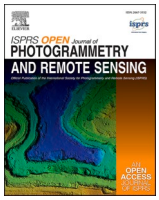
ISPRS Open Journal of Photogrammetry and Remote Sensing 6 (2022) 100024

Contents lists available at ScienceDirect

ISPRS Open Journal of Photogrammetry and Remote Sensing

journal homepage: www.journals.elsevier.com/isprs-open-journal-of-photogrammetry-and-remote-sensing



Instance segmentation of standing dead trees in dense forest from aerial imagery using deep learning

Abubakar Sani-Mohammed [a], Wei Yao [a,b,*], Marco Heurich [c,d,e]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
[b] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China
[c] Dept. for Visitor Management and National Park Monitoring, Bavarian Forest National Park, 94481, Grafenau, Germany
[d] Chair of Wildlife Ecology and Management, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany
[e] Department of Forestry and Wildlife Management, Campus Evenstad, Inland Norway University of Applied Sciences, Koppang, Norway

ARTICLE INFO

Keywords:
Carbon storage
CIR Aerial imagery
Forest management
Instance segmentation
Mask R-CNN
Standing dead tree

ABSTRACT

Mapping standing dead trees, especially, in natural forests is very important for evaluation of the forest's health status, and its capability for storing Carbon, and the conservation of biodiversity. Apparently, natural forests have larger areas which renders the classical field surveying method very challenging, time-consuming, labor-intensive, and unsustainable. Thus, for effective forest management, there is the need for an automated approach that would be cost-effective. With the advent of Machine Learning, Deep Learning has proven to successfully achieve excellent results. This study presents an adjusted Mask R-CNN Deep Learning approach for detecting and segmenting standing dead trees in a mixed dense forest from CIR aerial imagery using a limited (195 images) training dataset. First, transfer learning is considered coupled with the image augmentation technique to leverage the limitation of training datasets. Then, we strategically selected hyperparameters to suit appropriately our model's architecture that fits well with our type of data (dead trees in images). Finally, to assess the generalization capability of our model's performance, a test dataset that was not confronted to the deep neural network was used for comprehensive evaluation. Our model recorded promising results reaching a mean average precision, average recall, and average F1-Score of 0.85, 0.88, and 0.87 respectively, despite our relatively low resolution (20 cm) dataset. Consequently, our model could be used for automation in standing dead tree detection and segmentation for enhanced forest management. This is equally significant for biodiversity conservation, and forest Carbon storage estimation.

1. Introduction

Forest ecosystems are indispensable for the life of man; naturally, physically, and economically (Assessment, 2005). A healthy forest assists in providing a clean climate. Economically, wood alone contributes more than US $ 400 billion annually (Assessment, 2005; Gamfeldt et al., 2013). One of the most pressing phenomena that have been the focus of societies including renowned agencies like the United Nations (UN) has been Carbon storage and its cycles within the ecosystem as well as tree decomposition. It is important to note that dead trees have a great effect on Carbon storage and sequestration; the worst case of which could mar climatic conditions are large scale forest diebacks. Thus, understanding the condition of trees in the forest and their variation is imperative for effective forest management, healthy living, and prudent ecological decisions and policies. This will enable, for example, accurate forest carbon estimation and wood carbon fractions (Martin et al., 2021) for enhanced management. Consequently, a way to better understand the health of a forest is by identifying and quantifying its dead trees.

For the past few years, researchers have expounded on the causes and effects of forest tree health and disturbances. Microorganisms, insect infestation, droughts, wildfire, and extreme weather conditions have caused forest health deterioration and disturbances leading to ecological damages, increase in climate change and global warming (Laurance et al., 2000; Gamfeldt et al., 2013; Seidl et al., 2014; Zomer et al., 2016; Seibold et al., 2021; Einzmann et al., 2021). Likewise, numerous findings on Carbon storage and sequestration within the

* Corresponding author. Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.
E-mail addresses: abubakar.sanimohammed@connect.polyu.hk (A. Sani-Mohammed), wei.hn.yao@polyu.edu.hk (W. Yao), marco.heurich@npv-bw.bayern.de (M. Heurich).

https://doi.org/10.1016/j.ophoto.2022.100024
Received 14 August 2022; Received in revised form 28 October 2022; Accepted 28 October 2022
Available online 10 November 2022

forest ecosystem both regionally and globally have been reported. For example, healthy trees have been reported to play significant roles in carbon storage (Paniagua-Ramirez et al., 2021), as increase in tree cover resulted in global increase in biomass carbon (Zomer et al., 2016; Duque et al., 2021). Nowak and Crane (2002) analyzed the potential role of urban forests in reducing carbon dioxide in the atmosphere. Hulvey et al. (2013) advised for increase in tree species richness for improved carbon storage. Stephenson et al. (2014) analyzed 403 temperate and tropical tree species globally and showed that large old trees reserves larger amount of carbon compared to smaller trees. Further, Martin et al. (2021) showed that carbon fractions vary across forests based on tissue type, biomass type, stands and decay classes of dead woods. Also, forest age has significant effect on carbon storage and sequestration (Dai et al., 2021). Both young and old forests play important roles in storing the atmospheric carbon content. However, how well the forest can store the Carbon depends on the forest's health and its biodiversity. A healthy old forest plays a more significant role due to the magnitude of its biomass and canopy volume than a young forest that may contain less biomass because of young trees (Stephenson et al., 2014). But because a young forest is incorporating lots of carbon in wood that is building up, they play an important role in reducing carbon in the atmosphere (Nowak and Crane, 2002). On the other hand, in old stands there is a balance between building up biomass and the dieback of trees so that their net effect is not as big as from young stands. Therefore, avoiding and/or controlling disturbances and forest degradation is key for preserving a healthy environment and atmosphere. Forest disturbances arise through natural and anthropogenic activities. However, almost most of the disturbances in the European natural forests, especially in the central part, like the Bavarian Forest National Park (BFNP) are caused by insect infestation (Lausch et al., 2013; Latifi et al., 2014), which gradually spread affecting other healthy trees. Various methods have been used, ranging from classical to technological approaches, for dead tree identification and quantification (Senf et al., 2017). Although the classical (e.g., field survey) approach enables good observations and results, it is time-consuming, labor-intensive, and most importantly unsustainable for larger forest areas. Remote sensing techniques coupled with digital image analysis have been used to map forest structures and status, leveraging cost, time, and space. For example, satellite imagery was used to map tree mortality in the forest (Moustakas et al., 2006; Verbesselt et al., 2009; Garrity et al., 2013), indicating the effectiveness of satellite data in leveraging cost and time. Also, Lopes Queiroz et al. (2019) used aerial imagery to classify dead trees with the random forest algorithm. Other studies integrated the strength of the Light Detection And Ranging (LiDAR) with images to detect, classify and map dead trees (Kamińska et al., 2018; Meiforth et al., 2020; Maltamo et al., 2020). Despite these achievements from the remote sensing technologies, there is still the need for a more robust and efficient approach to dead tree detection and mapping, especially for individual trees(Yao et al., 2012).

With the advancement in computing, Artificial Intelligence (AI) through Machine/Deep Learning and Computer Vision (CV) has drastically improved object detection and segmentation with unprecedented accuracy (Zhu et al., 2017). Deep learning (DL) techniques have been proven to be more efficient in automation and predictions than the traditional Machine Learning (ML). Apparently, the need for an automated approach for tree detection in large coverage of land is very important because it is cost-effective and more efficient. Consequently, for the past few years, some DL models applied remote sensing data in tree detection and segmentation (Jiang et al., 2019; Polewski et al., 2020, 2021). However, to the best of our knowledge, there has been very little work done Chiang et al. (2020) in detecting and segmenting individual standing dead trees, especially in clumpy dense natural mixed forests. Cognizant of the fact that detecting and mapping dead trees in this forest would enable its efficient management, there is the need to apply a recent automated approach. This would guide the managers to control and keep healthy forests that can store more carbon from the atmosphere to mitigate global change. Thus, in this study, we propose a Mask Region Convolutional Neural Network (Mask R-CNN) framework (He et al., 2017) to detect and segment standing dead trees in the BFNP from Aerial Color Infrared (CIR) image. Our investigation seeks to provide answers to the following questions; (1) How well will DL, especially, the Mask R-CNN perform in the instance segmentation of standing dead trees in dense mixed forest from aerial imagery?; (2) How can we use limited dataset (per DL training custom) to achieve promising results?; (3) What hyperparameters would enhance the training process?, and (4) How will our model perform on unseen dataset? Our main objectives, therefore, include:

1. To adjust the Mask R-CNN for dead tree instance segmentation in dense mixed forests using limited training dataset with a relatively low resolution (20 cm) CIR aerial imagery
2. To exploit transfer learning technique and image augmentation in the training process to leverage the limitation of training dataset for improving the training process
3. To retrieve suitable hyperparameters for our type of dataset for improved model architectural performance
4. To apply the trained model on selected dead tree plots (test dataset that is not seen by the model during training) in clumpy dense mixed forest and comprehensively evaluate the results

Therefore, our adjusted Mask R-CNN model seeks to deal with the following challenges: (i) the samples used to train the main Mask R-CNN model (the pretrained model) include about 82 different scene features (e.g., cars, buildings, sky, human beings) excluding trees, not even urban trees. However, our samples (i.e., dead trees in dense forest) used for training vary greatly relative to the samples used for training the main Mask R-CNN model. Thus, we seek to know how well our model will adapt the transfer learning and be able to generalize our model's performance. (ii) In contrast to the required number of training samples needed for training deep learning models, we had had only a few (i.e., 195 images). Thus, in addition to the transfer learning, we seek to improve the performance of our model by exploiting the image augmentation technique. (iii) how well will our model perform, cognizant of the fact that the resolution of our image samples (datasets) is relatively low compared to the very high-resolution images (e.g., from UAVs and other sensors) used in similar deep learning architecture? (iv) Hyperparameters controls the learning process. Thus, we seek to also know how well is the performance and the generalization of our model as a result of our hyperparameter fine-tuning?

The remainder of this study is organized as follows: Accounts on related studies are presented in Section 2. Section 3 recounts the details of the materials used, including a description of the study area, and the methodological approach for achieving the results. The results are presented in Section 4 whilst a discussion of the results is elaborated in Section 5. Finally, the conclusion is stated in Section 6.

2. Related studies

This section accounts briefly for related studies that used Artificial Intelligence (AI). Accounts on applications using the traditional ML algorithms in contrast to the DL algorithms are provided. Also, brief descriptions of the Mask R-CNN model are presented, whiles recounting related studies that applied it in previous literature.

