

SegmentAnyTree: A sensor and platform agnostic deep learning model for tree segmentation using laser scanning data

Maciej Wielgosz ^{a,*¹}, Stefano Puliti ^{a,1}, Binbin Xiang ^b, Konrad Schindler ^b, Rasmus Astrup ^a

^a Norwegian Institute of Bioeconomy Research (NIBIO), 1433 Ås, Norway

^b Photogrammetry and Remote Sensing, ETH Zürich, 8093 Zürich, Switzerland

ARTICLE INFO

Editor: Jing M. Chen

Keywords:

3D deep learning
Instance segmentation
ITC
ALS
TLS
Drones

ABSTRACT

This study focuses on advancing individual tree crown (ITC) segmentation in lidar data, developing a sensor- and platform-agnostic deep learning model transferable across a spectrum of dense laser scanning datasets from drone (ULS), to terrestrial (TLS), and mobile (MLS) laser scanning data. In a field where transferability across different data characteristics has been a longstanding challenge, this research marks a step towards versatile, efficient, and comprehensive 3D forest scene analysis.

Central to this study is model performance evaluation based on platform type (ULS vs. MLS) and data density. This involved five distinct scenarios, each integrating different combinations of input training data, including ULS, MLS, and their augmented versions through random subsampling, to assess the model's transferability to varying resolutions and efficacy across different canopy layers. The core of the model, inspired by the *PointGroup* architecture, is a 3D convolutional neural network (CNN) with dedicated prediction heads for semantic and instance segmentation. The model underwent comprehensive validation on publicly available, machine learning-ready point cloud datasets. Additional analyses assessed model adaptability to different resolutions and performance across canopy layers.

Our results reveal that point cloud random subsampling is an effective augmentation strategy and improves model performance and transferability. The model trained using the most aggressive augmentation, including point clouds as sparse as 10 points m⁻², showed best performance and was found to be transferable to sparse lidar data and boosts detection and segmentation of codominant and dominated trees. Notably, the model showed consistent performance for point clouds with densities >50 points m⁻² but exhibited a drop in performance at the sparsest level (10 points m⁻²), mainly due to increased omission rates. Benchmarking against current state-of-the-art methods revealed boosts of up to 20% in the detection rates, indicating the model's superior performance on multiple open benchmark datasets. Further, our experiments also set new performance baselines for the other public datasets. The comparison highlights the model's superior segmentation skill, mainly due to better detection and segmentation of understory trees below the canopy, with reduced computational demands compared to other recent methods.

In conclusion, the present study demonstrates that it is indeed feasible to train a sensor-agnostic model that can handle diverse laser scanning data, going beyond current sensor-specific methodologies. Further, our study sets a new baseline for tree segmentation, especially in complex forest structures. By advancing the state-of-the-art in forest lidar analysis, our work also lays the foundation for future innovations in ecological modeling and forest management.

1. Introduction

Obtaining information on individual trees to support smaller-scale and multifunctional forest management has been a central area of

research for the last three decades. Trees are the basic unit based on which forest management is applied. The increased pressure to transition towards more granular, biodiversity-friendly, and multifunctional forest management practices set by policies such as the EU Biodiversity

* Corresponding author.

E-mail address: maciej.wielgosz@nibio.no (M. Wielgosz).

¹ These authors contributed equally to this study

strategy for 2030 (European Commission, 2020) and the New EU Forest strategy (EuropeanCommission 2021) calls for a more granular management thus shifting from stand-level management to individual tree management. In this context, individual tree information represents a cornerstone towards a more sustainable forest management. A wide variety of approaches, often referred to as individual tree crown (ITC) methods, has been proposed to segment individual trees in lidar data. However, the sub-optimal performances of these existing methods have consistently been a major obstacle, limiting advances towards individual tree inventories. We presently witness a renaissance of research into single tree methods, fueled by advancements in lidar technology and deep learning, which have increased access to high-resolution point clouds. Most existing ITC methods are tailored to a specific type of lidar data as input, with distinct approaches and traditions for airborne laser scanning (ALS) or for proximally sensed static terrestrial (TLS), mobile (MLS), or unmanned aerial vehicles (UAV) lidar (ULS). Comparatively, little effort went into developing platform- and sensor-agnostic methods or models that are transferable to any new lidar data without carefully re-tuning various key parameters and settings. Such an agnostic ITC method would offer unprecedented versatility during deployment, as it could be applied to any given lidar data and deliver consistent predictions across datasets. Once available, these models would streamline ITC method selection by eliminating the need for practitioners to sift through numerous options to find the one tailored to their increasingly similar lidar datasets.

