DIAMANTE: A data-centric semantic segmentation approach to map tree dieback induced by bark beetle infestations via satellite images

Giuseppina Andresini, Annalisa Appice, Dino Ienco, Vito Recchia

Noname manuscript No.
(will be inserted by the editor)

DIAMANTE: A data-centric semantic segmentation approach to map tree dieback induced by bark beetle infestations via satellite images

Giuseppina Andresini · Annalisa Appice · Dino Ienco · Vito Recchia

Received: date / Accepted: date

Abstract Forest tree dieback inventory has a crucial role in improving forest management strategies. This inventory is traditionally performed by forests through laborious and time-consuming human assessment of individual trees. On the other hand, the large amount of Earth satellite data that is publicly available with the Copernicus program and can be processed through advanced deep learning techniques has recently been established as an alternative to field surveys for forest tree dieback tasks. However, to realize its full potential, deep learning requires a deep understanding of satellite data since the data collection and preparation steps are essential as the model development step. In this study, we explore the performance of a data-centric semantic segmentation approach to detect forest tree dieback events due to bark beetle infestation in satellite images. The proposed approach prepares a multisensor data set collected using both the SAR Sentinel-1 sensor and the optical Sentinel-2 sensor and uses this dataset to train a multisensor semantic segmentation model. The evaluation shows the effectiveness of the proposed approach in a real inventory case study that regards non-overlapping forest scenes from the Northeast of France acquired in October 2018. The selected scenes host bark beetle infestation hotspots of different sizes, which originate from the mass reproduction of the bark beetle in the 2018 infestation.

Keywords Data-centric remote sensing, Sentinel-1 data, Sentinel-2 data, Data fusion, Data imbalance, Semantic segmentation, Bark beetle detection

A. Andresini, A. Appice, V. Recchia
Department of Informatics, Università degli Studi di Bari Aldo Moro, via Orabona, 4 - 70125 Bari - Italy
Consorzio Interuniversitario Nazionale per l'Informatica - CINI, Bari, Italy
E-mail: giuseppina.andresini@uniba.it,
annalisa.appice@uniba.it, vito.recchia@uniba.it

D. Ienco
INRAE, UMR TETIS, University of Montpellier,Montpellier, France
INRIA, University of Montpellier, Montpellier, France
E-mail: dino.ienco@inrae.fr

1 Introduction

Forests and woodlands cover roughly one-third of Earth's surface and play a critical role in providing many ecosystem services, including carbon sequestration, water flow regulation, timber production, soil protection, and biodiversity conservation. However, the accelerating pace of climate change and its impact on species distribution and biome composition are leading to an increase in various types of disturbances, whether biotic, abiotic, or a combination of both, which are now affecting this vital natural resource and resulting in forest loss. Consequently, the decline in key forest ecosystem services is becoming more and more apparent. Among all the disturbances, insect infestations and disease outbreaks (e.g., bark beetle infestations) can induce massive tree dieback and, subsequently, significantly disrupt ecosystem dynamics [18]. This is why forest surveillance is crucial to monitor, quantify and possibly prevent outbreak diseases and enable foresters to perform informed decision-making for effective environmental management. Nevertheless, common strategies used to evaluate the health of forested regions primarily rely on laborious and time-consuming field surveys [8]. Consequently, they are restricted in their ability to cover extensive geographical areas, thereby preventing large-scale analysis across vast territories. To this end, the substantial amount of remote sensing information collected today via modern Earth observation missions constitutes an unprecedented opportunity to scale up forest dieback assessment and surveillance over large areas. As an exemplar, the European Space Agency's Sentinel missions [7] provide a set of quasi-synchronous synthetic aperture radar (SAR) and optical data, systematically acquired worldwide, at high spatial (order of 10m) and temporal (an acquisition up to every five/six days) resolution. This information can be of paramount interest to support large-scale forest dieback assessment and surveillance systems.

While the research community is investigating the benefit related to exploiting multisensor remote sensing information via recent deep learning approaches [23, 29], there is still the necessity to design effective and well-tailored approaches to get the most out of multisensor remote sensing information [22]. This is the case for the large-scale assessment of tree dieback events induced by insect infestations and disease outbreaks where, to the best of our literature survey, existing works (e.g., [4, 8, 9, 15, 17, 45]) mainly focus on optical data analysis, while no works exist that achieve improvements by leveraging multisensor remote sensing data (e.g., SAR and optical data). In particular, the literature studies to monitor bark beetle infestation in optical data pay high attention to both the data engineering step, through the synthesis of spectral vegetation indices, and the model development step, through the test of various machine learning and deep learning algorithms. On the other hand, similar to research communities where data play a major central role (e.g., computer vision, machine learning, information retrieval), also researchers coming from the remote sensing field are investing efforts towards more systematic and effective exploitation of available data sources. To this end, research actions in this direction have been proposed under the umbrella of data-centric Artificial

Intelligence (AI) [43]. Under this movement, the attention of researchers and practitioners is gradually shifting from advancing model design (model-centric AI) to enhancing the quality, quantity and diversity of the data (data-centric AI). Moreover, when remote sensing data are considered, the data-centric AI perspective is even more important since it can steer the community towards developing a methodology to provide further improvements related to the use of highly heterogeneous information to ameliorate the generalization ability with impact on real-world relevant problems and applications [38]. Nevertheless, the two perspectives (model-centric and data-centric AI) play a complementary role in the larger remote sensing deployment cycle, since standard approaches still struggle to manage and exploit valuable data coming from different and heterogeneous sources as, for instance, in the case of leveraging multisensor complementary information.

With the objective to find a trade-off between data-centric and model-centric achievements in remote sensing and map bark beetle-induced tree dieback events in remote sensing data adopting a semantic segmentation approach (e.g., categorization of pixels into a class), in this paper, we propose DIAMANTE (Data-centrIc semAntic segMentation to mAp iNfestations in saTellite imagEs): a data-centric semantic segmentation approach to train a U-Net like model from a labelled remote sensing dataset prepared using both SAR Sentinel-1 (S1) and multi-spectral optical Sentinel-2 (S2) remote sensing data sources. In particular, for the model development, we compare the achievements of several multisensor data fusion schema that are performed via early, middle or late stages fusion in an underlining U-Net architecture [37]. The U-Net is considered thanks to its wide versatility and increasing popularity, as well as due to the fact that it has been recently used to map bark beetle-induced tree dieback in Sentinel-2 images [4,5,45]. In addition, in this study, we consider that model recycling is one of the achievements to be evaluated in developing a data-centric AI approach. Hence, we start a preliminary investigation of how the multisensor fusion approaches considered in this study may allow us to train a semantic segmentation model for bark beetle detection, which still achieves good performance in a future data setting. The following are the main contributions of this work:

- The definition of a remote sensing data collection and curation pipeline to prepare multisensor, Sentinel-1 and Sentinel-2 images of forest areas for which the ground truth map of the bark beetle infestation is available at a specific time. The defined pipeline pays particular attention to the quality of the Sentinel-1 and Sentinel-2 data prepared for the model development.
- The adoption and comparison of several multisensor data fusion schemes to combine Sentinel-1 and Sentinel-2 data via early, middle or late stages fusion considering the underlying U-Net architecture.
- The extensive assessment of our proposal using a ground truth map of tree dieback induced by bark beetle infestations in the Northeast of France in October 2018. The evaluation examines the performance of models trained and tested using images acquired over non-overlapping scenes in the same

period, as well as the temporal forecasting and transferability of the model to an upcoming data setting.

The rest of the manuscript is organized as follows. Related literature is reviewed in Section 2. The proposed methodology is described in Section 4 while the study site and the associated multisensor remote sensing dataset are introduced in Section 3. Section 5 reports the experimental evaluation and it discusses the related findings. Section 6 concludes.

2 Related work

This related work overview is organised into two main fronts. Firstly, we delve into recent remote sensing studies that incorporate machine learning and deep learning to map bark beetle infestation in Sentinel-1 (S1) and Sentinel-2 (S2) images. On the other front, we address the recent achievements of the data-centric artificial intelligence paradigm in remote sensing applications.

2.1 Bark beetle detection in remote sensing

Remote sensing studies to map forest stress related to bark beetle attacks have mainly focused on the analysis of Sentinel-2 data [16]. These studies are mainly inspired by the analysis conducted in [1] to explore the effect of several forest disturbances sources (comprising bark beetle infestation) on S2 data. This study shows that the bark beetle infestation, which may affect the biophysical and biochemical properties of trees, is commonly visible via Sentinel-2 multi-spectral imagery. In particular, the chlorophyll degradation and nitrogen deficiency lead to an increase in reflectance spectrum in the visible region (particularly, red and green bands). Changes caused by the reduction of chlorophyll and leaf water have also an effect on Near Infrared (NIR) and Water vapor bands, while diseased and insect attacks affect red-edge bands. This analysis has boosted a plethora of studies [4,5,8,9,15,17,25,26,45] that explore the ability of various spectral vegetation indices to enhance the accuracy of decision models trained on Sentinel-2 data. Notice that explored spectral vegetation indices mainly combine red, green, NIR and SWIR (short wave infrared) bands.

Regarding the classification algorithms used to map bark beetle infestations in Sentinel-2 images, the most recent studies have mainly used machine learning algorithms such as Random Forest [4,5,8,9,25], Support Vector Machine [4,9,15] and XGBoost [4,5]. Instead, [4,5,45] explore the performance of deep learning algorithms under semantic segmentation settings such as U-Net [4,5,45] and FCN-8 [4]. To handle the data imbalance situation, [4,5,15] use a cost-based learning strategy in combination with Random Forest and Support Vector Machine, while [4,5] use the Tversky loss in combination with U-Net and FCN-8. Finally, some studies consider Sentinel-2 time series data to train either Random Forest [5,8,17] or U-Net models [5].

On the other hand, only recently, few remote sensing studies have started exploring the potential of Sentinel-1 data to detect bark beetle infestations. Sentinel-1 data are traditionally used in deforestation detection on [21]. However, [22] has recently hypothesized that the joint exploitation of Sentinel-1 and Sentinel-2 satellite information can disclose useful information to detect bark beetle infestation hotspots. In particular, this study finds significant differences between Sentinel-1 values measured in infested and healthy sites, respectively. Similar conclusions are drawn in [2]. However, [2,22] perform a statistical analysis of Sentinel-1 data distribution without exploring how the use of the Sentinel-1 information can contribute to learning accurate decision models to characterise bark beetle infestations. In general, based on the literature survey, [22] highlights that significant research effort is still needed to explore the full potential of multisensor data in insect-induced forest disturbance mapping. In this direction, [25] shows that the joint analysis of Sentinel-1 and Sentinel-2 data marginally contributes to improving the performance of Random Forest models. This conclusion has been recently confirmed also by [28] where poor performances have been achieved for bark beetle infestation mapping exploiting only Sentinel-1 radar data and negligible amelioration by the joint exploitation of multisensor (Sentinel-1 and Sentinel-2) data considering both Bayesian and Random Forest classification models. Notably, in [28], the multisensor data are stacked in a single feature vector that is used as input space for training a classification model. This corresponds to an early fusion schema that concatenates pixel-wise the feature vectors which are acquired with the Sentinel-1 and Sentinel-2 sensors before starting the training stage.