2.1. Traditional Machine Learning applications

Unlike DL, in traditional ML, computers are made to learn without explicit structural coding. Thus, its algorithms are relatively simple and less complicated. Some of the frequently used algorithms for classification in forestry include Random Forest (RF) (Immitzer et al., 2012), Support Vector Machine (SVM) (Ye et al., 2021), K-Nearest Neighbor (KNN) (Gjertsen, 2007), and Artificial Neural Networks (ANN) (Zhang et al., 2021). Recently, these ML algorithms have been applied to either

classify tree species or tree health (Ye et al., 2021; DeCastro et al., 2022). For example, Furuya et al. (2020) in their effort to map forest vegetation in the Riparian Zones compared the performance of four different algorithms and proposed the decision tree for that Zone. They used Sentinel-2 images. Yu et al. (2021) also applied the random forest to detect pine wilt using hyperspectral imagery in Yiwu City, Zhejiang Province, China. They used UAV-based data with LiDAR data. Their overall accuracy improved from 66.86% to 73.96% when they combined their image with LiDAR data. Zhang et al. (2021) after comparing four algorithms also reported the prowess of RF in their study for mapping dead trees resulting from pinewood nematode outbreaks. They conducted this investigation in the Dangyang City, Hubei Province of China, using three types of datasets; Geographical Information System (GIS) database, Sentinel-2 at 20 m resolution, and Radar derive Digital Elevation Model (DEM) and 30m resolution. Furthermore, Liu et al. (2021) employed SVM and RF coupled with a deep neural network algorithm to map standing dead trees based on a pixel and object-based image fusion technique. Just recently, Junttila et al. (2022) mapped the decline in tree health due to European spruce bark beetle infestation using multispectral imagery and an RF algorithm. They classified, healthy, declined, and dead trees with an overall accuracy of 78.2% and 84.5% for spring and fall respectively. Although these algorithms are efficient and produce acceptable accuracies, they are not robust enough and less efficient compared to the DL algorithms. For example, unlike DL, traditional ML algorithms need to be guided by an expert to extract features. Thus, they are not fully automated. Also, traditional ML algorithms, compared to DL, produce less accuracy, especially, in classifying more complicated features like trees (Arce et al., 2021). Moreover, it seems ML algorithms are effective more in detection than segmentation (Kattenborn et al., 2021). Consequently, there is the need to consider a higher automated approach, especially for large-scale studies like natural forests.

2.2. Deep learning applications

DL is a subset of ML. DL uses several hidden layers of neural networks that work like the human brain. For some years now, the potential of DL to independently learn higher-level features of varying scales in CV tasks has rekindled the face of artificial intelligence, especially in handling complicated problems in various sciences and engineering fields (Najafabadi et al., 2015; Nielsen, 2015; Buduma et al., 2022). Several DL techniques have been applied in forestry, geared toward forest management and conservation for a healthy environment. However, DL application in these areas is still in its early stage (Hamedianfar et al., 2022) despite the achievements so far, especially, in determining forest damages. Recently, DL algorithms, especially, the Convolutional Neural Network (CNN) have been applied to detect and segment dead trees and sick trees (Fricker et al., 2019; Khan et al., 2021). It is important to note that most of these studies took advantage of very high spatial resolution data resulting from the advancements in remote sensing technology. For example, Jiang et al. (2019) used a very high-resolution CIR aerial image (10 cm) in their Fully Convolutional Network (FCN-DenseNet) architecture to semantically segment dead trees in the BFNP. Briechle et al. (2020) used a combination of very high-resolution (8.9 cm) multispectral images and UAV-based LiDAR in their deep neural network (PointNet++) to classify standing dead trees and some trees species with an overall accuracy of 90.2%. Wu et al. (2021) in their study, also used very high-resolution UAV-based images (4 cm) in their assessment of two CNNs (Faster R-CNN and YOLOv3) for detecting pine wilt disease in plants.

Apparently, CNN has been a breakthrough in remotely sensed image object recognition and classification (Hamdi et al., 2019). However, following the comprehensive reviews on remotely sensed vegetation data by Hamedianfar et al. (2022) and Kattenborn et al. (2021), previous investigations have pointed to the fact that one of the key factors for CNN to fully exploit features in remotely sensed images is its spatial resolution. Analysis of the types of data used in publications from the year 2017–2020, on the application of CNN for vegetation in remote sensing, shows a sharp increasing trend in the use of UAVs (Kattenborn et al., 2021). UAV-based data have been the most frequently used, followed by terrestrial data, then satellite data, while airborne data represented the least considered. Seemingly, UAVs could have taken the lead because of their very high spatial resolution data. Likewise for the terrestrial data. However, the spatial resolution of satellite images, in general, is less compared to the airborne images; but why were satellite images used more frequently than the airborne data? A plausible explanation could be that satellite data are often used for large-scale studies (e.g., mapping land covers at national levels) with promising accuracies. In contrast to satellite data, airborne images are mostly used for feature-level studies (e.g., plot-level, or even individual tree level). So, it could be that more studies have previously been considered for larger scale application like Land Use Land Cover mapping.

2.3. Mask R-CNN applications

The instance segmentation-based Mask R-CNN (He et al., 2017) is one of the most frequently used state-of-the-art DL architecture that has cut across various fields of sciences and engineering (Yekeen et al., 2020; Zhang et al., 2020; Soloy et al., 2020; Ullo et al., 2021; Zhao et al., 2021; Gella et al., 2022) due to its ease, flexibility, and unprecedented accuracy. Mask R-CNN is a build-up from an earlier object detection FCN architecture known as Faster R-CNN (Ren et al., 2015) with an additional arm, in parallel with the existing classification arm, for delineating masks based on pixel alignments.

Despite its successful applications in various fields of sciences and engineering, over the past few years, to the best of our knowledge, very few studies have considered the Mask R-CNN in vegetation remote sensing (Machefer et al., 2020; Lv et al., 2020; Hao et al., 2021; Yu et al., 2022), especially in forestry, and more so in instance segmentation of standing dead trees. The reason for this could fall in line with the argument propounded by Kattenborn et al. (2021), that the complexity and continuous transition of tree crowns in natural forestry render it not applicable for instance segmentation. Notwithstanding, Peculiar to dead trees segmentation, Chiang et al. (2020) used aerial imagery with a modified Mask R-CNN to detect dead trees in the forest. They reached a COCO (Common Object in Context) mean average precision (mAP) of 54%. Braga et al. (2020) also proposed the use of Mask R-CNN on very high-resolution satellite imagery to delineate tree crowns in tropical forests with a Precision, Recall, and F1 reaching 0.91, 0.81, and 0.86 respectively. Zhang et al. (2022) applied the Mask R-CNN on high-resolution UAV-based imagery to segment multi-species individual trees in a young mixed less dense forest. Their results indicated better performance of the Mask R-CNN relative to the U-Net and YOLOv3 in segmenting coniferous crowns than deciduous crowns.

3. Material and methods

In this section, we give a detailed description of how the materials used were acquired for our investigations, as well as our approach to adjusting the Mask R-CNN algorithm for promising results. Also, a brief description of the study area is presented.

3.1. Study area

The area (Fig. 1) we investigated is situated in the BFNP (242.5 square km), located in the southeastern part of Germany, sharing a border with the Czech Republic (49.10°*N*, 13.22°*E*) (Heurich et al., 2009). This is the oldest national park in Germany, founded in 1970 with the aim of protecting its natural process without human interference (Nielsen et al., 2014). The topography of the area is undulating and highly elevated, ranging between 600 and 1,453 m. Thus, field surveying in this kind of terrain would be a herculean task to

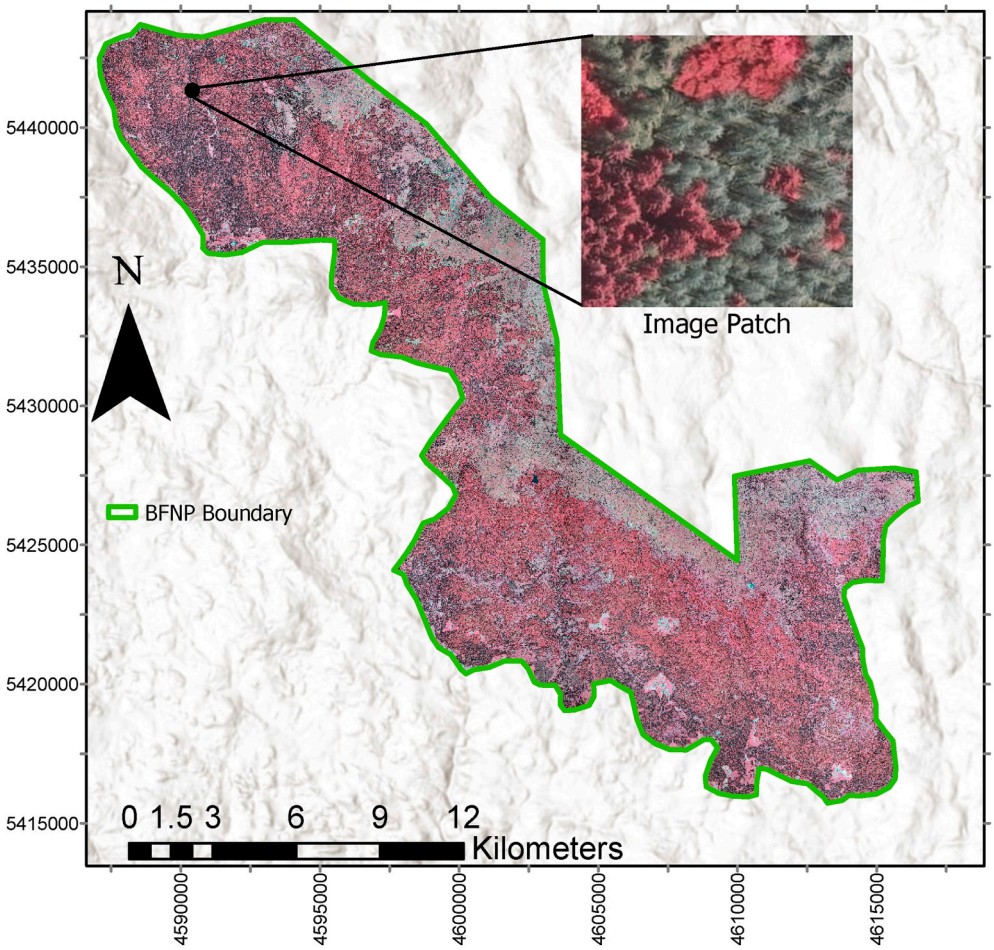


Fig. 1. Map of the study area, the Bavarian Forest National Park (BFNP) in gray to green extent. Insert is a sample CIR image patch (the dead trees in green while living trees in reddish to brown) of the 195 images used for the investigation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

accomplish; the more reason why an automated technological approach is required, especially, an airborne approach. The mixed forest is dense and clumpy with diverse species dominated by the Norway spruce (Picea abies) and European beech *(Fagus sylvatica)*. Other species include, White fir *(Abies alba)*, Common rowan *(Sorbus aucuparia)*, birches *(Betula pendula, Betula pubescens)*, and Sycamore maple *(Acer pseudo-platanus)* (Cailleret et al., 2014). The forest has witnessed some disturbances over the years, especially, insect infestation by Norway Spruce bark beetle *(Ips typographus)* (Lausch et al., 2013). Cognizant of the fact that the aim of the park was to protect the natural process of the forest, which could, in turn, preserve the forest for carbon storage leading to a clean environment, there is the need to find a cost-effective way of understanding the forest health and mapping its standing dead trees for better management.

3.2. Data acquisition

The dataset, CIR images, used for this study was acquired by a DMC 122 camera onboard a Cessna 207 aircraft. The flight campaign was conducted by ILV Remote Sensing GmbH on June 23, 2016, from 09:15 to 15:10 (with a fuel stop from 12:15 to 14:00) clock over the BFNP at an average altitude of 2918 m above Mean Sea Level. The images have three spectral bands (Near Infrared, Red, and Green) with a ground resolution of 20 cm. The images were radiometrically corrected and orthorectified. This was conducted using optimal camera calibration observations, transformation parameters and ground control points. This was achieved using the Trimble/INPHO company's program system OrthoBox (Orthovista, Orthomaster). Table 1 shows the flight and acquisition details.