Since the inception of ITC approaches for ALS data (Hyppa et al., 2001) the core of the methods has relied on a multi-step process that: i) identifies tree-tops, often referred to as ‘seeds,’ using rasterized canopy height models, ii) segments individual tree crowns using region-based image segmentation techniques starting from each seed. Following this, numerous alternative techniques have been developed with-variations to the original seeding and segmentation methods, e.g. Delponte and Coomes (2016). Additionally, newer methods have emerged that more effectively exploit the 3D capabilities of ALS data (e.g. Ayrey et al., 2017; Ferraz et al., 2012; Li et al., 2012). Amongst the latter, of particular interest are those methods operating at the point cloud level, such as the adaptive mean shift 3D (AMS3D) or the method presented by Li et al. (2012). While both have shown promising results, their operational use has been limited possibly by the need to perform parameter tuning for new datasets or by excessive computational needs. As a result, until recently most practitioners rely on simple 2D CHM methods. Despite the wealth of research done in ITC segmentation over the past 20 years, the state-of-the-art (SOTA) for tree detection and segmentation algorithms has improved only marginally. An evaluation of two decennially spaced ITC benchmark efforts—one by Kaartinen et al. (2012) and the other by Cao et al. (2023)—indicates only slight enhancements in metrics such as detection rates or F1-scores. Both studies highlighted the low detection rates (52% on average Kaartinen et al. (2012)), driven by the poor detection rates of co-dominant trees (i.e. trees in the dominant canopy but very close to each other) and the severe under-detection of suppressed or understory trees (i.e., trees entirely covered by the main canopy and not visible from above). The relatively large omission rates for smaller trees have been shown to produce biased stand volume estimates, thus resulting in the scarce uptake of ITC methods in operational forest planning. Further, it is important to highlight that most existing ITC segmentation methods rely on hyperparameters (e.g. window search size, crown dimensions). These generally relate to prior information on the tree spatial and size distribution, which is unknown when applying ITC methods over new data. Such dependency limits the transferability of ITC methods to new data, and while attempts have been made to estimate hyperparameters dynamically based on ALS data properties (e.g., Popescu et al. (2002)), empirical hyperparameter fine-tuning often remains a necessary step for new datasets, limiting transferability.

In parallel to the development of ALS ITC methods, we have, in the past ten years, also witnessed the development of methods to segment

individual trees in very detailed 3D point clouds (e.g., Tao et al. (2015) and Burt et al. (2019)). According to the benchmarking effort by Liang et al. (2018), TLS-based ITC segmentation generally outperforms ALS-based ones. Such result is likely due to the higher resolution of TLS data, which allows for more complex segmentation routines that leverage the fine detailed information on tree stems and crowns. Even so, crown segmentation results are often manually edited, making the ITC automation a main bottleneck to unlock the wealth of information available in proximally sensed lidar data either from the ground (TLS and MLS) or from drones (ULS).

In the past couple of years, driven by the advent of deep learning, convolutional neural networks (CNNs) and by the availability of open point cloud benchmark datasets (e.g. Puliti et al., 2023a; Weiser et al., 2022), significant progress has been made in ITC segmentation. Most recent research into lidar deep learning-based ITC segmentation employed very dense TLS, MLS, or ULS point clouds. The level of detail of such point clouds allows for a clear visual distinction between individual trees, enabling manual annotation of suitable training and evaluation data. With sufficient training data, these fully data-driven models, boost the automation by eliminating the need for hyper-parameter tuning for new datasets. Moreover, these models exhibit impressive transferability, as seen in the work by Wilkes et al. (2023), who successfully applied a semantic segmentation model developed by (Krisanski et al., 2021) using a small set of plots in an Australian forest, to a wide variety of forest types worldwide. Initially, methods like TLS2Trees (Wilkes et al., 2023), approached deep learning by leveraging its outputs in conventional TLS tree segmentation routines. While advancing terrestrial ITC segmentation, these methods are often custom-tailored to the characteristics of ground based lidar data, such as very high point density on stems and low vegetation (e.g. Wilkes et al., 2023). As such, they are, by design, not directly transferrable to airborne laser scanning data collected above the canopy.

