compact neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4388–4403, 2022.

[14] G. Andresini, A. Appice, and D. Malerba, "SILVIA: An explainable framework to map bark beetle infestation in Sentinel-2 images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 10 050–10 066, 2023.

[15] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019.

[16] J. S. Estrada, A. Fuentes, P. Reszka, and F. Auat Cheein, "Machine learning assisted remote forestry health assessment: a comprehensive state of the art review," *Frontiers in Plant Science*, vol. 14, 2023.

[17] A. Waśniewski, A. Hościło, B. Zagajewski, and D. Moukétou-Tarazewicz, "Assessment of Sentinel-2 satellite images and random forest classifier for rainforest mapping in gabon," *Forests*, vol. 11, no. 9, 2020.

[18] R. Gibson, T. Danaher, W. Hehir, and L. Collins, "A remote sensing approach to mapping fire severity in South-eastern Australia using Sentinel 2 and random forest," *Remote Sensing of Environment*, vol. 240, p. 111702, 2020.

[19] A. Candotti, M. De Giglio, M. Dubbini, and E. Tomelleri, "A sentinel-2 based multi-temporal monitoring framework for wind and bark beetle detection and damage mapping," *Remote Sensing*, vol. 14, no. 23, 2022.

[20] A. H. Y. Chan, C. Barnes, T. Swinfield, and D. A. Coomes, "Monitoring ash dieback (hymenoscyphus fraxineus) in British forests using hyperspectral remote sensing," *Remote Sensing in Ecology and Conservation*, vol. 7, no. 2, pp. 306–320, 2021.

[21] B. T. Fraser and R. G. Congalton, "Monitoring fine-scale forest health using unmanned aerial systems (UAS) multispectral models," *Remote Sensing*, vol. 13, no. 23, 2021.

[22] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proceedings of 10th European Conference on Machine Learning: ECML-98*, ser. Lecture Notes in Computer Science, C. Nédellec and C. Rouveirol, Eds., vol. 1398. Berlin, Heidelberg: Springer, 1998, pp. 137–142.

[23] R. Migas-Mazur, M. Kycko, T. Zwijacz-Kozica, and B. Zagajewski, "Assessment of Sentinel-2 images, support vector machines and change detection algorithms for bark beetle outbreaks mapping in the tatra mountains," *Remote Sensing*, vol. 13, no. 16, 2021.

[24] S. Bar, B. R. Parida, and A. C. Pandey, "Landsat-8 and Sentinel-2 based forest fire burn area mapping using machine learning algorithms on gee cloud platform over uttarakhand, western himalaya," *Remote Sensing Applications: Society and Environment*, vol. 18, p. 100324, 2020.

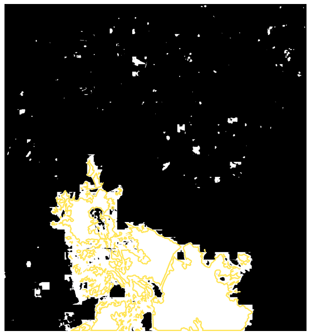
[25] J. Zhang, S. Cong, G. Zhang, Y. Ma, Y. Zhang, and J. Huang, "Detecting pest-infested forest damage through multispectral satellite imagery and improved unet++," *Sensors*, vol. 22, no. 19, 2022.

[26] J. Pyo, K.-j. Han, Y. Cho, D. Kim, and D. Jin, "Generalization of U-Net semantic segmentation for forest change detection in South Korea using airborne imagery," *Forests*, vol. 13, no. 12, 2022.

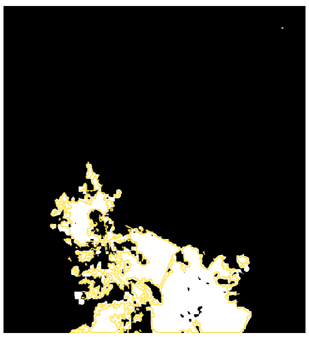
[27] X. Liu, J. Frey, M. Denter, K. Zielewska-Büttner, N. Still, and B. Koch, "Mapping standing dead trees in temperate montane forests using a