On the other hand, some recent studies have started to investigate how to combine multisensor remote sensing data (e.g., Sentinel-1 and Sentinel-2 data) for the underlying task of land use land cover mapping under a semantic segmentation setting [39]. The authors of [29] have surveyed recent deep learning architectures developed to handle multisensor data comprising Sentinel-1 and Sentinel-2 data. However, this survey mainly considers problems of change detection and biomass estimation without any attention to bark beetle detection problems. In addition, this study points out that the majority of deep neural architectures trained with multisensor satellite data adopt an early fusion mechanism to concatenate pixel-wise data acquired with the Sentinel-1 and Sentinel-2 satellites. The output of the concatenation step is subsequently used as input space for the deep neural model development. In particular, the authors of both [31] and [41] learn a U-Net model for land cover classification and flood detection via an early fusion of the Sentinel-1 and Sentinel-2 data. The authors of [3] introduce the Principal Component Analysis (PCA) to combine stacked Sentinel-1 and Sentinel-2 imagery before training a U-Net model for the downstream task of tropical mountain deforestation delineation. On the other hand, a few studies have recently started the investigation of late fusion mechanisms to combine Sentinel-1 and Sentinel-2 data through a deep learning architecture. For example, the authors of [24] describe a two-branch architecture that separately extracts features from data acquired with the two distinct satellites and perform the late convolutional fusion before the final de-

cision. A similar late fusion schema is also investigated in [19] for a problem of urban change detection. This study describes an architecture composed of two separate, identical U-Net architectures that process Sentinel-1 and Sentinel-2 image pairs in parallel, and lately fuses extracted features from both sensors at the final decision stage. A middle fusion mechanism is introduced in [6] to perform the fusion of Infrared-Red-Green (IRRG) images and Digital Surface Model (DSM) data extracted from the Lidar point cloud through a SegNet model. Middle fusion is performed at the encoder layers with a simple summation. Imagery data fusion schemes are also discussed in the survey paper [46].

In any case, to the best of our knowledge, no previous studies have been proposed yet to explore the opportunity of combining Sentinel-1 and Sentinel-2 data via modern deep learning architecture (i.e., U-Net) for the downstream bark beetle detection task. In addition, this is the first study that frames the investigation of different multisensor fusion schemes (i.e., early fusion, middle fusion and late fusion) in a U-Net development step performed under the umbrella of data-centric AI. On the other hand, neither previous studies have experimented with a fusion mechanism that operates at the encoder level of semantic segmentation models trained on Sentinel-1 and Sentinel-2 data, nor these studies have started the investigation of achievements of data fusion schemes for model development done under the possible lens of model recycling.

2.2 Data-centric artificial intelligence in remote sensing

Data plays a fundamental role in several remote sensing problems, comprising satellite imagery-based forest health monitoring. As a consequence, the emerging data-centric artificial intelligence paradigm [44] has recently started receiving attention in remote sensing where the big satellite image collections (e.g., the Earth Sentinel-1 and Sentinel-2 image collections acquired via the Copernicus programme) are freely available. [38] describes the main principles of the data-centric artificial intelligence paradigm in geospatial data applications by highlighting that data acquisition and curation should receive as much attention as data engineering and model development and evaluation. This study describes one of the first data-centric remote sensing pipelines experimented for land cover classification in satellite imagery. [35] describes a data-centric approach that uses deep feature extraction to prepare a Sentinel-2 dataset to improve the performance of insect species distribution models. [12] describes a data-centric approach that combines semantic segmentation and Geographical Information Systems (GIS) to obtain instance-level predictions of wind plants by using free orbital satellite images. Specifically, this study achieves an improvement of the model performance by including the wind plant shadows to increase the mapped area and facilitate target detection. [11] investigates the application of iterative sparse annotations for semantic segmentation in remote-sensing imagery, focusing on minimizing the labor-intensive and costly

data labeling process. Finally, [40] describes a data-centric approach for RGB imagery dataset creation that reduces annotation ambiguity for RGB images by combining semi-supervised classification and clustering. To the best of our knowledge, no previous studies have explicitly defined a data-centric semantic segmentation approach that pays specific attention to the data curation step, in addition to the model development step, to support bark beetle infestation mapping considering multisensor remote sensing data provided by Sentinel-1 and Sentinel-2 satellites.

3 Study area and data preparation

This section describes the pipeline realised to prepare the datasets used to train and test the semantic segmentation models. We used Microsoft Planetary Computer[1] that provides the API to access petabytes of environmental monitoring data comprising Sentinel-1 and Sentinel-2 images from 2016 to the present. Datasets are accessed via Azure Blob Storage. The study site denoted as *Northeast France*, situated in the northeastern region of France, is predominantly covered by coniferous forests. In 2018 and 2019, a significant proliferation of bark beetles occurred, leading to an estimation by the French National Forestry Office in late April 2019 that approximately 50% of spruce trees in France were infested, contrasting with the typical rate of 15% for dead or diseased trees under normal circumstances. Notably, preceding 2018, there were no instances of substantial windthrows in this area, suggesting that the observed regional-scale attacks were likely spurred by the hot summer droughts experienced in 2018. Satellite data covering the *Northeast France* study site consists of a Synthetic Aperture Radar (SAR) image acquired via the Sentinel-1 sensor and an optical multi-spectral image acquired via the Sentinel-2 sensor.

3.1 Sentinel-1 and Sentinel-2 data collection

The Sentinel-1 satellite constellation collects polarization data via a C-band synthetic-aperture radar instrument. The C-band denotes a nominal frequency range from 8 to 4 GHz (3.75 to 7.5 cm wavelength) within the microwave (radar) portion of the electromagnetic spectrum. Imaging radars equipped with C-band are generally not hindered by atmospheric effects. They are capable of imaging in all-weather (even through tropical clouds and rain showers), day or night. The constellation is composed of two satellites (Sentinel-1A and Sentinel-1B), and it offers a 6-day exact repeat cycle. This means that, over the same geographical area, one SAR can be accessed every 6 days. Due to the nature of the radar signal, the raw information needs calibration correction related to the terrain topography. For this reason, we adopt the level-1 Radiometrically Terrain Corrected (RTC) product available via the Microsoft

[1] https://planetarycomputer.microsoft.com/

Table 1: Sentinel-1 band description

Band	Spatial Resolution	Wavelength	Band Name
VV	10 m	C Band / 3.75 – 7.5 cm	Vertical-Vertical
VH	10 m	C Band / 3.75 – 7.5 cm	Vertical-Horizontal

Planetary platform[2]. This product provides SAR images at 10m of spatial resolution. Here we consider the two polarizations VV (Vertical-Vertical) and VH (Vertical-Horizontal). In particular, VV is a mode of radar polarisation where the microwaves of the electric field are oriented in the vertical plane for both signal transmission and reception by means of a radar antenna. VH is a mode of radar polarisation where the microwaves of the electric field are oriented in the vertical plane for signal transmission and where the horizontally polarised electric field of the back-scattered energy is received by the radar antenna. The list of Sentinel-1 bands considered in this study is reported in Table 1.

The Sentinel-2 satellite constellation retrieves multi-spectral radiometric data (13 bands) in the visible, near infrared, and short wave infrared parts of the spectrum through two satellites (Sentinel-2A and Sentinel-2B). The Sentinel-2 constellation permits covering the majority of the Earth's surface with a repeat cycle of 5 days. The optical imagery is acquired at high spatial resolution (between 10m and 60 m) over land and coastal water areas. The mission supports a broad range of services and applications such as agricultural monitoring, emergency management or land cover classification. Similarly to the SAR signal, also the optical signal collected by the Sentinel-2 sensors requires corrections. To this end, we adopt the level 2A product available via the Microsoft Planetary platform [3] that provides atmospherically corrected surface reflectances. Here we consider all the multi-spectral bands at a spatial resolution of 10m. While bands B2, B3, B4 and B8 are originally at a spatial resolution of 10m, for all the other bands we downscale them at 10m of spatial resolution via the nearest-neighbor resampling based interpolation [34]. This technique selects the value of the pixel that is nearby the surrounding coordinates of the intended interpolation point. Finally, we ignore the B10 (SWIR – Cirrus) band that is reserved for atmospheric corrections. The final list of Sentinel-2 bands considered in this study is reported in Table 2. In particular, for each Sentinel-2 band, we report the spatial resolution, the central wavelength, and the band name. The central wavelength refers to the midpoint wavelength at the centre of the spectral band range (barycenter) that the satellite sensor captures. For example, for the B1 band that captures wavelengths from 433 to 453 nanometers (nm), the central wavelength is 443 nm.

[2] https://planetarycomputer.microsoft.com/dataset/sentinel-1-rtc
[3] https://planetarycomputer.microsoft.com/dataset/sentinel-2-l2a

Table 2: Sentinel-2 band description.

Band	Spatial Resolution	Central Wavelength	Band Name
B1	60 m	443 nm	Coastal aerosol
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Red edge 1
B6	20 m	740 nm	Red edge 2
B7	20 m	783 nm	Red edge 3
B8	10 m	842 nm	Near Infrared Narrow
B8A	20 m	865 nm	Narrow Near Infrared Narrow
B9	60 m	940 nm	Water vapor
B11	20 m	1610 nm	SWIR 1
B12	20 m	2190 nm	SWIR

3.2 Multisensor data alignment

Let us consider a collection of scenes in *Northeast France* for which we know the coordinates of each scene geometry and the timestamp in which scenes were observed using both Sentinel-1 and Sentinel-2 sensors. For each scene, we perform two geospatial queries to select a Sentinel-1 and a Sentinel-2 image acquired in a given time interval. The two queries are performed over the Sentinel-1 and Sentinel-2 collections, respectively, using the coordinates of the selected scenes and the selected time interval as query filters. The queried Sentinel-1 and Sentinel-2 images are recorded in the World Geodetic System 1984 ensemble using metric units. As each query may return a resultset of images, we adopt a pipeline to select a representative image from each resultset.