Fewer examples exist in the literature for deep learning-based ITC from ALS data, primarily using very dense ALS or ULS data. Windrim and Bryson (2020) showed the first promising results of a deep learning model for ITC segmentation. Straker et al. (2023), proposed the use of the YOLOv5 model for ITC crown segmentation and found that, on the FOR-instance data (Puliti et al., 2023a), it outperformed the commonly used Voronoi segmentation (detection rate = 30.7%), while also not requiring any prior knowledge on the spatial distribution of trees. Xiang et al. (2024) proposed a more advanced 3D point cloud deep learning model that performs full panoptic (i.e., semantic and instance) segmentation of ULS forest scenes. That model outperformed the one of Straker et al. (2023) by nearly 13% points (detection rate 82.3% vs. 69.6%) and currently constitutes the SOTA on the FOR-instance data. The improvements are, however, limited to very dense ULS data (>1000 points m^{-2}). To broaden the scope and operational impact of ITC inventories, it is necessary to ensure applicability to sparser airborne data (<1000 points m^{-2}).

One of the main challenges when using deep learning on sparser point cloud data is the difficulty of generating suitable annotated training and validation point clouds since it is more complicated for human operators to separate individual trees in ALS point clouds than TLS and MLS data. One possible solution to obtain ALS-like annotated data is to approximately simulate it by down-sampling labelled dense lidar data for which labels do exists. Such point cloud augmentation, is a rather simple strategy that has been previously found to be useful in forest proximal sensing to make the models more robust against canopy occlusions and missing points (Krisanski et al., 2021; Xiang et al., 2024). Data augmentation is widely recognized (Shorten and Khoshgoftaar, 2019) as a valuable technique for enhancing the performance of CNN models by enhancing their robustness and transferability, particularly in scenarios where training data are not abundant, such as labelled 3D forest point clouds.

The primary objective of our study was to evaluate the possibility to leverage the augmentation of dense lidar data through random

subsampling, to train a sensor- and platform-agnostic model that can be applied to dense lidar data (i.e. TLS, ULS or MLS) as well as to sparser ALS. The secondary objective was to identify the factors affecting the segmentation performance and quantify their impact, including: the type of data (ULS or MLS) and degree of augmentation, canopy layering, and computational efficiency.

2. Materials

The study was conducted on a diverse collection of in-house and publicly available forest laser scanning scenes where unique tree identifiers have been manually annotated. The following sections describe in detail the used datasets. For the sake of terminological consistency, we introduced the definition of lidar data types in Table 1.

2.1. Drone laser scanning data

The openly available FOR-instance benchmark dataset (Puliti et al., 2023b) was used as the source of ULS data. That benchmark comprises fully annotated ULS point clouds (from Riegl VUX-1 and mini-VUX) collected over five sites in Norway, Czech Republic, Austria, Australia, and New Zealand and covering mature forests for following forest types: boreal (42% of the number of trees) and temperate coniferous (5%), temperate mixed deciduous (29%), coniferous plantation (13%), and dry sclerophyll forests (11%). The dataset prescribes a fixed split into development data for training and model validation (70% of the area) and test data (30% of the area) to enable meaningful comparisons. In total 796 trees were available for training (average tree density: 451 trees ha^{-1} and the mean tree height: 20.5 m). The point clouds included leaf-off and leaf-on data and had an average density of 6964 points m^{-2} . For each point, unique tree identifiers and semantic labels were manually labelled. In the context of this study, the semantic labels were flattened to a binary classification between tree (stems, branches, and leaves) and non-tree (ground, low vegetation) points.

2.2. Mobile laser scanning data

As ground-based laser scans, we used the MLS data from Wielgosz et al. (2023), consisting of point clouds collected with a GeoSLAM ZEB-HORIZON (GeoSLAM, 2020) at 16 circular plots (400 m^2) all collected in managed boreal coniferous forests settings. In total 649 trees (average density: 1567 trees ha^{-1} and mean tree height: 12.9 m) were available for training. As for the FOR-instance data, the point clouds (phenology: leaf-on; average point density: 18236 points m^{-2}) were manually labelled with instance IDs and semantic labels, which we again flatten

Table 1
Lidar data types.

Type	Description	Data Collection Platform	Typical Point Density (points m^{-2})
ALS	Standard airborne laser scanning	Manned aircraft	1–60
ALS-HD	High-density airborne laser scanning	Manned aircraft or helicopter	60–1000
MLS	Mobile laser scanning, collected from moving vehicles or people walking	Ground vehicles or walking of a man	500–30,000 Up to 50,000, depending on scanner
TLS	Terrestrial or static laser scanning	Tripod (static)	
	Drone-based laser scanning, also known as UAV Laser Scanning. UAV stands for Unmanned Aerial Vehicle.	Unmanned aircraft or copter (drone)	100–10,000
ULS			

into a binary tree (stems and crowns) vs. non-tree (ground, coarse woody debris, low vegetation) classification. To enable a direct comparison against the instance segmentation method proposed in Wielgosz et al. (2023), we use the same data split, with a randomly selected 25% of the area in each plot set aside for testing.