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2024.3460981

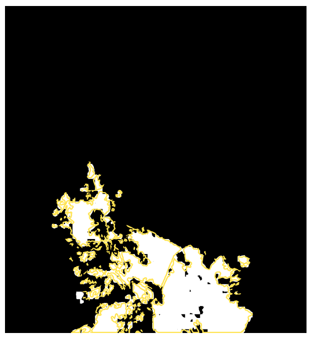
(a) U-Net



(b) Attention U-Net

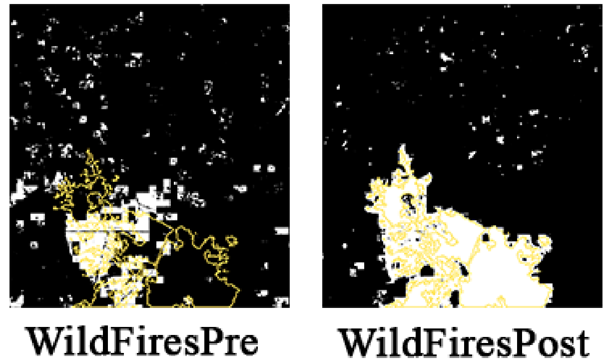


(c) Self-distilled U-Net

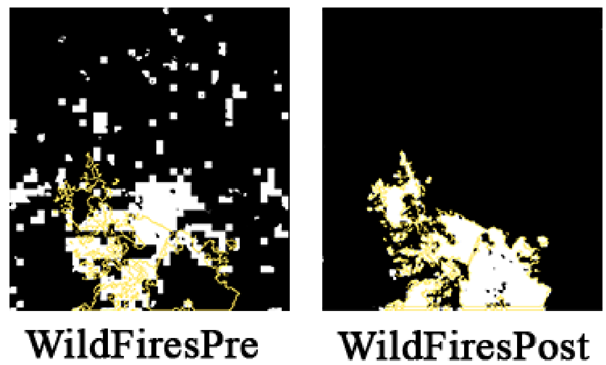


(d) TITANIA

Figure 9: Predicted inventory of the forest tree dieback due to wildfires in the sample scene of the testing set shown in Figure 4. The maps were produced with U-Net (9a), Attention U-Net (9b), Self-distilled U-Net (9c) and TITANIA (9d). The white pixels are the ones predicted in the "damaged" class. The yellow polygons delimit the boundary of the ground truth of the wildfire damages in the sample scene.



(a) Attention U-Net



(b) TITANIA

Figure 10: Predicted inventory maps of the forest tree dieback due to wildfires in the sample scene on the testing set shown in Figure 4. Maps were produced by Attention U-Net and TITANIA in the temporal study conducted using the imagery data collections WildFiresPre and WildFiresPost, respectively.

pixel- and object-based image fusion method and stereo worldview-3 imagery," *Ecological Indicators*, vol. 133, p. 108438, 2021.

[28] H. T. Nguyen, M. L. Lopez Caceres, K. Moritake, S. Kentsch, H. Shu, and Y. Diez, "Individual sick fir tree (abies mariesii) identification in insect infested forests by means of uav images and deep learning," *Remote Sensing*, vol. 13, no. 2, 2021.

[29] A. Sani-Mohammed, W. Yao, and M. Heurich, "Instance segmentation of standing dead trees in dense forest from aerial imagery using deep learning," *ISPRS Open Journal of Photogrammetry and Remote Sensing*, vol. 6, p. 100024, 2022.

[30] A. M. Al-Dabbagh and M. Ilyas, "Uni-temporal Sentinel-2 imagery for wildfire detection using deep learning semantic segmentation models," *Geomatics, Natural Hazards and Risk*, vol. 14, no. 1, p. 2196370, 2023.

[31] L. Zhang, C. Bao, and K. Ma, "Self-distillation: Towards efficient and compact neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4388–4403, 2022.

[32] N. Wang, S. Lin, X. Li, K. Li, Y. Shen, Y. Gao, and L. Ma, "MISSU: 3D medical image segmentation via self-distilling transunet," *IEEE Trans. Medical Imaging*, vol. 42, no. 9, pp. 2740–2750, 2023.

[33] I. Ahmed, A. Chehri, and G. Jeon, "A sustainable deep learning-based framework for automated segmentation of COVID-19 infected regions: Using U-Net with an attention mechanism and boundary loss function," *Electronics*, vol. 11, no. 15, 2022.

[34] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in *Machine Learning in Medical Imaging*, Q. Wang, Y. Shi, H.-I. Suk, and K. Suzuki, Eds. Cham: Springer International Publishing, 2017, pp. 379–387.

[35] G. E. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," CoRR, abs/1503.02531, pp. 1–9, 2015.

[36] D. Nardi, H. Jactel, E. Pagot, J.-C. Samalens, and L. Marini, "Drought and stand susceptibility to attacks by the european spruce bark beetle: A remote sensing approach," *Agricultural and Forest Entomology*, vol. 25, no. 1, pp. 119–129, 2023.