3.3. Our methodology

In this study, we used a very limited training dataset at 20 cm resolution to detect and delineate standing dead trees with our Mask R-CNN model. The following processes were followed to achieve our objectives: (i) First, the images were prepared to a specified size (256 by 256 pixels).

Table 1
Flight and data acquisition details.

Item	Details
Project	Bavarian Forest
Assignment	National Park Administration of Bavarian Forest
Flight date	June 23, 2016
Flight time	09:15–12:15; fuel stop; 14:00–15:10 clock
Ground resolution	20 cm
Average altitude	2918 m MSL
Longitudinal overlap	75%
Cross coverage	60%
Camera	DMC 122
Focal length	120 mm
Camera calibration	2014
Aircraft	Cessna 207
Exposure	Automatic
Navigation	CCNS4, AeroControl
Flight conduct and processing	ILV Remote Sensing GmbH

(ii) Then, the dataset (training, validation, and testing sets) was prepared. (iii) Next, the standing dead tree crowns were manually annotated (labeled) based on the visual inspection due to the spectral strength of the CIR images and expert advice. (iv) After that, the annotated dataset were used in our Mask R-CNN algorithm for training, validation, and testing for standing dead tree detection and delineation. In this step, we employed the image augmentation strategy, transfer learning, and appropriate hyperparameter definition to fine-tune our model. A summary of the framework approach is illustrated in Fig. 2.

3.3.1. Image preparation

In order to have an ease in the training process in the DL network, the processed image tiles acquired from the DMC 122 camera were further resized to patches of 256 by 256 pixels. There were no overlaps between the image tiles and patches used for the experiment. This was carried out using the IrfanView version 4.57 tool (Skiljan, 2022).

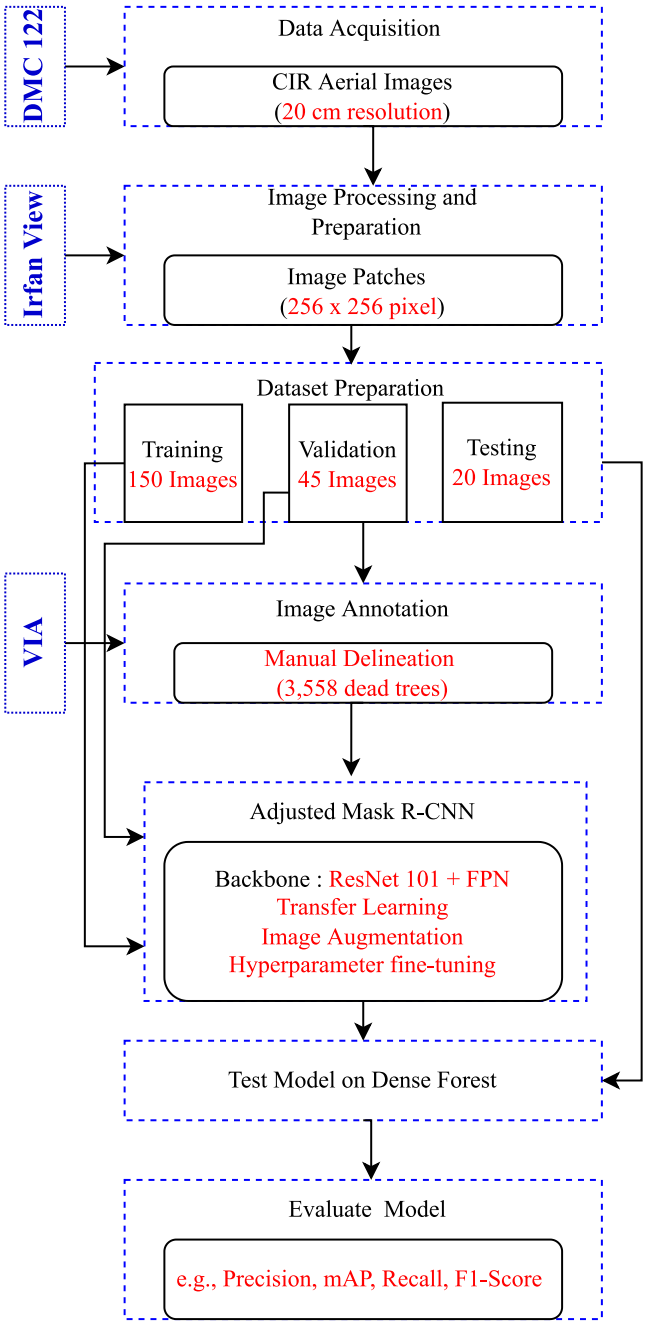


Fig. 2. Flowchart of our methodology framework for adjusting the Mask R-CNN algorithm for instance segmentation of standing dead trees in dense forests.

3.3.2. Dataset preparation

Before training our model, we randomly grouped 215 images into training (150 images), validation (45 images), and testing (20 images) sets. These sets were put in different folders. The training and validation sets were used for training the DL network, while the testing set did not see the network at all.

3.3.3. Dataset annotation

In order to train DL models, the network architecture would need the ground truth patterns from which it would learn, train and validate the predictions. Consequently, we manually annotated all dead tree crowns found on each patch of the images, to the best of our ability. At the borders of the image patches, we considered trees that were more than half of their full size for annotation. The approximate sizes of trees are between 7 by 7 pixels and 18 by 18 pixels in the image patches. All images in the three contained folders were annotated using the VGG Image Annotator (VIA) (Dutta and Zisserman, 2019), an open-source web-based image annotator that does not depend on any external libraries nor requires any installations (see Fig. 3). This tool was developed in such a way that it can be applied offline in some modern web browsers like Chrome and Firefox. Our annotated dead tree crowns were exported as a JSON file for convenient use in our scripts. The annotated images had at least one annotated dead tree and a maximum of 70 annotations.

3.3.4. Our mask R-CNN model

Mask R-CNN (He et al., 2017) builds up from Faster R-CNN (Ren et al., 2015) (which was proposed for detection problems) by adding an arm in parallel to that of Faster R-CNN for delineating masks. The masks are delineated for all areas of interest known as the Region of Interest (RoI). In contrast to most semantic segmentation DL architecture, the Mask R-CNN has less pooling operation (see Fig. 4). This is in order to minimize the possibility of losing some details from the image being investigated. With its backbone (ResNet), Mask R-CNN extracts features from the input images, whilst using its head to classify the features by describing bounding boxes (He et al., 2017). Fig. 4 illustrates the Mask R-CNN architectural design.

This works in two stages. In the first stage, the training image input into the network architecture is scanned to propose regions that are likely to contain objects. In our case, for instance, this proposed region could contain a standing dead tree. The regions are proposed referenced to the ground truth annotations. In the second stage, the proposed regions are classified by generating bounding boxes and masks for each region. Feature maps are extracted from the images with the backbone architecture made up of CNN. The Mask R-CNN applies ResNet (ResNet 50 or ResNet 101) and Feature Pyramid Network (FPN) (Lin et al., 2014) as the backbone. As the names sound, ResNet 50 has 50 CNN layers whiles ResNet 101 has 101 layers. In our case, we applied the ResNet 101 together with the FPN. The initial layers start by extracting low-level features (e.g., lines, and curves) while succeeding layers extract higher-level features until the required dead tree crowns. Furthermore, the proposed feature maps are then scanned by the Region Proposal Network (RPN) in a sliding window style to effectively determine the targets (e.g., dead trees) in the feature maps. These scans are known as ANCHORS. Predictions by the RPN indicate the possibility of targets existing in an anchor or not. Thus, bounding anchors with targets are further localized and resized based on the ones predicted with the highest score to exactly fit the targets (i.e., Region of Interest (ROI)); the remaining anchors are then rejected. Next, the ROIs predict the class (in our case we have two classes: dead tree or background) and the bounding boxes (with refinement). In Mask R-CNN, ROIAlign was applied to refine the bounding boxes with bilinear interpolations.

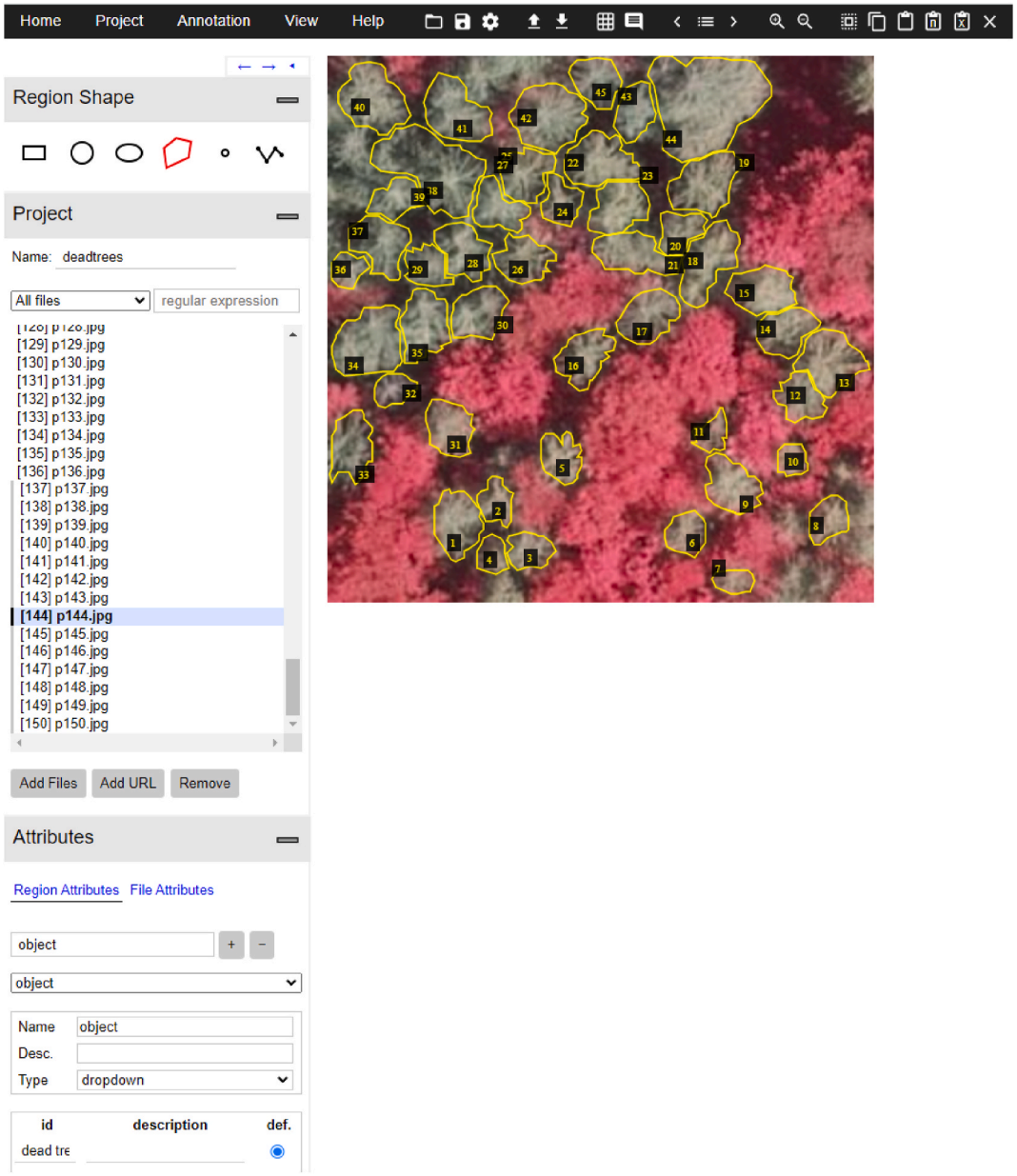


Fig. 3. The VGG Image Annotator interface displays a sample of the annotation on a CIR image. The dead trees are shown gray to green with the yellow polygons around them indicating the delineation, while the healthy trees are red. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Finally, masks are created for predicted ROIs with positive classes.