2.3. Augmented data

To obtain data from the same sites and recording times but with characteristics similar to ALS data, we synthetically augment the ULS and MLS point clouds. Given the complexity and computational needs to simulate long-range lidar-based only on short-range lidar points, similarly to Straker et al. (2023) we took the most straightforward approach and randomly subsampled the point clouds to point densities of 1000, 500, 100 and 10 points m^{-2} . Such strategy performs a uniform subsampling through the canopy vertical profile, thus aiming at replicating ALS-HD data with overall point densities that can be obtained under current airborne capture scenarios from low altitude helicopter surveys ALS-HD as well as ALS data.

It is nevertheless important to note that due to this simple approach this study does not account for differences in sensor (e.g. pulse frequency, beam divergence, number of returns, scan angle), acquisition (i.e. flight altitude, speed, flightline overlap) and forest characteristics (e.g. species, leaf-on vs. leaf-off, canopy penetration) that affect the size of the laser footprint and the occlusion and thus the vertical distribution of the points throughout the canopy. While methods to simulate some of these parameters exist (e.g. HELIOS++; Winiwarter et al., 2022), creating virtual scenes encompassing the above-mentioned sources of data variation for all of the training data would be computationally too intensive and was thus performed only for a portion of the test data. Fig. 1 illustrates the input data sources as well as the augmented versions.

2.4. Test data

To assess the methods proposed, we compiled a comprehensive test dataset that includes the earlier mentioned test portions of ULS data (Puliti et al., 2023b), MLS data, as well as additional publicly available, MLS and TLS annotated datasets (Table 2). Overall, 2182 trees were used for evaluation purposes.

To evaluate the model's performance with ALS and ALS-HD data we randomly subsampled the FOR-instance point clouds (see Section 2.4). We then assessed the effect of point cloud density on the models' performance metrics. While our method efficiently explores a wide range of densities, it relies on a basic simulation, which might not yield realistic results. Therefore, we compared its accuracy against the one obtained using advanced ray-tracing based point cloud simulator Helios++ (Winiwarter et al., 2022) that considers varying scanning and flight parameters. We conducted this comparison using an ALS-HD dataset that closely emulates a real data collection (about 1000 points per square meter) from a Riegl VQ1560II-S ALS scanner, flown at 1000 m above ground and with 70% lateral overlap. The initial 3D scenes for Helios++ simulations were based on the FOR-instance data, utilizing a voxel size of 3 cm, the smallest size permitted by the available 128 GB of RAM.

3. Methods

In this study, to assess the effectiveness of developing sensor- and platform-agnostic models, we adopted an approach whereby the model form and hyperparameters (e.g. layers structure, learning rate, batch size) were kept constant while we modified the input training data. We trained several models using either ULS or MLS data and different combinations of these and their respective augmented versions. The models were benchmarked against a comprehensive selection of open datasets. Fig. 2 provides an overview of the adopted workflow.

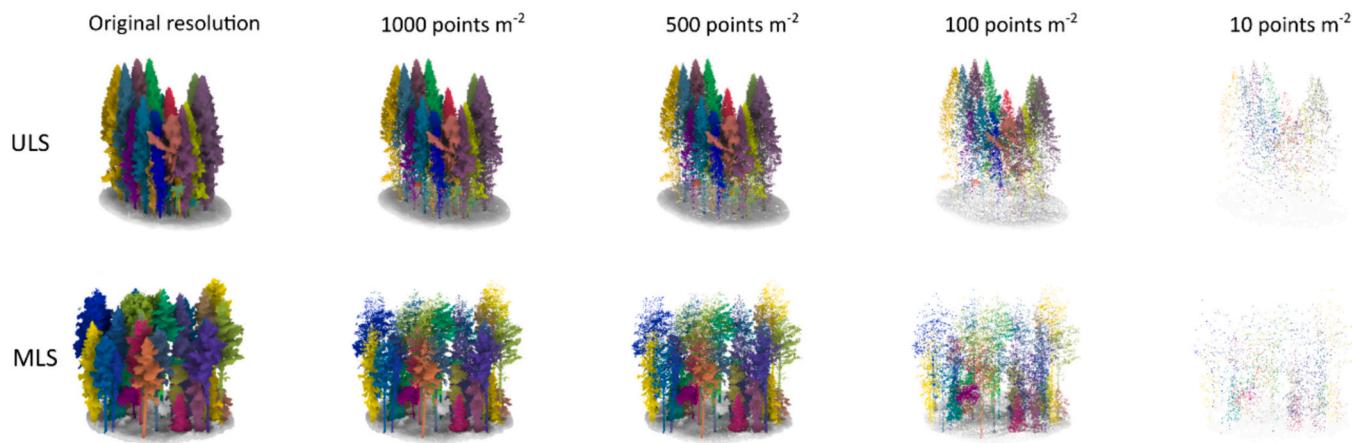


Fig. 1. Example of one plot from the MLS and one from the ULS data, including the augmented version of each dataset. The different colored trees represent the manually labelled individual trees.