In particular, images are downloaded from Planetary using the STAC API.[4] For each scene in the study area, we first retrieve the Sentinel-2 image of the scene in a given month by formulating a STAC API query with the parameters "catalogue", "bbox" and "datetime" set as follows: the value "sentinel-2-l2a" is used as "catalogue", the "list of the coordinates of the four vertices of the rectangular box of the scene" is used as value for "bbox", and the "date interval from the first day to the last day of a given month" is used as value for "datetime". As the Sentinel-2 satellite may record images of the Earth every five days, the resultset of such query may contain several Sentinel-2 images recorded in the sentinel-2-l2a catalogue, covered by the given bbox, and acquired by the satellite within the selected datetime interval. The motivation for querying the sentinel-2-l2a catalogue with a time interval (one month in this study) is that cloud cover, shadows and defective pixels are among the main issues that may affect the Sentinel-2 imagery. The assumption for the success of a model development step performed with Sentinel-2 images is that images have to be as much as possible cloud and defective pixels-free. For this reason, we query Sentinel-2 imagery on a time interval (of one month in this study), to improve the possibility of choosing low-affected Sentinel-2 images in terms of clouds and defective pixels. Hence, we select the Sentinel-2 image

[4] `https://planetarycomputer.microsoft.com/docs/quickstarts/reading-stac/`

of the resultset that achieves the lowest value of "cloud index". If several images achieve the minimum value of the cloud index in the resultset, then we select the most recent Sentinel-2 image of this selection. The cloud index is computed based on the output of the Scene Classification Level (SCL) algorithm [30]. This information is also recorded as a band in the sentinel-2-l2a catalogue. Specifically, the SCL algorithm uses the reflectance properties of imagery bands to establish the presence or absence of clouds or defective pixels in an image. In this way, it identifies clouds, snow and cloud shadows thus, generating a classification map, which consists of three different cloud classes (including cirrus), together with six additional classes covering shadows, cloud shadows, vegetation, not vegetated, water and snow land covers. For a candidate Sentinel-2 image, the index of cloud is computed as the percentage of imagery pixels that the SCL algorithm recognises as noise, defective, dark, cloud, cloud shadow or thin cirrus.

Given the Sentinel-2 image retrieved for a given scene in the given month, then we formulate the STAC API query to retrieve the Sentinel-1 image that is co-located in space and time with this Sentinel-2 image. The new query is performed by setting the "bbox" parameter as in the query performed to obtain the Sentinel-2 image while setting "catalogue" equal to "sentinel-1-rtc" and "datetime" equal to the "interval from three days before the date of the Sentinel-2 image and three days after the date of the Sentinel-2 image". The time interval of this query depends on the fact that we would extract a Sentinel-1 image that should be roughly co-located in time with the Sentinel-2 image. On the other hand, Sentinel-1 images are collected every three days with any weather by using a technology not affected by clouds or weather. In addition, we note that noise has been already removed from the Sentinel-1 images that are recorded in the "sentinel-1-rtc" catalogue of Planetary thanks to the application of the Radiometrically Terrain Corrected (RTC) process. This process has been performed before recording the images in the "sentinel-1-rtc" catalogue by using the Ground Range Detected (GRD) Level-1 products produced by the European Space Agency with the RTC processing performed by Catalyst [5]. Hence, we limit to search the Sentinel-1 images potentially collected before and after the Sentinel-2 image and select the Sentinel-1 image that is the closest in time to the respective Sentinel-2 image.

3.3 Ground truth data, datasets and statistics

We use the ground truth map of the bark beetle infestation hotspots that caused tree dieback in the Northeast of France in October 2018.[6] This map was commissioned by the French Ministry of Agriculture and Food to Sertit

[5] `https://catalyst.earth/`

[6] `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`

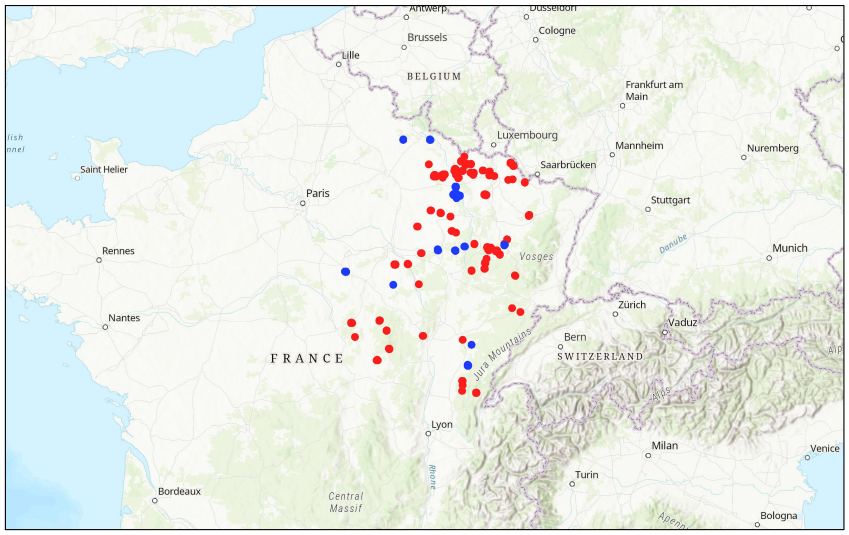


Fig. 1: Location of the centroids of the study 87 scenes in the Northeast of France area. The red circles correspond to scenes considered for training semantic segmentation models, while the blue circles correspond to scenes considered for evaluating semantic segmentation models.

(University of Strasbourg), to assess the damage in spruce forests of the Northeast of France following the 2018 bark beetle outbreak. The remote sensing company WildSense assessed and fixed the infestation hotspot polygons of this map. In particular, to avoid mixed reflectance from various causes in discoloration and defoliation of conifer, WildSense manually selected 87 squared, imagery tiles, covering spruce forestry areas fully under bark beetle attacks in October 2018. The scenes of the final collection cover 1004020 pixels at 10 square meters resolution. The size of the scenes varies from 27×16 to 296×319 pixels at 10 square meters resolution, while the percentage of infested territory per scene varies from 0.35% to 34.4% of the scene surface. The total percentage of damaged territory of the entire scene collection is 2.92%. For the experimental evaluation of this research work, we consider 71 scenes (covering 772844 pixels at 10 squared meters resolution) as training scene set and 16 scenes (covering 231176 pixels) as testing scene set. A map of the study scene location and their partitioning in the training set and testing set is depicted in Figure 1. In addition, WildSense identified an extra scene covering spruce forestry areas fully under bark beetle attacks, according to a ground truth map acquired in March 2020. The geographic location of this scene is shown in Figure 2. This scene is a tile with size 205×135 covering 27675 pixels with 10 squared meters as spatial resolution with a percentage of infested territory equal to 3.55%.

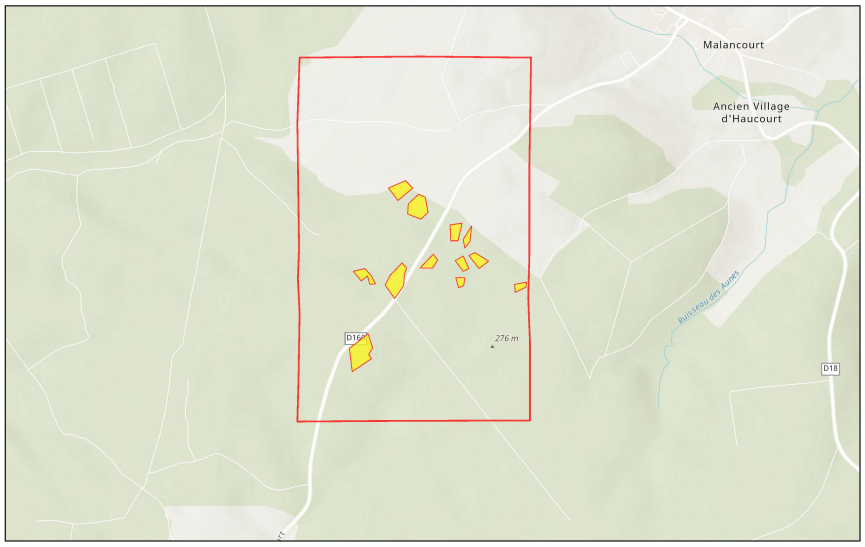


Fig. 2: Location of the scene for which the ground truth mask of the bark beetle infestation was acquired in March 2020. The yellow patches map the forest areas with tree dieback caused by the bark beetle.

In this study, we prepare four multisensor, satellite datasets populated with both the Sentinel-1 and Sentinel-2 images acquired for each scene in the study area in the Northeast of France. Hence, each dataset is populated with 87 Sentinel-1 images and 87 Sentinel-2 images roughly co-located in time within the same month. Specifically, the four multisensor satellite datasets were obtained by considering Sentinel-1 and Sentinel-2 images acquired monthly for the 87 study scenes in July 2018, August 2018, September 2018 and October 2018, respectively. We partition each imagery dataset into a training set and a testing set by using the same split ratio for each month. In particular, as mentioned above, we select 71 multisensor images as the training set and 16 multisensor images as the testing set for each of these four datasets. Notably, the multisensor images assigned to the four training sets were acquired for the same 71 training scenes although in different months. Similarly, the multisensor images assigned to the four testing sets were acquired for the same 16 testing scenes although in different months.

The dataset collected in October 2018 – the time at which the ground truth map of the bark beetle-induced tree dieback of the study scenes was produced – is elaborated to analyse the ability to map bark beetle-induced tree dieback in October, while datasets collected for the same scenes from July to September 2018 are elaborated to analyse the ability to predict as earlier as possible signs related to the bark beetle infestation (before trees start dying). Notice that the analysis of satellite imagery data collected in October 2018 follows some communications with foresters reported by [8], according to 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. On the other hand, the analysis of satellite imagery data collected in July, August and September 2018 is done to explore the performance of the proposed approach in predicting where bark beetle infestation disturbance events are likely to cause future tree dieback. This evaluation is done according to the considerations reported in [27] that the early detection symptoms of bark beetle infestation, which comprise the presence of entrance holes, resin flow from entrance holes and boring dust that occur when the beetles attack the tree, penetrate the bark, and excavate mating chambers and breeding galleries that can be observed through terrestrial fieldwork inventory. So, counting on manually produced labels in the summer months may help the training of semantic segmentation models for automated early detection in scenes uncovered by the forestry fieldwork.

Figure 3 shows the box plots of Sentinel-1 and Sentinel-2 data collected in the datasets prepared for this study. All bands are plotted independently of each other for the two opposite ground truth classes ("damaged" and "healthy"). The box plots show that the range of both Sentinel-1 and Sentinel-2 data changes over time. Sentinel-2 data, particularly B5, B6, B7, B8, B8A and B9, show a greater divergence between the opposite classes than Sentinel-1 data, over all the datasets. So, this visual data exploration confirms the general idea that Sentinel-2 contains the most important information to recognize bark beetle infestation hotspots, while Sentinel-1 data can be considered ancillary data that may be used to support analysis of Sentinel-2 data, to gain accuracy in the bark beetle infestation inventory.