(i) Transfer Learning: DL architecture like the Mask R-CNN requires a lot of training data to perform well. Previous research suggested the use of more than 10,000 well-labeled images for each class for attaining sound instance segmentation results (Zlateski et al., 2018). Due to the limitation of our dataset (195 images), transfer learning was employed. The concept of transfer learning is to initiate training with an already-trained model that has weights. It is expected that the learned model would have enough knowledge to transfer to the new model for easy training. Thus, we initiated our training with a model pre-trained with the Microsoft Common Object in Context (MS COCO) dataset (Lin et al., 2014). The MS COCO dataset contains more than 300,000 images with more than 2 million labeled instances used for training with about 90 different classes of natural scenes (e.g., human beings, cars, tables, and animals). However, there is no representation of our class (dead tree) in the MS COCO dataset. Notwithstanding, since the pre-trained model from MS COCO has learned about 90 classes within the natural environment, we presumed that this knowledge could help in improving the training of our limited data as indicated in previous research (Zlateski et al., 2018). Thus, we seek to know how well our model can adapt to the transfer learning strategy.

(ii) Image Augmentation: A strategy to artificially increase the number of training data is via image augmentation. To do this, we exploited the image augmenters (imgaug (Jung et al., 2020)) in the Python library. We executed zero to three (0–3) of the five given augmenters (they include, horizontally flipping 50% of the images; vertically flipping 50% of the images; one of Affine rotation at either 90, 180, or 270°; changing the brightness of images between 80 and 150%, and Gaussian blur with sigma between 0 and 2). This was done randomly, as augmentation applies to the images with their corresponding annotations. Thus,

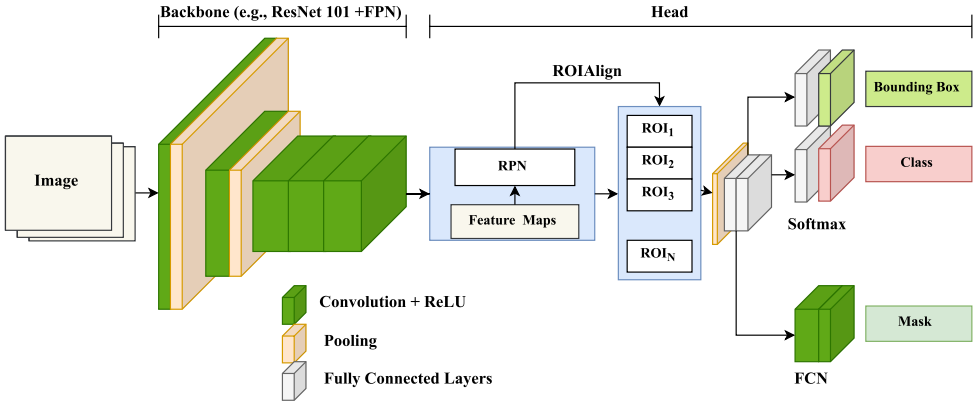


Fig. 4. The mask R-CNN architecture.

we applied this technique with the goal that it could help in leveraging the limitation of our dataset.

(iii) Hyperparameter fine-tuning: In contrast to traditional ML where the machine is allowed to automatically opt for the optimal model architecture, in the case of DL, one needs to search for the parameters to define the model's architecture. These parameters are known as hyperparameters. Hyperparameter fine-tuning is the means of exploring the ideal model architecture (Jordan, 2017). It is important to understand the difference between hyperparameters and model parameters. On one hand, the hyperparameters characterize how the model is structured. On the other hand, the model parameters describe how to get the required output by transforming the input data. In general, investigations are carried out to find what hyperparameters work best because there are no specific ways to update hyperparameters for loss reduction.

That said, the matterport's Mask R-CNN (Abdulla, 2017), from which our model was adjusted, used a training dataset (MS COCO dataset) that had great variations in scale because they had about 90 classes (cars, tables, chairs, sky, human beings, and other natural scene objects) in their dataset. This differs from ours in that our class consists of only trees with fewer variations in scale. Thus, in our case, applying the same hyperparameters implemented in the matterports's Mask R-CNN would not lead to good results (see Fig. 6). Thus, to fine-tune our hyperparameters, we re-scaled the anchors for the RPN (RPN Anchor Scales) by reducing the sizes. This was in order to have reasonable anchor sizes that could fit the standing dead trees, which are estimated to be approximately between 7 by 7 pixels and 18 by 18 pixels in the image patches. Furthermore, since our maximum annotated dead trees per image was not more than 100, we limited the maximum ground truth instances (Max GT Instances) to 100. Additionally, we limited the required region of interest necessary to be trained per image (Train ROIs Per Image). Equally, we also limited the required number of anchors necessary per image for the RPN (RPN Train Anchors Per Image), with an anchor stride (RPN Anchor Strides) of two since there is no possibility of having more than 60% tree crown overlap. Moreover, we scaled the dimensions of our images, (Image Min DIM) and (Image Max DIM), to the size of our patches (i.e., 256 by 256 pixels). Further, we increased the non-maximum suppression threshold (RPN NMS Threshold) to generate more proposals to filter the RPN proposals. Then, we limited the number of ROIs (for training (Post NMS ROIs Training) and inference (Post NMS ROIs Inference)) that would be kept after the NMS. We use a batch size (Batch Size) of eight and maintained all other parameters. Table 2 illustrates the details of our hyperparameter fine-tuning. This hyperparameter fine-tuning suits our dataset, leading to an improvement in the training to achieve promising results.

Table 2
The parameters used for our hyperparameter fine-tuning.

Hyperparameter	Modification	Description
Backbone	ResNet101 + FPN[a]	The backbone CNN used to extract the feature maps
Batch Size	8	Trained 8 images per GPU
Backbone Strides	[4, 8, 16, 32, 64]	Strides based on the resnet 101 backbone for each layer of the FPN pyramid
RPN Anchor Scale	(8, 16, 32, 64, 128)	Lengths of the different square anchor sides used, in pixel
RPN Anchor Stride	2	Anchors created for every other cell in the feature map
RPN NMS Threshold	0.8	Non-maximum suppression threshold to filter RPN proposals
Post NMS ROIs Training	1600	Number of ROIs kept after NMS for training
Post NMS ROIs Inference	600	Number of ROIs kept after NMS for inference
Image MIN Dimension	256	Minimum dimension of the images used for training,in pixel
Image MAX Dimension	256	Maximum dimension of the images used for training, in pixel
Train ROIs Per Image	100	Number of ROIs per image to pass to the classifier and mask heads
MAX GT Instances	100	The maximum quantity of ground truth instances to consider in one image
Detection MAX Instances	100	Maximum number of final detections
Detection MIN Confidence	0.6	The least probable value to accept a prediction

[a] Same parameters used in main architecture.

4. Results

In this section, we present a comprehensive evaluation of the experimental results of our investigations.

4.1. Training experiment

Experiments were investigated on the training and validation datasets. This was conducted on a desktop with 64 GB installed RAM; Intel® CoreTM i7-10700 CPU at 2.90 GHz and NVIDIA GeForce RTX 2070. The evaluation was not seen by the network and was therefore set aside to assess the generalization of our model, pursuant to the use of the transfer learning and data augmentation approach. The model was trained for 25 epochs with 100 steps per epoch for training and 25 steps per epoch for validation steps at a learning rate of 0.001. The Adam Optimizer was used for optimizing the training. Fig. 5 shows the loss graph.

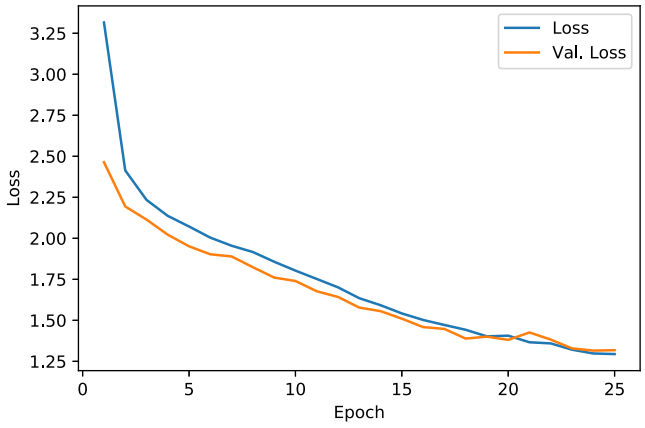


Fig. 5. The training loss.

4.1.1. Loss function

Whiles in the training process, the general loss function is calculated by computing the sum of the classification loss, bounding box regression loss, and the mask loss by applying per pixel sigmoid to each feature map (Equation (1)). The classification and bounding box loss are the same as described in the literature on Faster R-CNN (Ren et al., 2015). The mask loss is characterized by a log loss as an average binary cross-entropy loss from all the results output (Equation (2)). However, the results are considered only for regions with the ground truths (in this case the annotated dead trees).

$$L = l_{cls} + l_{box} + l_{mask} \tag{1}$$

where L is the total loss, l_{cls} is the classification loss of predictions, l_{box} is the bounding box regression loss of the predictions, and l_{mask} is the average binary cross-entropy loss.

$$l_{mask} = -1/y \sum_{i=1}^{n} [y_i^* log(p(y_i)) - (1 - y_i^*) log(1 - p(y_i))] \tag{2}$$

where y is the pixel number, y_i^* is the ground truth where the pixel is located, while $p(y_i)$ is the probability of the (y_i) predicted category (Wang et al., 2021).

4.2. Standing dead tree detection and segmentation

Fig. 6 shows the predictions obtained from the model without hyperparameter fine-tuning. The left side (a) shows the original images used for the evaluation, while the right side (b) shows the detections and segmentations from the model. All the 20 test images were evaluated. However, predictions/instances were seen for only the images (eight) displayed in Fig. 6 (labeled with their IDs). Even, most of these images show just a couple or three instances (see 1b, 3b, 11b, 16b, 18b and 20b), while 2b and 10 displayed only one instance each. The remainder of the images displayed zero predictions/instances, thus we did not provide them in this Fig. 6. This indicates the weakness of the model without suitable hyperparameters. Compare this to Fig. 7.