Table 2

Summary of the main characteristics for the test datasets used in this study.

Dataset	Country	Forest type and main tree species	n trees and tree density	Mean tree height (m)	Data Type (sensor)	Average point density (pts m^{-2})	Phenology
Wytham woods (Calders et al., 2022)	UK	Temperate mixed deciduous forest (<i>Fraxinus excelsior</i> , <i>Acer Pseudoplatanus</i> , and <i>Corylus avellana</i>)	835; 505 trees ha^{-1}	14.4	TLS (RIEGL VZ-400)	9378	Leaf-off
TreeLearn (Henrich et al., 2023)	Germany	Temperate deciduous forest (<i>Fagus sylvatica</i>)	235; 194 trees ha^{-1}	28.5	MLS (GeoSLAM ZEB-HORIZON)	1224	Leaf-off
LAUTx (Tockner et al., 2022)	Austria	Temperate broadleaved, coniferous, and mixed forests	521; 554 trees ha^{-1}	20.3	MLS (GeoSLAM ZEB-HORIZON)	30,000	Leaf-on
NIBIO MLS test data (Wielgosz et al., 2023)	Norway	Boreal coniferous forest (<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Betula pendula</i>)	258; 1467 trees ha^{-1}	14.7	MLS (GeoSLAM ZEB-HORIZON)	20,000	Leaf-on
FOR-instance test data (Puliti et al., 2023b)	Norway, Czech Republic, Austria, New Zealand, Australia	Boreal coniferous forest (<i>Picea abies</i>), temperate coniferous forests (<i>Pinus sylvestris</i>), temperate mixed deciduous forest (<i>Fagus sylvatica</i> , <i>Acer platanoides</i>), plantation forest (<i>Pinus radiata</i>), dry sclerophyll forest (<i>Eucalyptus</i> sp.)	334; 425 trees ha^{-1}	21.6	ULS (Riegl VUX1 UAV and miniVUX)	6747	Leaf-off and Leaf-on

3.1. Testing different input data and augmentation scenarios

To assess the performance of the different models based on the input of different point cloud data collected through different scanning methods and at different resolutions, we evaluated the following two aspects:

- **Platform:** here, we compared the performance of the model trained on only ULS data against the one trained on MLS data alone. This allows to better understand the versatility of models trained on either of these data sources regarding their transferability to a broad variety of laser scanning datasets. Further, this provides an understanding of the relative contributions of the different raw data sources to the creation of fully agnostic models.
- **Density:** here, we compared the effect of including augmented versions of the data in the training of the model as a form of augmentation with the aim to extend the model's range of transferability to ALS-HD or even to more commonly available ALS data with densities of 10 points m^{-2} .

Based on the above sources of data variation, we tested five different scenarios that differed regarding the input data for model training, including the following combination of ULS, MLS data, and their

augmented versions:

- **Scenario 1:** Only ULS data (i.e. FOR-instance training data split)
- **Scenario 2:** Only MLS data
- **Scenario 3:** The combination of scenarios 1 and 2 was used as a baseline as a consistent dataset to understand the role of the data augmentation strategy on the model's performance.
- **Scenario 4:** The combination of scenarios 1 and 2, plus their augmented version at 1000 points m^{-2} provided an intermediate scenario between the lack of augmentation (scenario 3) and the fully augmented model (scenario 5).
- **Scenario 5:** The combination of scenarios 1 and 2, plus their augmented versions at 1000, 500, 100, 75, 50, 25, and 10 points m^{-2} , represents the augmentation extreme where the training data was thinned to the level of ALS and ALS-HD data.