In addition, Figure 4 shows the results of the bivariate correlation analysis performed by computing the Spearman's rank correlation coefficient between Sentinel-1 and Sentinel 2 bands in images acquired between July and October 2018. Spearman's rank correlation coefficient is a non-parametric measure of rank correlation that assesses how well the relationship between two compared variables can be described using a monotonic function. It varies between -1 and +1 with 0 implying no correlation, -1 implying an exact monotonic relationship with negative correlation and +1 implying an exact monotonic relationship with positive correlation. This correlation analysis shows that the Sentinel-1 bands VV and VH are negatively correlated to the Sentinel-2 bands B1, B2, B3, B4, B5, B11 and B12, while they are positively correlated to Sentinel-2 bands B7, B8, B8A and B9. The Sentinel-2 band B6 passes from showing a low negative correlation with the Sentinel-1 bands VV and VH in July to showing a low positive correlation with the same Sentinel-1 bands in August, September and October. In general, the intensity of the correlation between the Sentinel-2 bands B6, B7, B8, B8A and B9 and the Sentinel-1 bands VV and VH increases from July to August, September and October. In any case, the correlation is close to zero independently of the sign, especially on the bands B6, B7, B8, B8A and B9, which are the Sentinel-2 bands that better separate the opposite classes in the box plot analysis of the same data. Hence, this visual inspection of the collected data confirms a limited correlation between Sentinel-1 and

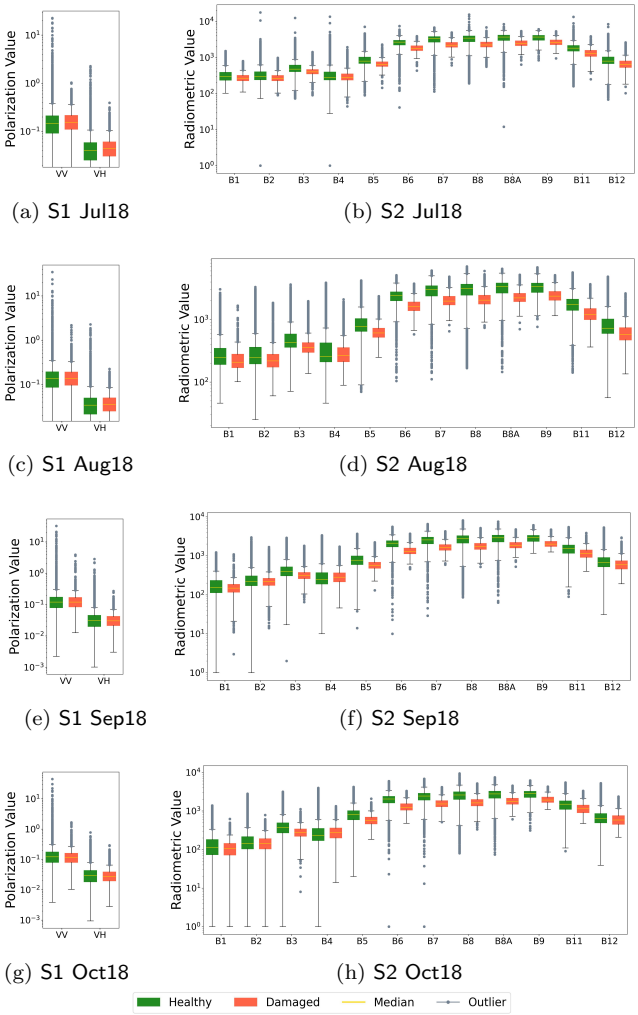
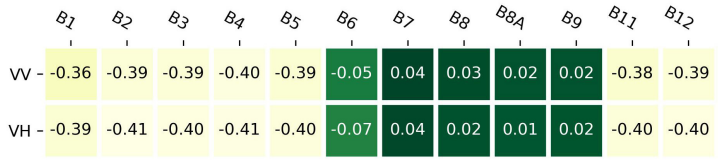


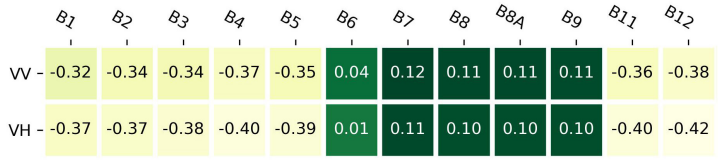
Fig. 3: Box plot distribution of the polarization values measured for the Sentinel-1 bands and the radiometric values measured for the Sentinel-2 bands recorded in the datasets of Sentinel-1 and Sentinel-2 images acquired in the study site in July, August, September and October 2018. Bands are plotted independently with respect to the two opposite classes in the logarithmic scale.

Sentinel-2 data, which is one of the prerequisites for taking advantage of a multisensor approach in model development.

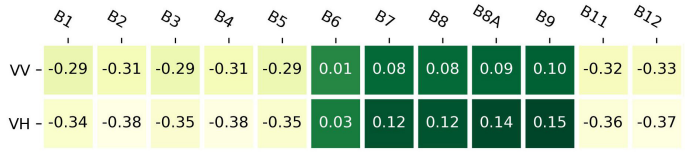
Figure 5 shows the box plot of the cloud index of the Sentinel-2 images selected for this study. This plot shows the high quality of Sentinel-2 images selected in each month. In fact, we are unable to select images with a cloud index lower than 5% only in one image in August 2018 and two images in



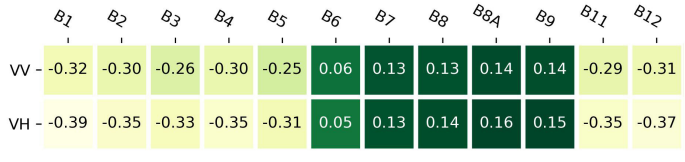
(a) Jul18



(b) Aug18



(c) Sep18



(d) Oct18

Fig. 4: Spearman's rank correlation coefficient computed between Sentinel-1 and Sentinel-2 bands in the images acquired in the study site in July, August, September and October 2018

October 2018. We also note that differences between the box-plot quartiles are slightly higher in October 2018 than in the period July-September 2018. This depends on the expected increase in the frequency of cloudiness as autumn advances.

Finally, we collect and prepare the pair of Sentinel-1 and Sentinel-2 images of the scene for which the ground truth map was acquired in March 2020. This pair of images is used in the evaluation stage only, to explore the transferability of the semantic segmentation model learned in October 2018 to subsequent periods. The Sentinel-2 image acquired for this scene in March 2020 and selected in this study has a low noise and cloud index equal to 0.16%. Finally,

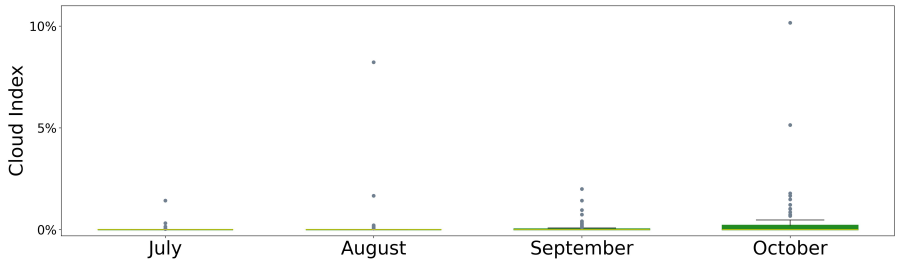
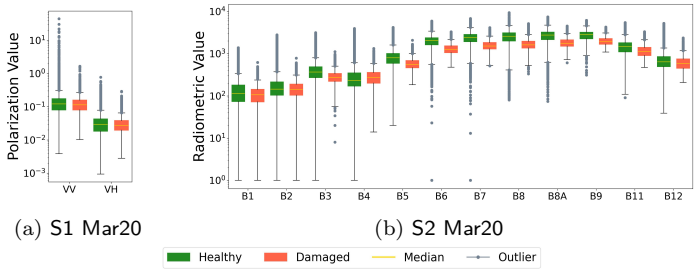


Fig. 5: Box plot of cloud index of Sentinel-2 images acquired in the study site in July, August, September and October 2018.



(a) S1 Mar20 (b) S2 Mar20

Fig. 6: Box plot distribution of the polarization values measured for the Sentinel-1 bands and the radiometric values measured for the Sentinel-2 bands recorded in the Sentinel-1 image and the Sentinel-2 image acquired in March 2020 for the scene seen in Figure 2. Bands are plotted independently to the two opposite classes in the logarithmic scale.

Figure 6 shows the box plots of both Sentinel-1 and Sentinel-2 data collected in March 2020 for this scene. We note that the outliers of Sentinel-1 data are spread across a lower heat range than that observed in the images collected in the summer and autumn months of 2018. On the other hand, B5, B6, B7, B8, B8A and B9 of Sentinel-2 data still show a remarkable divergence between the opposite classes as in the images collected in the summer and autumn months in 2018.

4 Semantic segmentation model development

The model development step is performed by leveraging the aligned Sentinel-1 and Sentinel-2 images of scenes for which the ground truth mask of bark beetle infestation is available. Let us consider $\mathcal{D} = \{(\mathbf{X_{S1}}, \mathbf{X_{S2}}, \mathbf{Y}) | \mathbf{X_{S1}} \in \mathbb{R}^{H \times W \times 2}, \mathbf{X_{S2}} \in \mathbb{R}^{H \times W \times 12}, \mathbf{Y} \in \mathbb{R}^{H \times W \times 1}\}$ a collection of labelled Sentinel-1 and Sentinel-2 images of forest scenes, where every ground truth mask $\mathbf{Y}$ is associated with the images $\mathbf{X_{S1}}$ and $\mathbf{X_{S2}}$, acquired from Sentinel-1 and

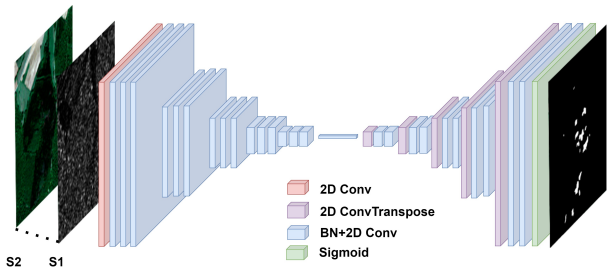


Fig. 7: Early fusion. Abbreviations: 2D Conv = 2D Convolutional layer; BN=Batch Normalization; S1=Sentinel-1; S2=Sentinel-2

Sentinel-2 satellites, respectively. For each scene, H and W denote the spatial extent of the monitored scene in terms of scene height and scene width, respectively. The model development step trains a semantic segmentation network from $\mathcal{D}$ through a U-Net-like architecture that is also in charge of learning the data fusion.