Despite the challenges with relatively low resolution CIR images used, our model produced promising results with very high accuracy, exceeding some previous studies in the literature. Fig. 7 displays our model's results for standing dead tree instance segmentation with masks for all the 20 evaluated test images. As illustrated in Fig. 7, the left side (a) shows the original images used for the evaluation, while the right side (b) shows the detections and segmentations from our new model. From visual inspection, it could be seen that our model's performance looks promising. Apparently, our new model could detect and segment almost 90% and even more of the standing dead trees available in some images (See Fig. 7; 1b, 2b, 3b, and 10b). Likewise, in the same figure, for evaluated images with IDs 11b, 15b, 16b, 17b, 18b, and 20b. Interestingly, also in the remainder of the images, almost all standing dead trees available are detected and segmented (see Figs. 7, 4b and 5b, 6b, 7b, 8b, 9b,12b, 13b, and 14b). This indicates our model's improved strength in detecting and segmenting standing dead trees in the mixed, dense, and clumpy Bavarian Forest. Notwithstanding, we had some challenges, especially, in annotating samples for training. Despite the expert guide, annotation in some images proved quite challenging due to the tree patterns coupled with the image resolution. Consequently, some standing dead trees were intentionally ignored, thus not obtaining ground truth for them. Also, there were areas where the ground's spectral color looks like that of the dead trees (gray to green) due to the presence of fallen dead trees. All this could be the reason for seeing low Average Precision (*AP*) and *F*1 − *Score* for Images 3b and 6b, as well as 7b and 9b in Fig. 7 (see Table 3) even though the results are acceptable. This may have also influenced the accuracy assessments seen in Table 3, despite the tremendous performance displayed in Fig. 7.

4.3. Accuracy assessment

We evaluated the performance of our model's predictions on the detection and segmentation of standing dead trees in the BFNP. Thus, we measured the Average Precision (*AP*) for each image and mean Average Precision (*mAP*), *Recalls*, and *F*1 − *Scores* for the entire images. Here, the *Precision* measures our model's accuracy in correctly classifying the standing dead trees in the image. It is expressed as the number of correctly classified standing dead trees (True Positive) relative to the entire classified trees in an image whether correctly classified or not (True Positive + False Positive). The *Precision* is expressed as in Equation (3):

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

where *TP* is True Positive, and *FP* is False Positive. The *mAP* is calculated as described in Equation (4); the results are shown in Table 3:

$$mAP = 1/n \sum \frac{TP}{TP + FP} \tag{4}$$

where n is the total number of tested images.

The *Recall* defines our model's ability to correctly classify the standing dead trees. It is expressed as the number of correctly classified standing dead trees (True Positive) relative to the entire number of standing dead trees available in an image (True Positive + False Negative). Therefore, if the *Recall* value is high, it means our model's tendency to correctly detect and classify standing dead trees is very high. The *Recall* is expressed as in Equation (5):

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

where *FN* is the False Negative.

*F*1 − *Score* is a balance between *Precision* and *Recall*. It is also known as the harmonic mean of *Precision* and *Recall*. Thus, a high *F*1 − *Score* value means the model's Precision and *Recall* are high as well. *F*1 − *Score* is expressed as in Equation (6):

$$F1 = 2\left(\frac{Precision * Recall}{Precision + Recall}\right) \tag{6}$$

Table 3 presents results for our model's entire evaluated dataset. The Image IDs, Average Precisions (*AP*), *Recalls*, and *F*1 − *Scores* for the test images are displayed in columns one, two, three, and four respectively. Also presented is the overall *mAP*, mean Recall (*mR*), mean F1-Score (*mF*1) for the entire images. Each row display results for an image whose ID is stated in the first column. The results range between 0 and 1; the closer the value is to 1 the better the performance, whereas values

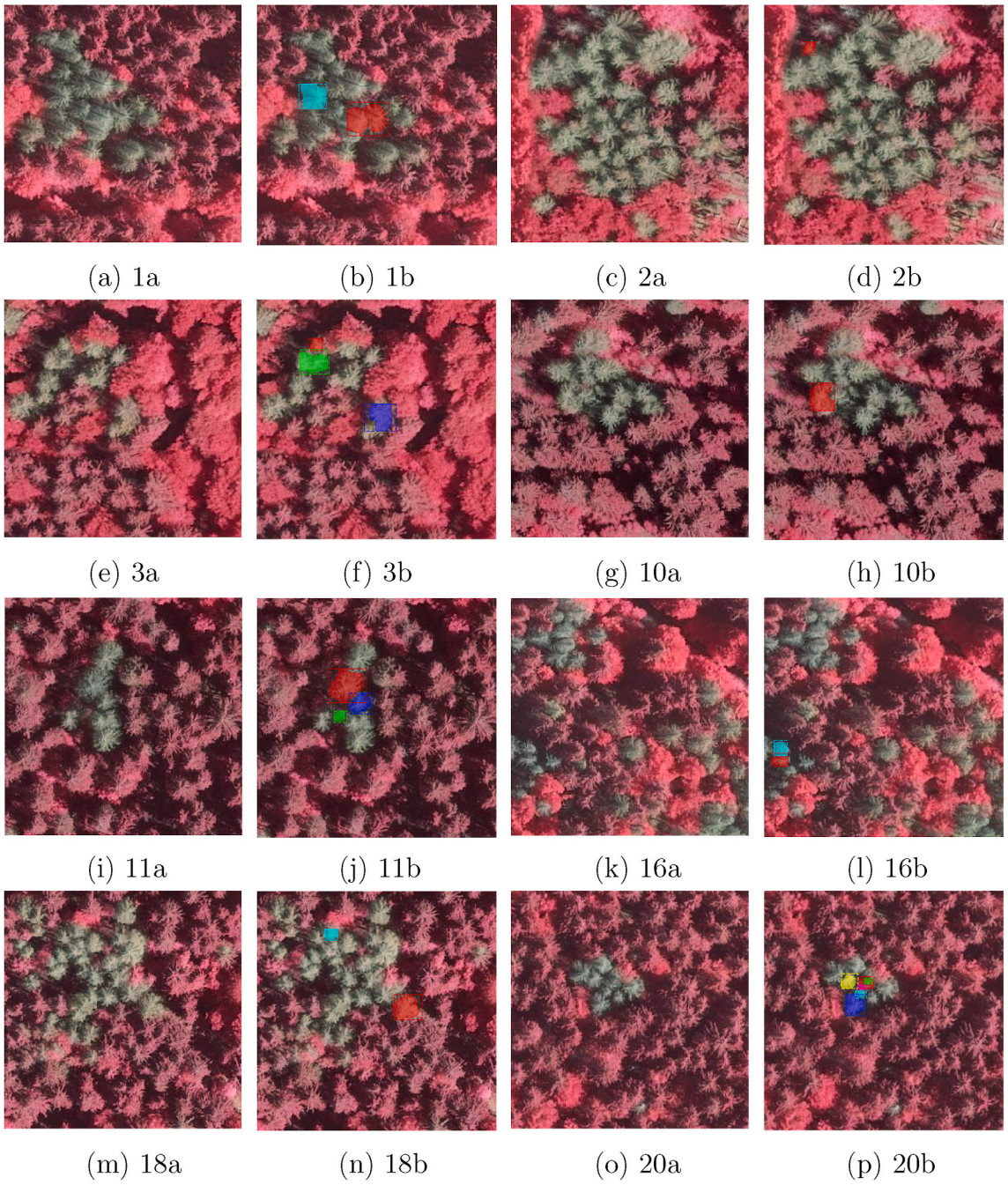


Fig. 6. Instance segmentation of standing dead trees (b; right); corresponding tested images (a; left), for model without hyperparameter fine-tuning.

close to zero indicate weak performance. As indicated in Table 3, about 90% of the evaluated dataset had a *AP* greater than 0.7 while 35% reaching *AP* of greater than 0.90. Apparently, 80% of the evaluated dataset recorded *AP* greater than 0.8. Interestingly, our model recorded an *mAP* of 0.85 for the entire evaluated dataset. Further to the promising *mAP* results we have, we also recorded promising Recalls. The least recorded Recall was 0.67 for only two evaluated images, while the remaining dataset recorded Recall values between 0.7 and 1 with 85% of the data recording Recalls greater than 0.8. An (*mR*) of 0.88 was recorded for the entire evaluated dataset. This, therefore, indicates that our model's tendency to correctly detect and classify standing dead trees is very high (0.88). Like the promising results recorded for the *mAP* and *Recalls*, our model equally recorded promising results for the *mF*1 at 0.87 for the entire evaluated dataset. However, the model without suitable hyperparameters recorded an overall *mAP*, *mR*, and *mF*1 reaching 0.05, 0.06 and 0.05 respectively (see Fig. 6 for display of its predictions).

5. Discussion

In this investigation, we have attempted to detect and segment standing dead trees in the mixed, dense, and clumpy BFNP from aerial CIR imagery using a DL approach. We adjusted the Mask R-CNN to suit our dataset pattern for improved performance (Compare Fig. 6 with Fig. 7). Apparently, for a DL architecture to perform very well there is a need for tens of thousands of training datasets, which, most often, is quite challenging. For a supervised DL technique like ours, it means all these datasets must be labeled, thus an additional challenge. Further to this, most often, DL requires very high-resolution datasets to be fed into the neural networks for robust model performance (Kattenborn et al., 2021), especially for mapping individual trees. We were challenged with