3.2. Comparison of predictions at different resolutions

To evaluate the models' ability to transfer to ALS and ALS-HD data, we compared the performance of the best-performing scenario on the FOR-instance dataset and the respective augmented versions at all resolution steps. Through such analysis, we aimed to understand the extent to which the model can be transferred to airborne data with decreasing

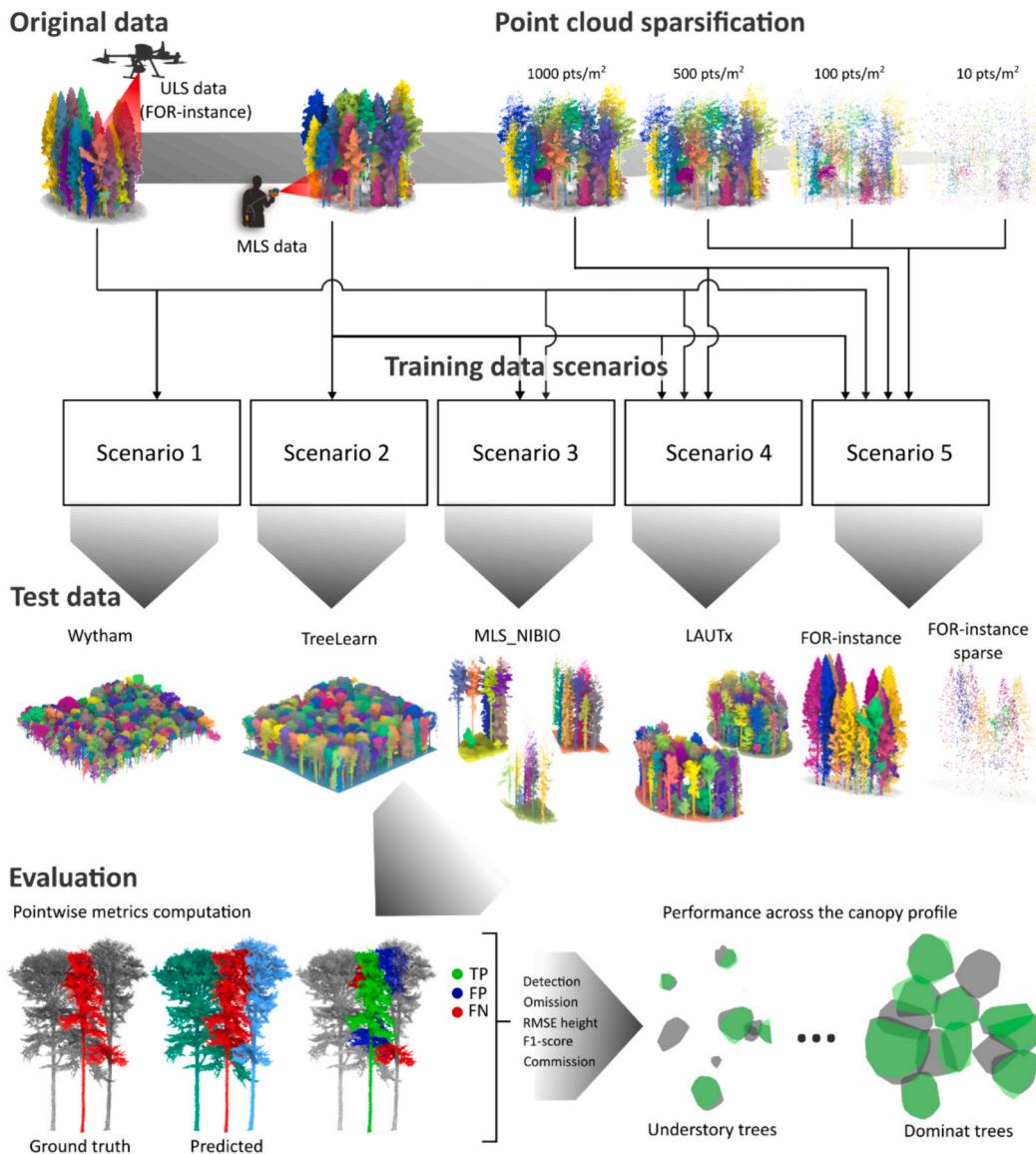


Fig. 2. Schematic visualization of the implemented workflow for the training and evaluation of the proposed models.

point densities without compromising performance. This analysis was limited to these data because the same analysis would not be meaningful for terrestrial datasets, which are by nature very dense and rarely come in the form of sparse data.

Further, to check the validity of our results we tested whether the simple random subsampling of a point cloud can provide a suitable and simple solution to simulate test data for evaluation of ITC segmentation at varying resolution. The test was done by comparing our results on the FOR-instance point clouds augmented at $1000 \text{ points m}^{-2}$ against the Helios++ simulated data with similar point density (see section 2.3).

3.3. Performance across the canopy profile

We evaluated the best model's performance across the test dataset for different layers of the forest canopy. Given the lack of information for each tree on the canopy social status, similarly to the recent benchmark study by Cao et al. (2023) and Xiang et al. (2024), we used tree height to define the layers each tree belonged to. For this purpose, we subdivided each test dataset into 5 m high vertical height bins and computed the metrics for each bin. Further, we visually assessed these results to better understand the relationship between the selected evaluation metrics and

the model's predictions for the different layers.