The U-Net architecture is composed of an encoder part and a decoder part. The encoder extracts features. It consists of multiple blocks, where each block is composed of a Batch Normalization layer and a 2D Convolutional layer followed by Max-Pooling for downsampling. At each downsampling step, the height and width of the tensor halves, while the number of channels remains unchanged. The decoder part upsamples the encoded feature maps to the original input shape. It consists of one transposed Convolutional layer for upsampling, followed by multiple blocks, each of which each block consists of a Batch Normalization layer and a 2D Convolutional layer. Skipping connections between the decoder part and the encoder part are used to propagate the spatial information from the earlier layers to the deeper layers to alleviate the vanishing gradients problem [42]. The final classification of each imagery pixel is obtained by using the Sigmoid activation function. The U-Net used in this study is trained via the Tversky loss, which is commonly used to handle imbalanced data [20].

The data fusion mechanism is implemented through three different strategies, namely, Early fusion, Middle fusion and Late fusion, which are defined according to the general classification of multimodal data fusion methods reported in the survey of [46]. The Early fusion strategy is the first mechanism adopted in literature for the multimodal data fusion in the deep neural scenario [13]. It is implemented via a simple concatenation, performed at an early stage, of features from different modalities (i.e., sensors in this study). The concatenation produces a single input space for the model development. In our study, the Early fusion strategy, shown in Figure 7, concatenates each pair of images $\mathbf{X_{S1}}$ and $\mathbf{X_{S2}}$ obtaining a single hypercube with dimension $H \times W \times 14$. A traditional U-Net architecture is trained on the newly stacked hypercubes.

The Middle fusion strategy combines features learned with the separate branches of a multi-input deep neural network that takes data acquired from

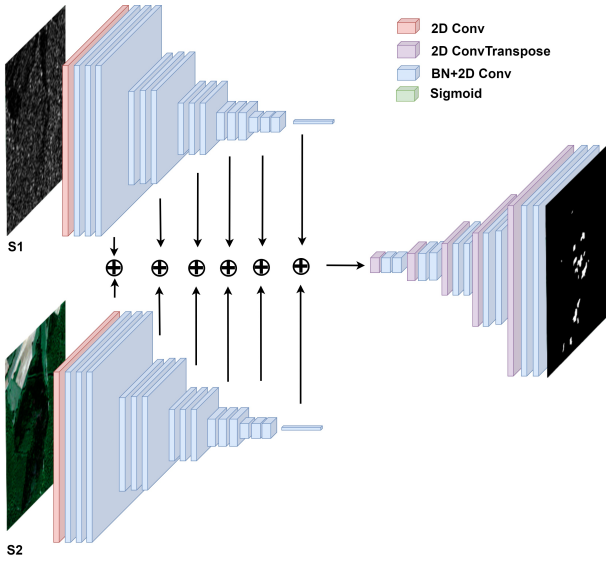


Fig. 8: Middle fusion. Abbreviations: 2D Conv = 2D Convolutional layer; BN=Batch Normalization; S1=Sentinel-1; S2=Sentinel-2

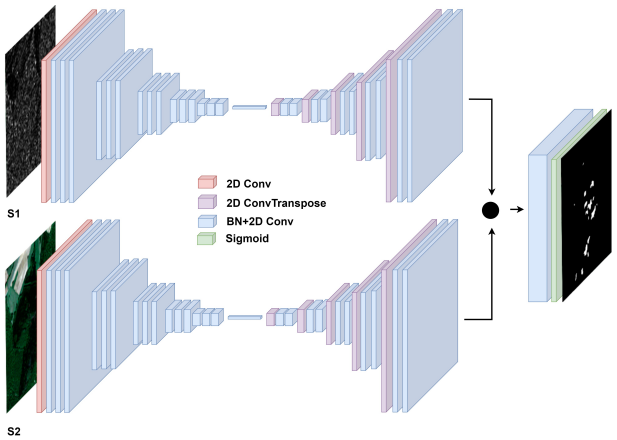


Fig. 9: Late fusion. Abbreviations: 2D Conv = 2D Convolutional layer; BN=Batch Normalization; S1=Sentinel-1; S2=Sentinel-2

different modalities as separate inputs. The fusion is performed at an intermediate layer of the deep neural network. The output of this combination performed at the fusion layer is processed across the subsequent layers of the network until the decision layer. In our study, the Middle fusion strategy, depicted in Figure 8 uses an architecture with two encoder branches, each taking $\mathbf{X_{S1}}$ and $\mathbf{X_{S2}}$ as input, respectively. The output of these branches is fed into a

single decoder. The two encoder branches are mapped into a common feature space via a fusion operation and the fusion output is used for the skipping connections. Two fusion operators, named SUM and CONC, are considered in this work for the middle fusion. The SUM operator performs an element-wise summation between the outputs of two parallel blocks in the encoder parts. The CONC operator produces a single hypercube by stacking the outputs of two parallel blocks in the encoder parts. Subsequently, it employs a 2D Convolutional layer to halve the channel size of the output hypercube. This is done to align with the number of channels of the corresponding decoder block for skipping connections. Both the concatenation [13,47] and the element-wise summation [33,36] are two common fusion operators used in the literature to fuse multimodal features enclosed in RGB images and Depth images by using CNN-based algorithms. We select these two operators for the Middle fusion strategy performed in this study since they implement two different mechanisms in terms of information retention. In particular, the concatenation operator (CONC) allows us to keep all the information from both Sentinel-1 and Sentinel-2 data, where each feature is entirely preserved. On the other hand, the summation operator (SUM) provides a more compact representation than the concatenation. In fact, it fuses the features originated from the two sensors into a single vector having the same size of the combined vectors. This operator can be particularly useful when the features are aligned and represent the same spatial locations or attributes.

The Late fusion strategy processes separately input data provided by each modality through distinct deep neural models, and their outputs are combined at the later stage, usually at the classification stage. In our study, the Late fusion strategy, illustrated in Figure 9, uses an architecture with two identical, parallel encoder and decoder paths that take as input $\mathbf{X_{S1}}$ and $\mathbf{X_{S2}}$, respectively. The outputs returned by the two decoders are stacked into a single hypercube and the Sigmoid activation function is employed in the final layer. Final considerations concern the expected behaviour of the three data fusion schemes. According to the discussion reported in [46], the Early fusion strategy is expected to better leverage cross-modal information interaction as early as possible in the learning stage. On the other hand, the Late fusion strategy is considered flexible, but it may lack sufficient cross-modal correlation. Finally, the Middle fusion strategy is expected to find a trade-off between Early fusion and Late fusion, with possible advantages in terms of final performances.

5 Empirical evaluation and discussion

5.1 Implementation details

We implemented DIAMANTE in Python 3.0. The source code is available online.[7] In this study, we consider a U-Net architecture optimized for satellite

[7] https://github.com/gsndr/DIAMANTE

images implemented using the Keras 2.15 and TensorFlow as back-end [8]. Both encoder and decoder components of the different variants of U-Net architectures tested in this study are composed of five main blocks. In the encoder part, each block consists of 3 blocks containing a Batch Normalization layer and a 2D Convolutional layer, followed by a 2×2 Max-Pooling operation or downsampling. The stride of the Max-Pooling operation was set equal to 2. In the decoder part, each main block consists of a transposed Convolutional layer (for upsampling) followed by 3 blocks containing a Batch Normalization layer and a 2D Convolutional layer. The kernel size of each Convolutional layer was set equal to 3×3. In all hidden layers the Rectified Linear Unit function (ReLU) was used as the activation function, while the Sigmoid activation function was used in the final semantic segmentation layer. The SUM operator was implemented using the Add layer available in TensorFlow.[9] The training of the U-Net architectures was performed using imagery tiles of size 32×32 extracted from the imagery scenes by using tiler library.[10] Both Sentinel-1 and Sentinel-2 data were scaled between 0 and 1 using the Min-Max scaler (as it is implemented in the Scikit-learn 0.22.2 library) In addition, we considered a tile augmentation strategy to improve the performance of the U-Net architecture by using the Albumentations library [11]. Specifically, we quintupled the number of training imagery tiles by creating new tiles applying traditional computer vision augmentation operators (i.e., Horizontal Flip, Vertical Flip, Random Rotate, Transpose and Grid Distortion). We used the tree-structured Parzen estimator algorithm to optimize hyper-parameters of U-Net architectures (i.e., mini-batch size in $\{2^2, 2^3, 2^4, 2^5, 2^6\}$, learning rate between 0.0001 and 0.01 and image augmentation in {True, False}), by using 20% of the training set as the validation set. In particular, the hyper-parameter configuration that achieves the highest F1 score on the minority class ("damaged") in the validation set was automatically selected as the best semantic segmentation model. We performed the gradient-based optimisation using the Adam update rule. Finally, each U-Net model was trained with a maximum number of epochs equal to 150, using an early stopping approach to retain the best semantic segmentation model.

5.2 Metrics

To evaluate the accuracy of the semantic segmentation masks, we measured the following metrics: F1 score (F1) computed for the two opposite classes, Macro F1 score (Macro F1) averaged on the two opposite classes and Intersection-over-Union (IoU). Specifically, the F1 score measures 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

[8] https://github.com/karolzak/keras-unet/tree/master

[9] https://www.tensorflow.org/api_docs/python/tf/keras/layers/Add

[10] https://github.com/the-lay/tiler

[11] https://albumentations.ai/

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" (F1(h)) and "damaged" (F1(d)). Macro F1 measures the average of each F1 score value per class, that is, Macro F1=$\frac{F1(h)+F1(d)}{2}$. The IoU 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 to evaluate the accuracy of models trained in both semantic segmentation and object detection problems. All metrics are reported in percentages and computed on the images collected for the testing scenes. For each metric, the higher the value, the better the performance of the semantic segmentation masks predicted.

5.3 Results

The illustration of results is organised as follows. Section 5.3.1 presents the results achieved by processing the multisensor imagery dataset collected in the study area in October 2018. This analysis is done to evaluate the performance of the data fusion strategies at the same time the ground truth masks of the study scenes were created. Section 5.3.2 presents a temporal study where we explore the performance of the models trained and evaluated considering satellite images acquired in July, August and September 2018. This analysis is done to explore the ability of the considered data fusion strategies to learn a model capable to perform early detection of tree dieback phenomena. Finally, Section 5.3.3 illustrates the results achieved by considering multisensor semantic segmentation models trained from satellite images acquired in October 2018 to predict the mask of tree dieback caused by a bark beetle infestation in a new scene located in the Northeast of France, but monitored in March 2020. This analysis explores the transferability over time of a semantic segmentation model.