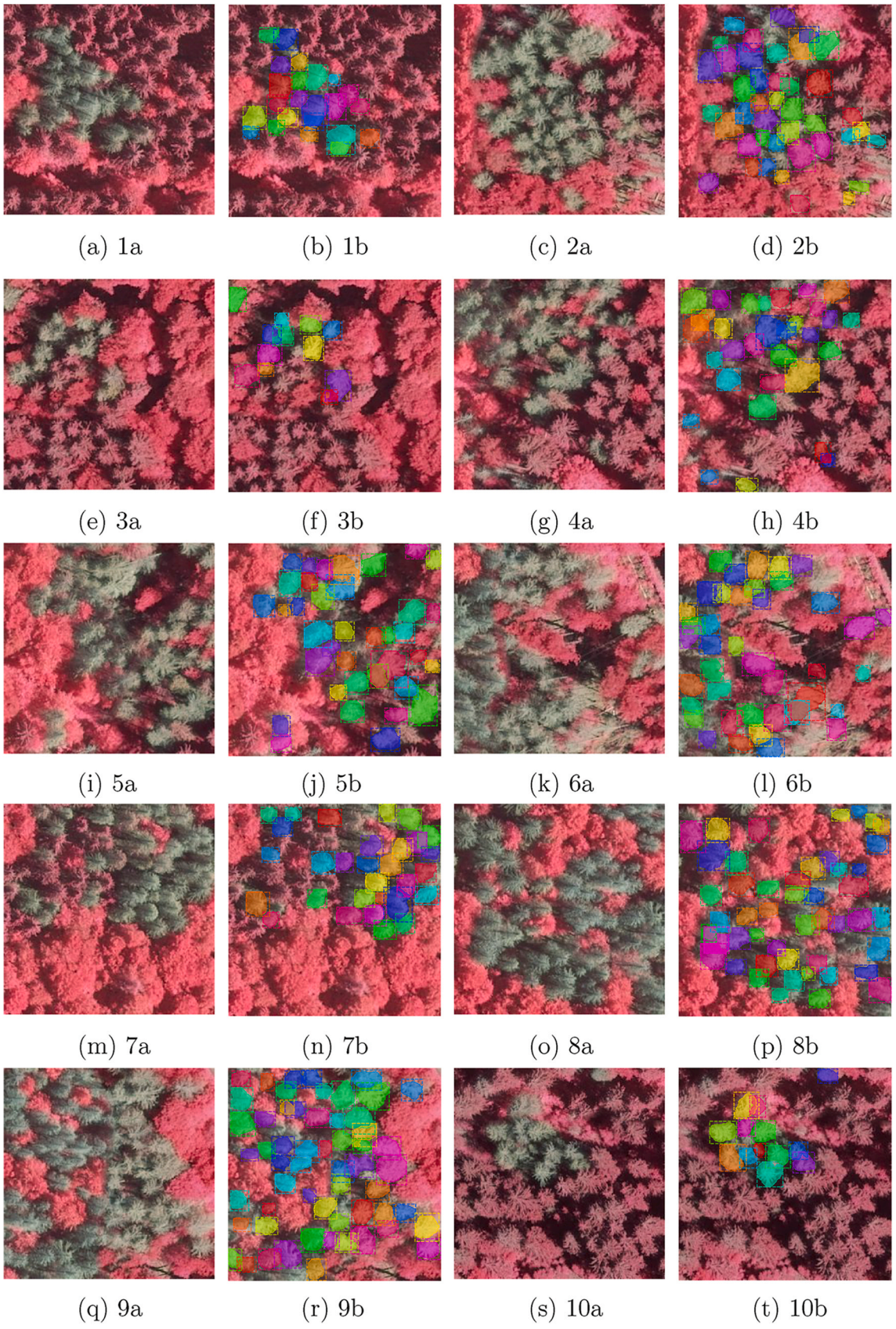


Fig. 7. Our new model's instance segmentation of standing dead trees (b; right); corresponding tested images (a; left).

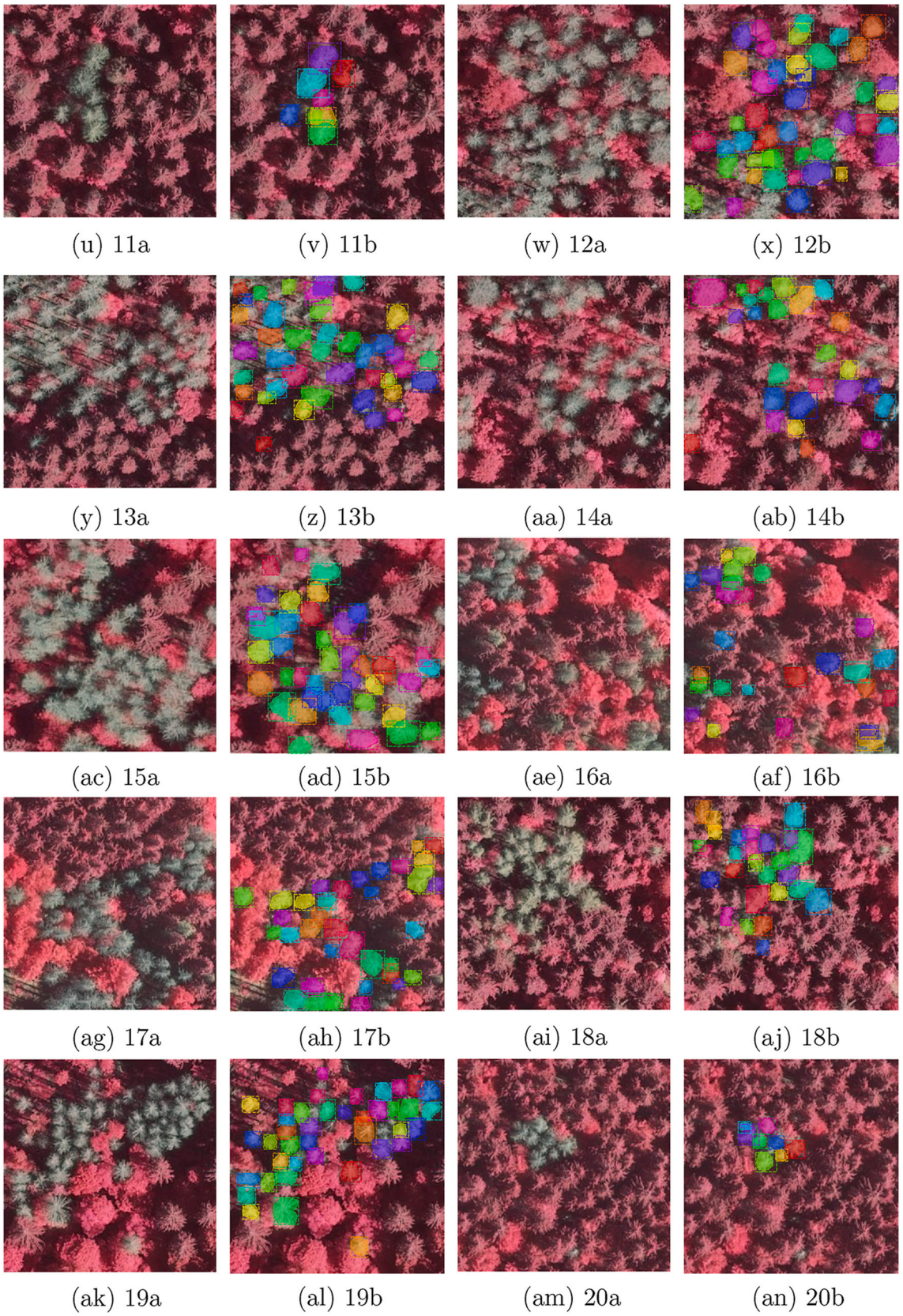


Fig. 7. (*continued*).

very limited training dataset. Despite this challenge, we employed techniques to leverage this limitation, leading to promising results.

We noted quite a consistent improvement in the training process while the epochs increase from 1 through to 25 as the training loss decreased (Fig. 5). This corroborates with the results we have seen in our evaluation of the test dataset (Table 3). The worst performance was seen

Table 3
Our model's evaluation performance of the Precision, Recall and F1-Score.

Image ID	**Average** Precision (AP)	**Recall**	**F1**-Score
1	0.85	0.83	0.84
2	0.86	0.86	0.86
3	0.67	0.67	0.67
4	0.86	0.96	0.91
5	0.84	0.90	0.87
6	0.66	0.67	0.66
7	0.71	0.77	0.74
8	0.89	0.91	0.90
9	0.73	0.84	0.78
10	0.90	0.90	0.90
11	1.00	1.00	1.00
12	0.94	0.95	0.95
13	0.88	0.96	0.92
14	0.93	1.00	0.96
15	0.86	0.87	0.86
16	0.91	0.92	0.91
17	0.94	0.93	0.93
18	0.94	0.96	0.95
19	0.81	0.87	0.84
20	0.88	0.88	0.88
Overall	**mAP**	**mean** Recall (mR)	**mean** F1-Score (mF1)
	0.85	**0**.88	**0**.87

in two images (3b and 6b). The reason for such low performance could be due to the challenge in labeling standing dead trees in the images, as the trees were very clumpy, making them difficult for manual delineation. Thus, some trees were not labeled, so not having ground truths. However, our observations show that our model detected and segmented most of the standing dead trees in every image. This could be the reason for that low performance, even though the results are acceptable. This, therefore, indicates our model's robustness for detecting and segmenting standing dead trees in dense forests. Contrary to the overall accuracy performance (0.05, 0.06, and 0.05) of the model without suitable hyperparameters, our new model recorded an overall mean Average Precision (*mAP*), *mR*, and *mF*1 of 0.85, 0.88, and 0.87 respectively in detecting and segmenting approximately 561 standing dead trees from the test dataset. This is very promising and higher than the 0.82 recorded in (Lv et al., 2020). Chiang et al. (2020) also reached a COCO *mAP* of 0.54. This is very significant, especially, in investigations relating to mapping standing dead trees in a natural dense forest like the Bavarian in Germany. This further indicates that our transfer learning and image augmentation approach worked successfully in leveraging the limitation in our data. Likewise, our selected hyperparameters have suited well the structure of our training dataset even though the COCO dataset used for transfer learning were quite different in scale, nature, and variation. Our remarkable results also point to the generalization capability of our model. Observation of our new model's training (see Fig. 5) indicates that longer training time could have enabled it to reach even better results than the ones presented, as the training and validation losses were still lowering. Thus, in our future works we will rather implement early stopping criterion in order to let the algorithm decide, automatically, when to stop training instead of the fixed approach we considered.

One of the challenges we had, however, was the difficulty in annotating the standing dead trees. It was quite challenging trying to delineate all the crowns due to uncertainty in knowing exact boundaries in some instances. Thus, some few crowns were not annotated. There were, also, some areas where the ground's spectral color and the dead trees (gray to green) looks similar due to the presence of fallen dead trees. Some of these ground areas (although very few) were predicted by our model as dead trees. This may have contributed to the low Average Precision (*AP*) and $F1 - Score$ for Images 3b and 6b, as well as 7b and 9b in Fig. 7 (see Table 3) even though the results are acceptable. Also, all this is likely to influence the accuracy assessments seen in Table 3, despite the tremendous performance displayed in Fig. 7, as the base for the assessments are the ground truths. Thus, we argue, that If we had attained perfection in our manual annotation, our results could have been higher than in Table 3. Another limitation is our inability to find dataset from other regions (outside Germany) to equally test our model's performance elsewhere. Thus, we are not very sure how it will perform on forests in regions outside Germany. Moving forward, we hope that data accessibility would become easy in the future, so that we could ascertain this performance in later studies.

Regarding the instance segmentation of standing dead trees, especially in dense forests using DL, to the best of our knowledge, the study we could find in literature that could be compared with our investigation was the work of Chiang et al. (2020). However, we could not have a quantitative comparison with their work as they used the COCO *mAP* (reaching COCO *mAP* of 0.54), which is a bit different metric to ours. This was their best out of eight fine-tuned models. It was, however, interesting to note that, qualitatively, we observed that our masks delineated our tree crowns quite better, depicting a real sense of instance segmentation compared to most of their test images as displayed in their publication (see their Fig 20). This indicates that, unlike ours, perhaps their focus was not really on the individual dead tree instance segmentation. Like us, they also used aerial imagery, with synthetic dataset for training, in their investigation, without stating the resolution of their image. Also, visual inspection of the lower left corners of our Images 1b, 5b, 6b, 8b, and 9b in Fig. 7, for instance, indicate the segmentation of overlapping crowns. There are other few studies that applied DL instance segmentation, in vegetation, in general. For instance, Lv et al. (2020) modified the Mask R-CNN to delineate and grade crop production using remote sensing data. They used Landsat 8 OLI and Land Use Change vector database for their investigation in the Da'an City, Jilin Province of China. They recorded *mAP* of 0.82 which is a bit low compared to ours. It is also important to note that their investigation was not focused on individual trees. Machefer et al. (2020), also refitted the Mask R-CNN to individually segment plants from RGB UAV imagery recording *mAP* of 0.66 for lettuce. Their datasets were acquired in the UK and Australia. Despite their use of very high-resolution (1.7 cm–2 cm) imagery and a young fine vegetable farm, our model performed quite better per the recorded *mAP*.