[37] L. Colomba, A. Farasin, S. Monaco, S. Greco, P. Garza, D. Apiletti, E. Baralis, and T. Cerquitelli, "A dataset for burned area delineation and severity estimation from satellite imagery," in *31st ACM International Conference on Information & Knowledge Management, CIKM 2022*. AM, 2022, p. 3893–3897.

[38] R. Duraciova, M. Munko, I. Barka, M. Koren, K. Resnerova, J. Holusa, M. Blazenec, M. Potterf, and R. Jakus, "A bark beetle infestation predictive model based on satellite data in the frame of decision support system TANABBO," *iForest - Biogeosciences and Forestry*, no. 3, pp. 215–223, 2020.

[39] M. AL-Essa, G. Andresini, A. Appice, and D. Malerba, "Panacea: a neural model ensemble for cyber-threat detection," *Machine Learning*, 2024.

[40] Y. Zhang, J. Liu, and W. Shen, "A review of ensemble learning algorithms used in remote sensing applications," *Applied Sciences*, vol. 12, no. 17, 2022.

[41] L. Zhang, J. Song, A. Gao, J. Chen, C. Bao, and K. Ma, "Be your own teacher: Improve the performance of convolutional neural networks via self distillation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 3712–3721.

[42] M. Kautz, F. Peter, L. Harms, S. Kammen, and H. Delb, "Patterns, drivers and detectability of infestation symptoms following attacks by the european spruce bark beetle," *Journal of Pest Science*, vol. 96, p. 403–414, 2022.

[43] U. Oliveira1, B. Soares-Filho, H. Rodrigues, D. Figueira, L. Gomes, W. Leles, C. Berlinck, F. Morelli, M. Bustamante, J. Ometto, and H. Miranda, "A near real-time web-system for predicting fire spread across the cerrado biome," *Scientific Reports*, vol. 13, no. 4829, pp.

1–12, 2023.

VI. Biography Section

Giuseppina Andresini received the Ph.D. degree in Computer Science, University of Bari Aldo Moro, Italy, in March 2022. She is an Assistant Professor (non-tenure) at the Department of Computer Science, the University of Bari Aldo Moro. Her current research interests mainly concern deep learning, XAI and adversarial learning with applications in cybersecurity and remote sensing. On these topics, she has published several papers in international journals and international conferences. She is on the program committee of several top conferences (e.g., ECML-PKDD, ICDE, IJCAI, Discover Science, ECAI) and a member of the editorial board of "Machine Learning" Journal.

Annalisa Appice received the Ph.D. degree in computer science from the University of Bari Aldo Moro, Bari, Italy, in 2005. She is a Full Professor at the Department of Computer Science, University of Bari Aldo Moro, Italy. Her current research interests include data mining with spatio-temporal data, data streams and event logs with applications to remote sensing, process mining and cybersecurity. She has has authored and co-authored more than 190 papers in international journals and conferences. She has been the Program co-Chair of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2015, International Symposium on Methodologies for Intelligent System (ISMIS) 2017 and 2024, and Discovery Science (DS) 2020. She is Associate Editor of Engineering Applications of Artificial Intelligence Journal (EAAI) and Machine Learning Journal (MACH). She is a member of the Editorial board of Data Mining and Knowledge Discovery Journal (DAMI) and Journal of Intelligent Information Systems (JIIS). She is responsible for the local research unit of the University of Bari Aldo Moro in the EU project SWIFTT.

Donato Malerba (Senior Member, IEEE) received the M.Sc. degree in computer science from the University of Bari, Italy, in 1987. He is a Full Professor at the Department of Computer Science, University of Bari Aldo Moro, Italy. He is a Full Professor with the Department of Computer Science, University of Bari Aldo Moro, Italy. He has authored and co-authored more than 330 papers in international journals and conferences. His research interests include machine learning, data mining, Big Data analytic, and their applications. He has been responsible for the local research unit of several European and national projects, and was the recipient of an IBM Faculty Award in 2004. He was Program (co-)Chair of International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA-AIE) 2005, International Symposium on Methodologies for Intelligent System (ISMIS) 2006, Symposium on Advanced Database System (SEBD) 2007, and European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2011, and the General Chair of Algorithmic Learning Theory (ALT) and Discovery Science (DS) 2016 and International Joint Conference on Learning & Reasoning (IJCLR) 2023. He is scientific responsible of Spoke 6 - Symbiotic AI of the Future AI Research (FAIR) project. He is in the Editorial Board of several international journals.