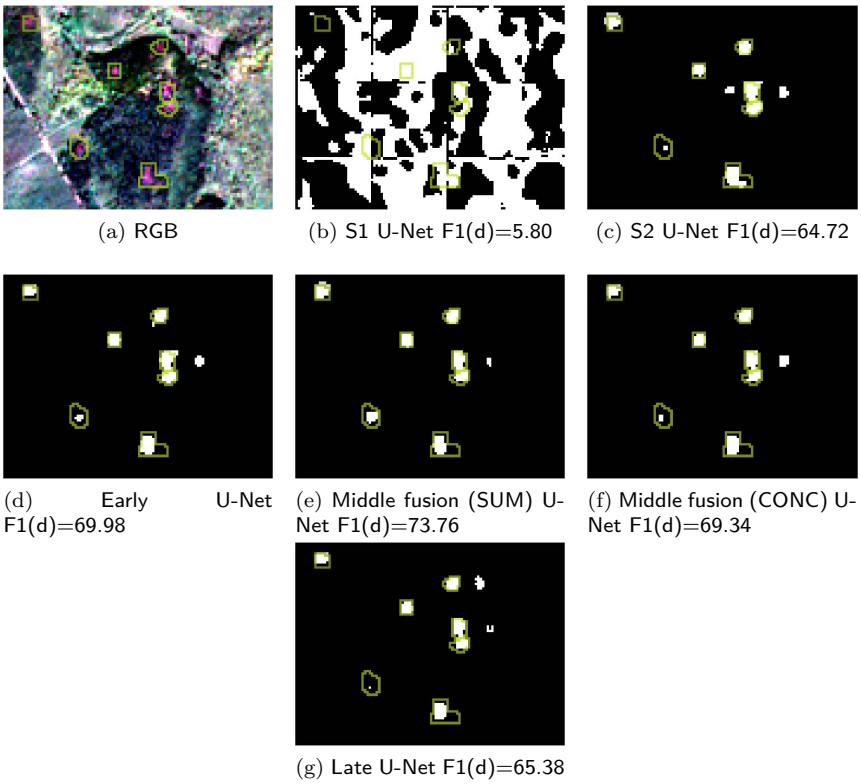
5.3.1 Performance Analysis

In this Section, we analyse the performance of the semantic segmentation masks produced for the testing scenes of the *Northeast France* study by using the multisensor semantic segmentation models trained via the three data fusion schemes illustrated in Section 4. As baselines, we consider the single-sensor semantic segmentation models trained with a traditional U-Net by processing either the Sentinel-1 images (S1 U-Net) or the Sentinel-2 images (S2 U-Net) alone. With regard to the Middle fusion strategy, we report the results achieved with the two fusion operators: SUM and CONC. This evaluation was conducted by processing the dataset of images acquired in October 2018 for both the training and evaluation stages. The accuracy metrics measured on the semantic segmentation masks produced for the images of the testing scenes are reported in Table 3. As we expected, the output of the stand-alone use

Table 3: Accuracy performance of semantic segmentation produced with S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net in the imagery collections acquired in October 2018. The best results are in bold

Configuration	F1(h)	F1(d)	IoU	Macro F1
S1 U-Net	61.90	5.47	2.81	33.68
S2 U-Net	99.14	64.65	47.76	81.89
Early fusion U-Net	**99.42**	69.09	52.78	84.25
Middle fusion (SUM) U-Net	99.35	**70.93**	**54.92**	**85.13**
Middle fusion (CONC) U-Net	99.40	69.90	53.73	84.65
Late fusion U-Net	99.29	67.66	51.25	83.47

of Sentinel-1 images is unsatisfactory for this inventory task. In fact, the configuration S1 U-Net achieves the lowest performance in all accuracy metrics. Better performance can be achieved by processing Sentinel-2 images in place of Sentinel-1 images. However, this evaluation study shows that the data fusion of Sentinel-1 and Sentinel-2 images can help us to improve the performance of the semantic segmentation model regardless of the type of data fusion strategy employed. In fact, the Early fusion U-Net, Late fusion U-Net and Middle fusion U-Net all achieve better performance than S2 U-Net that considers Sentinel-2 images only. More in detail, the best configuration in terms of F1(d), IoU and Macro F1 is achieved with the Middle fusion schemes having Middle fusion (CONC) U-Net as runner-up of Middle fusion (SUM) U-Net. These conclusions are consistent with the observations on the expected behaviour of the data fusion schemes reported in Section 4. Figures 10b-10g show the semantic segmentation masks of a sample testing scene predicted by the compared models, while Figure 10a shows the RGB image of this sample scene. The masks highlight how the use of a data fusion strategy helps us to reduce the number of false alarms in this case. Specifically, the bark beetle infestation masks predicted using the multisensor U-net trained with both Early fusion and Middle fusion schemes show only one false infested patch, while the U-Net trained from Sentinel-1 data shows large extensions of false infested areas and the U-Net trained from Sentinel-2 data shows two false infested patches. Notably, the multisensor U-Net trained with Late fusion strategy removes one of the false patches discovered by S2 U-Net, but, at the same time, it alerts a new false patch that is undetected in the other masks. We note that the Late fusion strategy is the worst-performing fusion strategy of this experiment. This result suggests that although the Late fusion strategy may allow us to correct some false patches detected processing Sentinel-2 data only, it may also produce some artefacts at the decision level, which may cause false alarms unseen in the remaining configurations. Finally, the masks of this example show that the use of SUM operator performs better than the CONC operator in delineating the large damaged patch located on the left side of the scene.



(a) RGB (b) S1 U-Net F1(d)=5.80 (c) S2 U-Net F1(d)=64.72

(d) Early U-Net F1(d)=69.98 (e) Middle fusion (SUM) U-Net F1(d)=73.76 (f) Middle fusion (CONC) U-Net F1(d)=69.34

(g) Late U-Net F1(d)=65.38

Fig. 10: RGB of the Sentinel-2 image acquired in October 2018 for a testing scene of the study area in the Northeast of France (10a). Inventory masks of tree dieback areas caused by bark beetle hotspots in this scene as they are predicted by S1 U-Net (10b), S2 U-Net (10c), Early fusion U-Net (10d), (10g), Middle fusion U-Net with operators SUM (10e) and CONC (10f) and Late fusion U-Net trained on the imagery set acquired in October 2018 for the training scenes of the study area.

5.3.2 Temporal Analysis

To complete this investigation, we illustrate the results of a temporal study conducted to explore the accuracy performance of the semantic segmentation maps produced when the Sentinel-1 and Sentinel-2 images were acquired in the middle of summer (i.e., July 2018) and the late summer (i.e., August 2018 and September 2018), while the ground truth map of the tree dieback was observed in early autumn (October 2018). This analysis is done to explore the performance of the presented data fusion strategies in the early detection of areas where bark beetle infestation disturbance events are likely to cause (near-)future tree dieback. The temporal snapshots of this experiment were selected according to the recent achievements of the analysis on the spectral

Table 4: Accuracy performance of the semantic segmentation models produced with S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net in the multisensor images acquired with both Sentinel-1 and Sentinel-2 satellites from July 2018 to October 2018. For each configuration, the best results are in bold, while the runner-up configuration is underlined.

Imagery set	Configuration	F1(h)	F1(d)	IoU	Macro F1
Jul18	S1 U-Net	73.53	4.03	2.06	38.68
	S2 U-Net	97.23	30.18	17.78	63.71
	Early fusion U-Net	98.26	31.49	18.69	64.87
	Middle fusion (SUM) U-Net	**98.39**	**36.46**	**22.29**	**67.43**
	Middle fusion (CONC) U-Net	98.13	35.27	21.41	66.70
	Late fusion U-Net	98.36	34.20	20.63	66.28
Aug18	S1 U-Net	40.87	4.05	2.07	22.46
	S2 U-Net	95.58	26.46	15.25	61.01
	Early fusion U-Net	98.22	32.49	19.40	65.35
	Middle fusion (SUM) U-Net	98.92	**37.41**	**23.00**	**68.16**
	Middle fusion (CONC) U-Net	**99.35**	36.37	22.22	67.29
	Late fusion U-Net	98.02	36.15	22.06	67.08
Sep18	S1 U-Net	72.66	5.26	2.70	38.96
	S2 U-Net	99.32	58.92	41.77	79.13
	Early fusion U-Net	99.14	63.42	46.44	81.28
	Middle fusion (SUM) U-Net	**99.35**	64.28	47.36	81.81
	Middle fusion (CONC) U-Net	99.29	63.16	46.15	81.22
	Late fusion U-Net	99.28	**64.99**	**48.14**	**82.14**
Oct18	S1 U-Net	61.90	5.47	2.81	33.68
	S2 U-Net	99.14	64.65	47.76	81.89
	Early fusion U-Net	**99.42**	69.09	52.78	84.25
	Middle fusion (SUM) U-Net	99.35	**70.93**	**54.92**	**85.13**
	Middle fusion (CONC) U-Net	99.40	69.90	53.73	84.65
	Late fusion U-Net	99.29	67.66	51.25	83.47

separability between the healthy and bark beetle attacked trees illustrated in [14]. In particular, this study shows that bark beetle attacks commonly occur in the summer, while the spectral separability between the two opposite classes ("Healthy" and "Damaged") increases moving from July to October. In addition, it highlights that a time span of approximately one month commonly occurs between the attack of the beetles to a tree and the development of the first symptoms (green-attack) in the tree. Hence, based on the conclusions drawn in this study, the green attack detection stage can reasonably arise in the summer period spanned from July to August. Based on these premises, the accuracy metrics measured on the semantic segmentation maps produced for the testing scenes of this study in each month between July and October 2018 are reported in Table 4.

These results show that the data fusion of Sentinel-1 and Sentinel-2 continues to help us to gain accuracy also when the multisensor semantic segmentation model is trained to forecast tree dieback areas caused by the bark beetle infestation. Notably, Middle fusion (SUM) U-Net achieves the highest F1(d), IoU and Macro F1 in segmentation maps produced in experiments performed in July 2018, August 2018 and October 2018. The only exception is

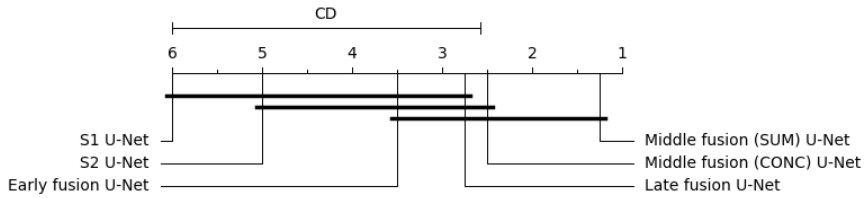


Fig. 11: Comparison of the configurations: Macro F1 measured for S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net, performed with the Friedman-Nemenyi test run on Macro F1 measured in the temporal analysis performed from July 2018 to October 2018 (computed $pvalue = 0.013$)

observed in the segmentation maps produced for the evaluation in September 2018. However, also in the experiment conducted in September 2018, the Middle fusion (SUM) U-Net still achieves good performance by ranking as the runner-up of the Late fusion U-Net. To draw conclusive conclusions on the better data fusion strategy, we perform the Friedman-Nemenyi test to compare the Macro F1 measured for S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net on the multiple segmentation maps produced for the testing data of the multisensor datasets of this temporal analysis. This non-parametric test ranks the model configurations compared for each dataset separately, so the best-performing model is given a rank of 1, the second-best rank of 2 and so on. The results of the Friedman-Nemenyi test reported in Figure 11 shows that the test groups the configurations adopting a multisensor data fusion strategy as statistically different from the configurations that consider either Sentinel-1 data only (S1 U-Net) or Sentinel-2 data only (S2 U-Net). In addition, the Middle fusion (SUM) U-Net achieves the highest rank by having the Middle fusion (CONC) U-Net as runner-up. Notably, these results of the comparative test support the conclusions already drawn in 5.3.1 and 5.3.3 on the superior performance of a Middle fusion strategy to combine Sentinel-1 and Sentinel-2 data for bark beetle infestation detection.