Although we used relatively low-resolution (20 cm) images compared to the very high-resolution (5 cm or better) images normally from UAVs and other powerful sensors, our model's results look very promising. This unravels the prowess of the Mask R-CNN architecture in forestry applications, even with relatively low resolution datasets. Additionally, this demonstrates the potential of our model in detecting and segmenting standing dead trees in the dense BFNP forest from aerial imagery. Consequently, our model could be used for automation in standing dead tree detection and segmentation for enhanced forest management. In so doing, it would ease forestry assessments relating to tree mortality, making it more efficient and cost-effective with less manual work even in larger forest areas. Finally, we would like to state that these findings could be useful and very significant for accurate Carbon storage estimation.

In future investigations, we hope to look more into which hyperparameters led to most improvements. It is important to note that the applications of DL, despite its recent achievements and growth with unprecedented accuracies in solving complex problems, in forestry is still young. Moving forward, we would like to recommend more applications of DL in assessing and quantifying forest health, as this would be more sustainable and cost-effective for enhanced forest management. A combination of different remote sensors (e.g., images and point cloud from LiDAR) in DL training is likely to achieve better results. Therefore, this sensor combination should be considered. We would also recommend that the dataset used in such investigations be published in order to reduce the data burden for more research. Especially, due to the fact that one of the challenges for training supervised DL models is dataset acquisition and preparation.

6. Conclusion

In this study, we exhibited a DL automation approach by adjusting the Mask R-CNN to train a new model for detecting and segmenting individual dead tree crowns in a mixed dense forest. With limited training dataset, we applied techniques that enabled our model to reach promising results, exceeding other studies that used relatively higher resolution datasets compared to ours. Our model recorded an overall mean Average Precision (*mAP*), *mR*, and *mF*1 of 0.85, 0.88, and 0.87 respectively. This demonstrates the robustness of our model and a great potential to apply it in forestry management, as it would be cost-effective and sustainable even for larger forest areas. This study is equally significant for forest Carbon storage estimation, having known the living tree and their heights, in the sense that how well a forest can store Carbon is determined by the forest's growth and health coupled with its biodiversity. Therefore, getting to know the quantity of standing dead trees in forests would provide firsthand information to forest managers to facilitate the assessment of the general forest health while carrying out their routine inventory. This kind of information could as well enlighten forest managers to reorient their strategies in keeping the forest in a good state; because the healthier and denser the forest, the more carbon it can store. This is very important for keeping a clean climate and healthy environment. In the future, we intend to investigate the decay levels of standing dead trees using a combination of different remotely sensed data, like imagery and LiDAR data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Project No. 42171361), the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project PolyU 25211819.

References

Abdulla, W., 2017. Mask R-cnn for Object Detection and Instance Segmentation on Keras and Tensorflow.

Arce, L.S.D., Osco, L.P., Arruda, M.d.S.d., Furuya, D.E.G., Ramos, A.P.M., Aoki, C., Pott, A., Fatholahi, S., Li, J., Araújo, F.F.d., et al., 2021. Mauritia flexuosa palm trees airborne mapping with deep convolutional neural network. Sci. Rep. 11, 1–13. https://doi.org/10.21203/rs.3.rs-374682/v1.

Assessment, M.E., 2005. Ecosystems and Human Well-Being: Wetlands and Water. World Resources Institute.

Braga, J.R.G., Peripato, V., Dalagnol, R., Ferreira, M.P., Tarabalka, Y., Aragão, L.E.O.C., de Campos Velho, H.F., Shiguemori, E.H., Wagner, F.H., 2020. Tree crown delineation algorithm based on a convolutional neural network. Rem. Sens. 12, 1288. https://doi.org/10.3390/rs12081288.

Briechle, S., Krzystek, P., Vosselman, G., 2020. Classification of tree species and standing dead trees by fusing uav-based lidar data and multispectral imagery in the 3d deep neural network pointnet++. ISPRS Ann. photogramm. Rem. Sens. Spatial Inf. Sci 2, 203–210. https://doi.org/10.5194/isprs-annals-V-2-2020-203-2020.

Buduma, N., Buduma, N., Papa, J., 2022. Fundamentals of Deep Learning. O'Reilly Media, Inc..".

Cailleret, M., Heurich, M., Bugmann, H., 2014. Reduction in browsing intensity may not compensate climate change effects on tree species composition in the bavarian forest national park. For. Ecol. Manag. 328, 179–192. https://doi.org/10.1016/j.foreco.2014.05.030.

Chiang, C.Y., Barnes, C., Angelov, P., Jiang, R., 2020. Deep learning-based automated forest health diagnosis from aerial images. IEEE Access 8, 144064–144076. https://doi.org/10.1109/access.2020.3012417.

Dai, L., Zhang, Y., Wang, L., Zheng, S., Xu, W., 2021. Assessment of carbon density in natural mountain forest ecosystems at northwest China. Int. J. Environ. Res. Publ. Health 18, 2098. https://doi.org/10.3390/ijerph18042098.

DeCastro, A.L., Juliano, T.W., Kosović, B., Ebrahimian, H., Balch, J.K., 2022. A computationally efficient method for updating fuel inputs for wildfire behavior models using sentinel imagery and random forest classification. Rem. Sens. 14, 1447. https://doi.org/10.3390/rs14061447.

Duque, A., Peña, M.A., Cuesta, F., González-Caro, S., Kennedy, P., Phillips, O.L., Calderón-Loor, M., Blundo, C., Carilla, J., Cayola, L., et al., 2021. Mature andean forests as globally important carbon sinks and future carbon refuges. Nat. Commun. 12, 1–10. https://doi.org/10.1038/s41467-021-22459-8.

Dutta, A., Zisserman, A., 2019. The via Annotation Software for Images, Audio and Video. https://doi.org/10.48550/ARXIV.1904.10699.

Einzmann, K., Atzberger, C., Pinnel, N., Glas, C., Böck, S., Seitz, R., Immitzer, M., 2021. Early detection of spruce vitality loss with hyperspectral data: results of an experimental study in bavaria, Germany. Rem. Sens. Environ. 266, 112676 https://doi.org/10.1016/j.rse.2021.112676.

Fricker, G.A., Ventura, J.D., Wolf, J.A., North, M.P., Davis, F.W., Franklin, J., 2019. A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. Rem. Sens. 11, 2326. https://doi.org/10.3390/rs11192326.

Furuya, D.E.G., Aguiar, J.A.F., Estrabis, N.V., Pinheiro, M.M.F., Furuya, M.T.G., Pereira, D.R., Gonçalves, W.N., Liesenberg, V., Li, J., Marcato Junior, J., et al., 2020. A machine learning approach for mapping forest vegetation in riparian zones in an atlantic biome environment using sentinel-2 imagery. Rem. Sens. 12, 4086. https://doi.org/10.3390/rs12244086.

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M.C., Fröberg, M., Stendahl, J., Philipson, C.D., et al., 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. Nat. Commun. 4, 1–8. https://doi.org/10.1038/ncomms2328.

Garrity, S.R., Allen, C.D., Brumby, S.P., Gangodagamage, C., McDowell, N.G., Cai, D.M., 2013. Quantifying tree mortality in a mixed species woodland using multitemporal high spatial resolution satellite imagery. Rem. Sens. Environ. 129, 54–65. https://doi.org/10.1016/j.rse.2012.10.029.

Gella, G.W., Wendt, L., Lang, S., Tiede, D., Hofer, B., Gao, Y., Braun, A., 2022. Mapping of dwellings in IDP/refugee settlements from very high-resolution satellite imagery using a mask region-based convolutional neural network. Rem. Sens. 14, 689. https://doi.org/10.3390/rs14030689.

Gjertsen, A.K., 2007. Accuracy of forest mapping based on landsat tm data and a knn-based method. Rem. Sens. Environ. 110, 420–430. https://doi.org/10.1016/j.rse.2006.08.018.

Hamdi, Z.M., Brandmeier, M., Straub, C., 2019. Forest damage assessment using deep learning on high resolution remote sensing data. Rem. Sens. 11 https://doi.org/10.3390/rs11171976. URL: https://www.mdpi.com/2072-4292/11/17/1976.

Hamedianfar, A., Mohamedou, C., Kangas, A., Vauhkonen, J., 2022. Deep Learning for Forest Inventory and Planning: a Critical Review on the Remote Sensing Approaches So Far and Prospects for Further Applications. https://doi.org/10.1093/forestry/cpac002.

Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y., Yu, K., Liu, J., 2021. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask r-CNN). ISPRS J. Photogrammetry Remote Sens. 178, 112–123. https://doi.org/10.1016/j.isprsjprs.2021.06.003.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.322.

Heurich, M., Ochs, T., Andresen, T., Schneider, T., 2009. Object-orientated image analysis for the semi-automatic detection of dead trees following a spruce bark beetle (ips typographus) outbreak. Eur. J. For. Res. 129, 313–324. https://doi.org/10.1007/s10342-009-0331-1.

Hulvey, K.B., Hobbs, R.J., Standish, R.J., Lindenmayer, D.B., Lach, L., Perring, M.P., 2013. Benefits of tree mixes in carbon plantings. Nat. Clim. Change 3, 869–874. https://doi.org/10.1038/nclimate1862.

Immitzer, M., Atzberger, C., Koukal, T., 2012. Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. Rem. Sens. 4, 2661–2693. https://doi.org/10.3390/rs4092661.

Jiang, S., Yao, W., Heurich, M., 2019. Dead wood detection based on semantic segmentation of vhr aerial cir imagery using optimized fcn-densenet. The International Archives of the Photogrammetry. Rem. Sens. Spatial Inf. Sci. XLII-2/W16 127–133. https://doi.org/10.5194/isprs-archives-XLII-2-W16-127-2019.

Jordan, J., 2017. Hyperparameter tuning for machine learning models. Online. URL: https://www.jeremyjordan.me/hyperparameter-tuning/.

Jung, A.B., Wada, K., Crall, J., Tanaka, S., Graving, J., Reinders, C., Yadav, S., Banerjee, J., Vecsei, G., Kraft, A., Rui, Z., Borovec, J., Vallentin, C., Zhydenko, S., Pfeiffer, K., Cook, B., Fernández, I., De Rainville, F.M., Weng, C.H., Ayala-Acevedo, A., Meudec, R., Laporte, M., et al., 2020. Imgaug. https://github.com/aleju/imgaug. Online. (Accessed 1 February 2020).

Junttila, S., Näsi, R., Koivumäki, N., Imangholiloo, M., Saarinen, N., Raisio, J., Holopainen, M., Hyyppä, H., Hyyppä, J., Lyytikäinen-Saarenmaa, P., et al., 2022. Multispectral imagery provides benefits for mapping spruce tree decline due to bark beetle infestation when acquired late in the season. Rem. Sens. 14, 909. https://doi.org/10.3390/rs14040909.

Kamińska, A., Lisiewicz, M., Stereńczak, K., Kraszewski, B., Sadkowski, R., 2018. Species-related single dead tree detection using multi-temporal als data and cir imagery. Rem. Sens. Environ. 219, 31–43. https://doi.org/10.1016/j.rse.2018.10.005.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS J. Photogrammetry Remote Sens. 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.

Khan, A., Asim, W., Ulhaq, A., Ghazi, B., Robinson, R.W., 2021. Health assessment of eucalyptus trees using siamese network from google street and ground truth images. Rem. Sens. 13, 2194. https://doi.org/10.3390/rs13112194.