3.4. Model

We adopted the model architecture and implementation proposed by Xiang et al. (2023), a panoptic point cloud segmentation network leveraging a 3D CNN as its core and enhanced by three parallel prediction heads (see Fig. 3). Our network relies on the Minkowski Engine to balance performance and computational efficiency. The model training pipeline begins with a feature extraction process utilizing a 3D U-Net architecture. This backbone network extracts feature vectors from the voxelized point cloud data.

To ensure that these features are transmitted without loss through the network, our model employs several key strategies. Skip connections, inspired by the U-Net architecture, play a crucial role by transmitting feature maps from earlier to deeper layers, preserving both high-level and low-level feature integrity throughout the network. Furthermore, our architecture incorporates multi-scale feature aggregation, which maintains spatial hierarchies and contextual information essential for accurate segmentation.

From the feature extractor, the process splits into three pathways:

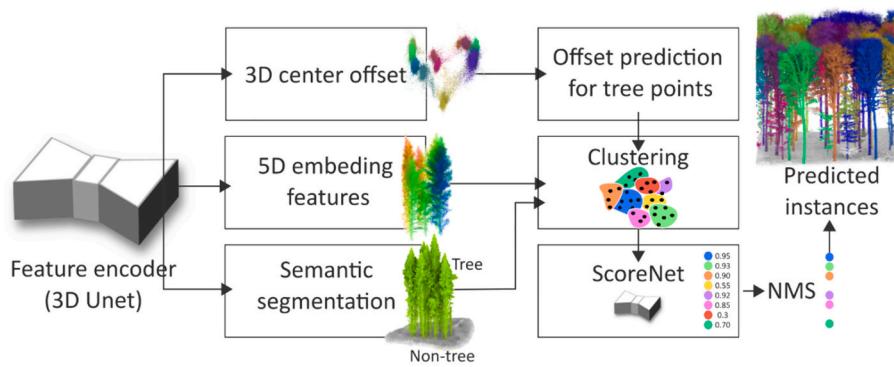


Fig. 3. Schematic representation of the model architecture.

one for semantic segmentation (tree or non-tree), and two dedicated to generating the following embeddings for instance segmentation. The semantic segmentation branch employs a multi-layer perception (MLP) to calculate class probabilities. For the instance segmentation, one branch predicts a 3D offset vector for centering trees, while another branch maps the points into a 5-dimensional embedding space useful for identifying distinct trees. The former, consists of vectors directed from the points to the center of their respective tree instances and that help in clustering the points belonging to the same tree. Further, points belonging to the same tree instance are expected to cluster together in the 5D embedding space. The instance embeddings are then clustered into individual tree candidates using region-growing and mean-shift methods, respectively, and further refined by ScoreNet, a neural network that filters and merges tree candidates based on their intersection over union (IoU) with ground truth data. The model is trained in an end-to-end fashion, utilizing a combined loss function comprising semantic, direction, regression, and score losses. During inference, non-maximum suppression (NMS) is performed on the clusters with the scores predicted by ScoreNet, pruning the candidate list and outputting the final instance predictions (Jiang et al., 2020).

In the model's training process, we utilized several additional data augmentations. Noise with a sigma of 0.01 m was used to adds slight perturbations to the data points, mimicking sensor inaccuracies or environmental disturbances typical in 3D scanning, enhancing the model's resilience against minor data inconsistencies and less reliant on very high-quality data. Further, we relied on random scale anisotropic (Scales: [0.9, 1.1]) and random symmetry augmentations to scale objects within a 90% to 110% range and apply asymmetrical reflection across specific axes, teaching the model to recognize objects amidst size fluctuations and various symmetrical orientations. As previously mentioned, such model hyperparameters were kept constant throughout the tested scenarios.

During the model's training, we implemented several other standard point cloud data augmentations to enhance robustness. We added noise with a sigma of 0.01 m to simulate sensor inaccuracies or environmental disturbances, making the model less dependent on high-quality data. Additionally, we used random scale anisotropic augmentations (scales between 0.9 and 1.1) and random symmetry transformations to teach the model to recognize objects despite size variations and different symmetrical orientations. These hyperparameters remained constant across all testing scenarios.

Model training was done on the NIBIO HPC using a virtual machine with an Intel® Xeon® Gold 6246R CPU and equipped with a NVIDIA GRID V100S-16Q GPU.