5.3.3 Transferability Analysis

In this Section, we examine the accuracy of the semantic segmentation models learned in October 2018 when used to detect the tree dieback events caused by bark beetle infestations in March 2020. The accuracy metrics measured in this experiment are reported in Table 5. These results show that also in this evaluation scenario, the data fusion of Sentinel-1 and Sentinel-2 may help us to improve the performance of a semantic segmentation model even when it was trained on past images and used for mapping the bark beetle infestation in future images. The only exception is observed for the Late fusion strategy

Table 5: Accuracy performance of semantic segmentation produced with S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net in the pair of Sentinel-1 and Sentinel-2 images acquired in March 2020. The best results are in bold.

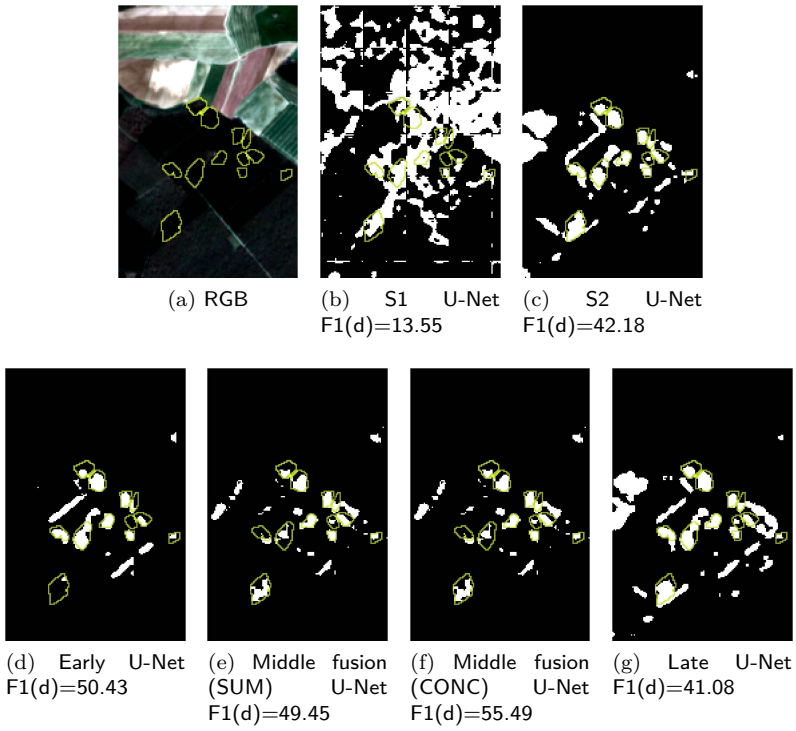
Configuration	F1(h)	F1(d)	IoU	Macro F1
S1 U-Net	82.04	13.55	7.27	37.22
S2 U-Net	96.20	42.18	26.73	69.20
Early fusion U-Net	**98.07**	50.43	33.72	74.25
Middle fusion (SUM) U-Net	98.03	49.45	32.84	73.74
Middle fusion (CONC) U-Net	97.65	**55.49**	**38.40**	**76.57**
Late fusion U-Net	96.00	41.08	25.85	68.54

that achieves lower performance than S2 U-Net. In general, the highest F1(d), IoU and Macro F1 are achieved with the Middle fusion (CONC) U-Net schema having Middle fusion (SUM) U-Net as runner-up. This confirms the conclusions on the better performance of the Middle fusion strategy already drawn in Section 5.3.1. Finally, Figures 12b-12g show the semantic segmentation masks predicted for the scene under evaluation. The RGB image of the scene in March 2020 is shown in Figure 12a. The extracts show that the data fusion schemes, except for Late fusion, allow us to reduce the extension of the false alarm areas detected. In both Early fusion and Middle fusion (SUM) schemes, the higher precision is achieved at the cost of a lower recall. Both data fusion configurations allow us to map correctly a percentage of the infested area that is lower than the one mapped processing Sentinel-2 data only. Instead, the use of the Middle fusion (CONC) strategy allows us to achieve the best trade-off between precision and recall in detecting the tree dieback areas caused by the bark beetle infestation. In general, these maps confirm the idea that also when the semantic segmentation model is trained on historical data, the main contribution to the correct detection of the bark beetle infestation is given by Sentinel-2 data, while Sentinel-1 data can aid in reducing false alarms and better delimiting infested areas.

5.4 Considerations and Findings

The experimental assessment highlights the general advantages of using multisensor data over a single data source in various scenarios of bark beetle detection, including early disease detection and out-of-year temporal transfer. While Sentinel-1 alone is not suitable for the considered downstream mapping task, using Sentinel-2 alone yields satisfactory results. However, the combined use of these two publicly available and freely accessible remote sensing data sources provides the best overall results.

More specifically, the joint use of Sentinel-1 and Sentinel-2 data significantly reduces false alarms and improves the delineation of infested areas in the resulting binary maps. Regarding the early detection of bark beetle attacks (Section 5.3.2), signs of the attack can be detected with reasonable accuracy



(a) RGB (b) S1 U-Net F1(d)=13.55 (c) S2 U-Net F1(d)=42.18

(d) Early U-Net F1(d)=50.43 (e) Middle fusion (SUM) U-Net F1(d)=49.45 (f) Middle fusion (CONC) U-Net F1(d)=55.49 (g) Late U-Net F1(d)=41.08

Fig. 12: RGB of the Sentinel-2 image acquired in March 2020 (12a). Inventory masks of tree dieback areas caused by bark beetle hotspots in this scene as they are predicted by S1 U-Net (12b), S2 U-Net (12c), Early fusion U-Net (12d), Middle fusion U-Net with operators SUM (12e) and CONC (12f) and Late fusion U-Net (12g) trained on the imagery set acquired in October 2018 for the training scenes of the study area

one month before the acquisition of ground truth data (September 2018). However, the disease's early stages (before July 2018) are weakly detectable via satellite imagery.

An additional challenge is represented by the out-of-year transfer of the model trained on 2018 data to 2020 data. Recent studies in the domain of remote sensing analysis have highlighted that spatial and temporal distribution shifts can hinder the direct deployment of a model trained on a particular area or time period to a different area or time period [10,32]. The results obtained in Section 5.3.3 confirm this point, indicating that there is still room for research activities in the way historical data can be leveraged in order to improve current mapping results.

Finally, the comparison of the different approaches indicates that all fusion strategies are statistically significant compared to single source analysis, with the Middle fusion (SUM) U-Net model exhibiting the best average performance.

This finding underscores once more the importance of combining multisensor satellite data for mapping tree dieback induced by bark beetle infestation.

6 Conclusion

In this study, we investigate the effectiveness of a data-centric semantic segmentation approach to map forest tree dieback areas caused by bark beetle hotspots. First, we define a data-centric pipeline to collect and prepare images acquired from both the SAR Sentinel-1 sensor and the optical Sentinel-2 sensor. Then, we explore the accuracy performance of several data fusion strategies, namely Early fusion, Middle fusion and Late fusion adopted for the development of a U-Net-like model combining both Sentinel-1 and Sentinel-2 images acquired in the Northeast of France. Finally, we investigate the performance of the proposed strategies in multisensor imagery data acquired in Northeast of France with the map of bark beetle infestation available in October 2018. We conducted the evaluation with imagery data prepared according to the data curation pipeline presented in this study. The experimental results show that multisensor data can actually help us to improve the ability of the U-Net model to detect tree dieback areas caused bark beetle infestations. The evaluation also explores the transferability of the output of the model development step, as well as the performance of the proposed approach in early detection of infestations that will cause tree dieback.

As future work, we plan to continue the investigation of multisensor data fusion strategies in combination with ecological and weather data, as well temporal data trend information. In addition, we plan to extend the investigation of the transferability of the semantic segmentation model, trained with the described multisensor data fusion techniques to unseen data settings. In particular, we intend to start a systematic exploration of some transfer learning approaches to obtain the transferability of a "general" semantic segmentation model trained for a specific disturbance agent to different disturbance agents. For example, we intend to investigate the transferability of a semantic segmentation model trained for mapping forest tree die-back hotspots caused by bark beetle infestation to perform the inventory of tree die-back hotspots caused by different families of fungal forest pathogens. In addition, we hope to be able to acquire large-scale data within the experimental phase of the EU project SWIFTT to be able of investigating, on large scale, the transferability of a semantic segmentation model trained in a geographic area to a different geographic area, in addition to a future time.

Acknowledgment

Annalisa Appice acknowledges support from the SWIFTT project, funded by the European Union under Grant Agreement 101082732. Dino Ienco acknowledges support from the Eco2Adapt project, funded by the European Union

under Grant Agreement 101059498. Giuseppina Andresini and Vito Recchia are supported by the project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI, under the NRRP MUR program funded by the NextGenerationEU. The authors wish to thank the remote sensing company WildSense for preparing the ground truth masks of the evaluation study.

Declarations

6.1 Conflict of Interests

The authors declare that they have no conflict of interest.

6.2 Competing Interests

No competing interests.

6.3 Ethics Approval

We declare that this submission follows the policies as outlined in the Guide for Authors. The current research involves no Human Participants and/or Animals.