Latifi, H., Fassnacht, F.E., Schumann, B., Dech, S., 2014. Object-based extraction of bark beetle (ips typographus l.) infestations using multi-date landsat and spot satellite

imagery. Prog. Phys. Geogr. 38, 755–785. https://doi.org/10.1177/0309133314550670.
Laurance, W.F., Delamônica, P., Laurance, S.G., Vasconcelos, H.L., Lovejoy, T.E., 2000. Rainforest fragmentation kills big trees. Nature 404. https://doi.org/10.1038/35009032, 836–836.
Lausch, A., Heurich, M., Fahse, L., 2013. Spatio-temporal infestation patterns of ips typographus (l.) in the bavarian forest national park, Germany. Ecol. Indicat. 31, 73–81. https://doi.org/10.1016/j.ecolind.2012.07.026.
Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P., 2014. Microsoft Coco: Common Objects in Context arXiv:1405.0312.
Liu, X., Frey, J., Denter, M., Zielewska-Büttner, K., Still, N., Koch, B., 2021. Mapping standing dead trees in temperate montane forests using a pixel-and object-based image fusion method and stereo worldview-3 imagery. Ecol. Indicat. 133, 108438 https://doi.org/10.1016/j.ecolind.2021.108438.
Lopes Queiroz, G., McDermid, G.J., Castilla, G., Linke, J., Rahman, M.M., 2019. Mapping coarse woody debris with random forest classification of centimetric aerial imagery. Forests 10, 471. https://doi.org/10.3390/f10060471.
Lv, Y., Zhang, C., Yun, W., Gao, L., Wang, H., Ma, J., Li, H., Zhu, D., 2020. The delineation and grading of actual crop production units in modern smallholder areas using RS data and mask r-CNN. Rem. Sens. 12, 1074. https://doi.org/10.3390/rs12071074.
Machefer, M., Lemarchand, F., Bonnefond, V., Hitchins, A., Sidiropoulos, P., 2020. Mask r-CNN refitting strategy for plant counting and sizing in UAV imagery. Rem. Sens. 12, 3015. https://doi.org/10.3390/rs12183015.
Maltamo, M., Räty, J., Korhonen, L., Kotivuori, E., Kukkonen, M., Peltola, H., Kangas, J., Packalen, P., 2020. Prediction of forest canopy fuel parameters in managed boreal forests using multispectral and unispectral airborne laser scanning data and aerial images. Eur. J. Rem. Sens. 53, 245–257. https://doi.org/10.1080/22797254.2020.1816142.
Martin, A.R., Domke, G.M., Doraisami, M., Thomas, S.C., 2021. Carbon fractions in the world's dead wood. Nat. Commun. 12, 1–9. https://doi.org/10.1038/s41467-021-21149-9.
Meiforth, J.J., Buddenbaum, H., Hill, J., Shepherd, J.D., Dymond, J.R., 2020. Stress detection in New Zealand kauri canopies with worldview-2 satellite and lidar data. Rem. Sens. 12, 1906. https://doi.org/10.3390/rs12121906.
Moustakas, A., Guenther, M., Wiegand, K., Mueller, K.H., Ward, D., Meyer, K.M., Jeltsch, F., 2006. Long-term mortality patterns of the deep-rooted acacia erioloba: the middle class shall die. J. Veg. Sci. 17, 473–480. https://doi.org/10.1111/j.1654-1103.2006.tb02468.x.
Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. Journal of big data 2, 1–21. https://doi.org/10.1186/s40537-014-0007-7.
Nielsen, M.A., 2015. Neural Networks and Deep Learning, 25. Determination press, San Francisco, CA, USA.
Nielsen, M.M., Heurich, M., Malmberg, B., Brun, A., 2014. Automatic mapping of standing dead trees after an insect outbreak using the window independent context segmentation method. J. For. https://doi.org/10.5849/jof.13-050.
Nowak, D.J., Crane, D.E., 2002. Carbon storage and sequestration by urban trees in the USA. Environ. Pollut. 116, 381–389. https://doi.org/10.1016/s0269-7491(01)00214-7.
Paniagua-Ramirez, A., Krupinska, O., Jagdeo, V., Cooper, W.J., 2021. Carbon storage estimation in a secondary tropical forest at ciee sustainability center, monteverde, Costa Rica. Sci. Rep. 11, 1–8. https://doi.org/10.1038/s41598-021-03004-5.
Polewski, P., Shelton, J., Yao, W., Heurich, M., 2020. Segmentation of single standing dead trees in high-resolution aerial imagery with generative adversarial network-based shape priors. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. XLIII-B2–2020, 717–723. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-717-2020.
Polewski, P., Shelton, J., Yao, W., Heurich, M., 2021. Instance segmentation of fallen trees in aerial color infrared imagery using active multi-contour evolution with fully convolutional network-based intensity priors. ISPRS J. Photogrammetry Remote Sens. 178, 297–313. https://doi.org/10.1016/j.isprsjprs.2021.06.016.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-Cnn: towards Real-Time Object Detection with Region Proposal Networks. https://doi.org/10.48550/ARXIV.1506.01497.
Seibold, S., Rammer, W., Hothorn, T., Seidl, R., Ulyshen, M.D., Lorz, J., Cadotte, M.W., Lindenmayer, D.B., Adhikari, Y.P., Aragón, R., et al., 2021. The contribution of insects to global forest deadwood decomposition. Nature 597, 77–81. https://doi.org/10.1038/s41586-021-03740-8.
Seidl, R., Schelhaas, M.J., Rammer, W., Verkerk, P.J., 2014. Increasing forest disturbances in europe and their impact on carbon storage. Nat. Clim. Change 4, 806–810. https://doi.org/10.1038/nclimate2318.
Senf, C., Seidl, R., Hostert, P., 2017. Remote sensing of forest insect disturbances: current state and future directions. Int. J. Appl. Earth Obs. Geoinf. 60, 49–60. https://doi.org/10.1016/j.jag.2017.04.004.
Skiljan, I., 2022. Irfan view. Online. URL: https://www.irfanview.com/.
Soloy, A., Turki, I., Fournier, M., Costa, S., Peuziat, B., Lecoq, N., 2020. A deep learning-based method for quantifying and mapping the grain size on pebble beaches. Rem. Sens. 12, 3659. https://doi.org/10.3390/rs12213659.
Stephenson, N.L., Das, A., Condit, R., Russo, S., Baker, P., Beckman, N.G., Coomes, D., Lines, E., Morris, W., Rüger, N., et al., 2014. Rate of tree carbon accumulation increases continuously with tree size. Nature 507, 90–93. https://doi.org/10.1038/nature12914.
Ullo, S., Mohan, A., Sebastianelli, A., Ahamed, S., Kumar, B., Dwivedi, R., Sinha, G.R., 2021. A new mask r-CNN-based method for improved landslide detection. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 14, 3799–3810. https://doi.org/10.1109/jstars.2021.3064981.
Verbesselt, J., Robinson, A., Stone, C., Culvenor, D., 2009. Forecasting tree mortality using change metrics derived from modis satellite data. For. Ecol. Manag. 258, 1166–1173. https://doi.org/10.1016/j.foreco.2009.06.011.
Wang, T., Zhang, K., Zhang, W., Wang, R., Wan, S., Rao, Y., Jiang, Z., Gu, L., 2021. Tea picking point detection and location based on mask-RCNN. Inf. Process. Agric. https://doi.org/10.1016/j.inpa.2021.12.004.
Wu, B., Liang, A., Zhang, H., Zhu, T., Zou, Z., Yang, D., Tang, W., Li, J., Su, J., 2021. Application of conventional uav-based high-throughput object detection to the early diagnosis of pine wilt disease by deep learning. For. Ecol. Manag. 486, 118986 https://doi.org/10.1016/j.foreco.2021.118986.
Yao, W., Krzystek, P., Heurich, M., 2012. Tree species classification and estimation of stem volume and dbh based on single tree extraction by exploiting airborne full-waveform lidar data. Rem. Sens. Environ. 123, 368–380. https://doi.org/10.1016/j.rse.2012.03.027.
Ye, N., Morgenroth, J., Xu, C., Chen, N., 2021. Indigenous forest classification in New Zealand–a comparison of classifiers and sensors. Int. J. Appl. Earth Obs. Geoinf. 102, 102395 https://doi.org/10.1016/j.jag.2021.102395.
Yekeen, S.T., Balogun, A.L., Yusof, K.B.W., 2020. A novel deep learning instance segmentation model for automated marine oil spill detection. ISPRS J. Photogrammetry Remote Sens. 167, 190–200. https://doi.org/10.1016/j.isprsjprs.2020.07.011.
Yu, K., Hao, Z., Post, C.J., Mikhailova, E.A., Lin, L., Zhao, G., Tian, S., Liu, J., 2022. Comparison of classical methods and mask r-CNN for automatic tree detection and mapping using UAV imagery. Rem. Sens. 14, 295. https://doi.org/10.3390/rs14020295.
Yu, R., Luo, Y., Zhou, Q., Zhang, X., Wu, D., Ren, L., 2021. A machine learning algorithm to detect pine wilt disease using uav-based hyperspectral imagery and lidar data at the tree level. Int. J. Appl. Earth Obs. Geoinf. 101, 102363 https://doi.org/10.1016/j.jag.2021.102363.
Zhang, C., Zhou, J., Wang, H., Tan, T., Cui, M., Huang, Z., Wang, P., Zhang, L., 2022. Multi-species individual tree segmentation and identification based on improved mask r-CNN and UAV imagery in mixed forests. Rem. Sens. 14, 874. https://doi.org/10.3390/rs14040874.
Zhang, W., Liljedahl, A.K., Kanevskiy, M., Epstein, H.E., Jones, B.M., Jorgenson, M.T., Kent, K., 2020. Transferability of the deep learning mask r-CNN model for automated mapping of ice-wedge polygons in high-resolution satellite and UAV images. Rem. Sens. 12, 1085. https://doi.org/10.3390/rs12071085.
Zhang, Y., Dian, Y., Zhou, J., Peng, S., Hu, Y., Hu, L., Han, Z., Fang, X., Cui, H., 2021. Characterizing spatial patterns of pine wood nematode outbreaks in subtropical zone in China. Rem. Sens. 13, 4682. https://doi.org/10.3390/rs13224682.
Zhao, W., Persello, C., Stein, A., 2021. Building outline delineation: from aerial images to polygons with an improved end-to-end learning framework. ISPRS J. Photogrammetry Remote Sens. 175, 119–131. https://doi.org/10.1016/j.isprsjprs.2021.02.014.
Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine 5, 8–36. https://doi.org/10.1109/mgrs.2017.2762307.
Zlateski, A., Jaroensri, R., Sharma, P., Durand, F., 2018. On the importance of label quality for semantic segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE. https://doi.org/10.1109/cvpr.2018.00160.
Zomer, R.J., Neufeldt, H., Xu, J., Ahrends, A., Bossio, D., Trabucco, A., Van Noordwijk, M., Wang, M., 2016. Global tree cover and biomass carbon on agricultural land: the contribution of agroforestry to global and national carbon budgets. Sci. Rep. 6, 1–12. https://doi.org/10.1038/srep29987.