6.4 Data, Material, and/or Code Availability

The source code is available at `https://github.com/gsndr/DIAMANTE`

6.5 Authors' contributions

Giuseppina Andresini: Conceptualization, Methodology, Software, Validation, Investigation, Supervision, Writing - original draft, Writing - review & editing **Annalisa Appice**: Conceptualization, Methodology, Validation, Visualization, Investigation, Writing - original draft, Writing - review & editing, Supervision, Project administration. **Dino Ienco**: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Vito Recchia**: Conceptualization, Methodology, Data curation, Software, Validation, Visualization, Investigation, Writing - review & editing

References

1. Abdullah, H., Skidmore, A.K., Darvishzadeh, R., Heurich, M.: Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (ips typographus, l.) compared with landsat-8. Remote Sensing in Ecology and Conservation **5**(1), 87–106 (2019)

2. Alshayef, M.S., Musthafa, M.: Identification of bark beetle infestation in part of bohemian forest using sentinel-1 time series insar. In: 2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS), pp. 5–9 (2021)
3. Altarez, R.D.D., Apan, A., Maraseni, T.: Deep learning u-net classification of sentinel-1 and 2 fusions effectively demarcates tropical montane forest's deforestation. Remote Sensing Applications: Society and Environment **29**, 1–21 (2023)
4. Andresini, G., Appice, A., Malerba, D.: 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 **16**, 10050–10066 (2023)
5. Andresini, G., Appice, A., Malerba, D.: Leveraging sentinel-2 time series for bark beetle-induced forest dieback inventory. In: The 39th ACM/SIGAPP Symposium on Applied Computing, SAC 2024, pp. 875–882. CAM (2024)
6. Audebert, N., Le Saux, B., Lefèvre, S.: Beyond rgb: Very high resolution urban remote sensing with multimodal deep networks. ISPRS Journal of Photogrammetry and Remote Sensing **140**, 20–32 (2018)
7. Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F.: Esa's sentinel missions in support of earth system science. Remote Sensing of Environment **120**, 84 – 90 (2012)
8. Bárta, V., Lukeš, P., Homolová, L.: Early detection of bark beetle infestation in norway spruce forests of central europe using Sentinel-2. International Journal of Applied Earth Observation and Geoinformation **100**, 102335 (2021)
9. Candotti, A., De Giglio, M., Dubbini, M., Tomelleri, E.: A Sentinel-2 based multi-temporal monitoring framework for wind and bark beetle detection and damage mapping. Remote Sensing **14**(23) (2022)
10. Capliez, E., Ienco, D., Gaetano, R., Baghdadi, N.N., Hadj-Salah, A.: Temporal-domain adaptation for satellite image time-series land-cover mapping with adversarial learning and spatially aware self-training. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. **16**, 3645–3675 (2023)
11. Ferreira de Carvalho, O.L., Olino de Albuquerque, A., Luiz, A.S., Henrique Guimarães Ferreira, P., Mou, L., e Silva, D.G., Abílio de Carvalho Junior, O.: A data-centric approach for rapid dataset generation using iterative learning and sparse annotations. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, pp. 5650–5653 (2023)
12. de Carvalho, O.L.F., de Carvalho Junior, O.A., de Albuquerque, A.O., Orlandi, A.G., Hirata, I., Borges, D.L., Gomes, R.A.T., Guimarães, R.F.: A data-centric approach for wind plant instance-level segmentation using semantic segmentation and gis. Remote Sensing **15**(5), 1–23 (2023)
13. Couprie, C., Farabet, C., Najman, L., Lecun, Y.: Indoor Semantic Segmentation using depth information. In: First International Conference on Learning Representations (ICLR 2013), pp. 1–8 (2013)
14. Dalponte, M., Cetto, R., Marinelli, D., Andreatta, D., Salvadori, C., Pirotti, F., Frizzera, L., Gianelle, D.: Spectral separability of bark beetle infestation stages: A single-tree time-series analysis using planet imagery. Ecological Indicators **153**, 110349 (2023)
15. Dalponte, M., Solano-Correa, Y.T., Frizzera, L., Gianelle, D.: Mapping a European spruce bark beetle outbreak using Sentinel-2 remote sensing data. Remote Sensing **14**(13) (2022)
16. Estrada, J.S., Fuentes, A., Reszka, P., Auat Cheein, F.: Machine learning assisted remote forestry health assessment: a comprehensive state of the art review. Frontiers in Plant Science **14** (2023)
17. Fernandez-Carrillo, A., Patočka, Z., Dobrovolný, L., Franco-Nieto, A., Revilla-Romero, B.: Monitoring bark beetle forest damage in central europe. a remote sensing approach validated with field data. Remote Sensing **12**(21) (2020)
18. Gomez, D.F., Riggins, J.J., Cognato, A.I.: Forest Entomology and Pathology: Volume 1: Entomology – Bark Beetles, pp. 299–337. Springer International Publishing, Cham (2023)
19. Hafner, S., Nascetti, A., Azizpour, H., Ban, Y.: Sentinel-1 and sentinel-2 data fusion for urban change detection using a dual stream u-net. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022)

20. Hinton, G.E., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. CoRR, abs/1503.02531 (2015)
21. Hoekman, D., Kooij, B., Quiñones, M., Vellekoop, S., Carolita, I., Budhiman, S., Arief, R., Roswintiarti, O.: Wide-area near-real-time monitoring of tropical forest degradation and deforestation using sentinel-1. Remote Sensing **12**(19), 1–32 (2020)
22. Hollaus, M., Vreugdenhil, M.: Radar satellite imagery for detecting bark beetle outbreaks in forests. Current Forestry Reports **5**, 240–250 (2019)
23. Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B.: More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. IEEE Transactions on Geoscience and Remote Sensing pp. 1–15 (2020)
24. Hu, J., Mou, L., Schmitt, A., Zhu, X.X.: Fusionet: A two-stream convolutional neural network for urban scene classification using polsar and hyperspectral data. In: 2017 Joint Urban Remote Sensing Event (JURSE), pp. 1–4 (2017)
25. Huo, L., Persson, H.J., Lindberg, E.: Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & swir (ndrs). Remote Sensing of Environment **255**, 112240 (2021)
26. Jamali, S., Olsson, P.O., Ghorbanian, A., Müller, M.: Examining the potential for early detection of spruce bark beetle attacks using multi-temporal sentinel-2 and harvester data. ISPRS Journal of Photogrammetry and Remote Sensing **205**, 352–366 (2023)
27. Kautz, M., Peter, F., Harms, L., Kammen, S., Delb, H.: Patterns, drivers and detectability of infestation symptoms following attacks by the European spruce bark beetle. Journal of Pest Science **96**, 403–414 (2022)
28. Konig, S., Thonfeld, F., Forster, M., Dubovyk, O., Heurich, M.: Assessing combinations of landsat, Sentinel-2 and Sentinel-1 time series for detecting bark beetle infestations. GIScience & Remote Sensing **60**(1), 2226515 (2023)
29. Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., Chanussot, J.: Deep learning in multimodal remote sensing data fusion: A comprehensive review. Int. J. Appl. Earth Obs. Geoinformation **112**, 102926 (2022)
30. Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., Gascon, F.: Sentinel-2 sen2cor: L2a processor for users. In: Proceedings of the Living Planet Symposium 2016, pp. 1–8. Spacebooks Onlin (2016)
31. Muszynski, M., Hölzer, T., Weiss, J., Fraccaro, P., Zortea, M., Brunschwiler, T.: Flood event detection from sentinel 1 and sentinel 2 data: Does land use matter for performance of u-net based flood segmenters? In: 2022 IEEE International Conference on Big Data (Big Data), pp. 4860–4867 (2022)
32. Nyborg, J., Pelletier, C., Lefèvre, S., Assent, I.: Timematch: Unsupervised cross-region adaptation by temporal shift estimation. ISPRS Journal of Photogrammetry and Remote Sensing **188**, 301–313 (2022)
33. Park, S.J., Hong, K.S., Lee, S.: Rdfnet: Rgb-d multi-level residual feature fusion for indoor semantic segmentation. In: The IEEE International Conference on Computer Vision (ICCV) (2017)
34. Patil, M.: Interpolation techniques in image resampling. International Journal of Engineering and Technology **7**, 567–570 (2018)
35. Phillips, J., Zhang, C., Williams, B., Jarvis, S.: Data-driven sentinel-2 based deep feature extraction to improve insect species distribution models. In: EGU General Assembly 2022, pp. 1–1 (2022)
36. Qian, X., Lin, X., Yu, L., Zhou, W.: Fasflnet: feature adaptive selection and fusion lightweight network for rgb-d indoor scene parsing. Opt. Express **31**(5), 8029–8041 (2023). DOI 10.1364/OE.480252
37. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, pp. 234–241. Springer (2015)
38. Roscher, R., Rußwurm, M., Gevaert, C., Kampffmeyer, M., dos Santos, J.A., Vakalopoulou, M., Hänsch, R., Hansen, S., Nogueira, K., Prexl, J., Tuia, D.: Data-centric machine learning for geospatial remote sensing data. CoRR **abs/2312.05327** (2023)
39. Sainte Fare Garnot, V., Landrieu, L., Chehata, N.: Multi-modal temporal attention models for crop mapping from satellite time series. ISPRS Journal of Photogrammetry and Remote Sensing **187**, 294–305 (2022)

40. Schmarje, L., Santarossa, M., Schröder, S.M., Zelenka, C., Kiko, R., Stracke, J., Volkmann, N., Koch, R.: A data-centric approach for improving ambiguous labels with combined semi-supervised classification and clustering. In: S. Avidan, G. Brostow, M. Cissé, G.M. Farinella, T. Hassner (eds.) Computer Vision – ECCV 2022, pp. 363–380. Springer Nature Switzerland, Cham (2022)
41. Solórzano, J.V., Mas, J.F., Gao, Y., Gallardo-Cruz, J.A.: Land use land cover classification with u-net: Advantages of combining sentinel-1 and sentinel-2 imagery. Remote Sensing **13**(18), 1–23 (2021)
42. Wu, J., Zhang, Y., Wang, K., Tang, X.: Skip connection u-net for white matter hyperintensities segmentation from mri. IEEE Access **7**, 155194–155202 (2019)
43. Zha, D., Bhat, Z.P., Lai, K., Yang, F., Jiang, Z., Zhong, S., Hu, X.: Data-centric artificial intelligence: A survey. CoRR **abs/2303.10158** (2023)
44. Zha, D., Lai, K., Yang, F., Zou, N., Gao, H., Hu, X.: Data-centric AI: techniques and future perspectives. In: A.K. Singh, Y. Sun, L. Akoglu, D. Gunopulos, X. Yan, R. Kumar, F. Ozcan, J. Ye (eds.) Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, pp. 5839–5840. ACM (2023)
45. Zhang, J., Cong, S., Zhang, G., Ma, Y., Zhang, Y., Huang, J.: Detecting pest-infested forest damage through multispectral satellite imagery and improved UNet++. Sensors **22**(19) (2022)
46. Zhang, Y., Sidibé, D., Morel, O., Mériaudeau, F.: Deep multimodal fusion for semantic image segmentation: A survey. Image and Vision Computing **105**, 1–17 (2021)
47. Zhou, W., Yue, Y., Fang, M., Qian, X., Yang, R., Yu, L.: Bcinet: Bilateral cross-modal interaction network for indoor scene understanding in rgb-d images. Information Fusion **94**, 32–42 (2023)