3.5. Evaluation metrics

All the above analyses and comparisons were evaluated against the test datasets based on a point-wise matching of ground truth and predicted tree instance identifiers (see Fig. 2). Given the total number of

predicted trees (PT) and ground truth trees (GT), we matched tree instances based on the point-wise intersection over union (IoU). According to this approach, a tree was considered a true positive (TP) if it had an $\text{IoU} > 0.5$ with a ground truth instance. Based on this principle we constructed the confusion matrix and obtained the counts for the TP, false positives (FP), and false negatives (FN) required to compute a selection of commonly used metrics for the evaluation of tree instance segmentation, which included:

$$\text{Detection rate} = \frac{\text{TP}}{\text{GT}} \quad (1)$$

$$\text{Omission rate} = \frac{\text{FN}}{\text{GT}} \quad (2)$$

$$\text{Commission rate} = \frac{\text{FP}}{\text{PT}} \quad (3)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

In addition, we included the root mean square error (RMSE) for the tree height (H ; m above ground), computed as the difference between the highest and lowest point for each tree instance. The inclusion of the tree height RMSE is motivated by the fact that it represents a complementary measure to the detection rates and F1-score, that captures the ability of the model to segment the whole length of the tree. The RMSE was computed according to:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (H_{\text{GT},i} - H_{\text{Pred},i})^2} \quad (7)$$

where N is the total number of trees for which the metric is calculated, $H_{\text{GT},i}$ is the height of the i -th tree in the ground truth, and $H_{\text{Pred},i}$ is the predicted height of the i -th tree.

Further we evaluated the computational efficiency (CE) as a consistent measure of the resources required for inference, useful to compare the processing efficiencies of various methods under varying compute configurations. For this purpose, we proposed a metric describing the data processed per core per time unit or the efficiency with which a system processes data. To compute this metric, three essential elements were considered: the data size processed, measured in megabytes (MB); the number of cores (N_{cores}) which indicates the computational resources employed; and the processing time (T) in

minutes. The CE was computed as:

$$CE = \frac{MB}{N_{cores} * Time} \quad (8)$$

4. Results and discussion

4.1. Influence of data input selection and augmentation on instance segmentation performance

4.1.1. Comparison between airborne and terrestrial data as source for model training

Our initial analysis compared scenarios 1 and 2 to gauge the impact of ULS versus MLS on platform-agnostic models. The results (see Fig. 4 and Table 3) showed that the ULS model generally outperformed the MLS model, although the differences were slight across most datasets, with a notable exception being the TUWIEN dataset. Here, the ULS model boosted detection rates and reduced commission errors by about 20% compared to MLS. This difference is partly attributed to the complex, unmanaged mixed deciduous forest in the TUWIEN dataset, in contrast to the simpler, managed boreal forests used in MLS training data.

Interestingly, the transferability of the ULS model on the MLS datasets (TreeLearn, NIBIO_MLS, and LAUTx) and of the MLS model on the FOR-instance data suggests that both terrestrial and airborne lidar data can be used to train models transferable beyond their platform-domains. This is significant as it shows the models' capability to identify general tree characteristics thus broadening the use and scope of any available dense lidar data. Overall, these findings indicate that the selected model architecture represents a more robust tree segmentation approach compared to platform-specific methods such as TLS2trees (Wilkes et al., 2023).

4.1.2. Comparison of augmentation strategies

To address the effect of the augmentation strategy on the performance of the instance segmentation across the available spectrum of point cloud densities, we conducted a comparative analysis of scenarios 1 to 5, each representing a unique augmentation strategy of increasing intensity.

Our analysis (see Table 3) showed that scenarios 4 and 5, resulted consistently in larger detection rates (i.e., on average 0.76 and 0.77) compared to scenarios 1–3 (on average 0.70, 0.67, and 0.72). This indicates that point cloud random sub-sampling positively impacts the detection and segmentation performance. Scenario 5, which represented the most aggressive augmentation with point clouds as sparse as 10 pts. m⁻², stood out by improving detection rates across various datasets, including increases for airborne ULS datasets like SCION (6% points) and TUWIEN (14.3% points), as well as dense terrestrial datasets like Wytham Woods (10% points), LAUTx (5% points), and NIBIO_MLS (6% points). Overall, we found that training with abstracted point cloud representations enhanced the model's capability to effectively segment trees, especially in complex forest structures like those in Wytham Woods and TUWIEN. This suggests that including sparser data representations can increase the model's robustness and adaptability to complex 3D forest scenes.

Our experiments revealed the value of simple augmentation strategies such as random subsampling resulted in enhancing the generalization of our models. By exposing the model to varying sparsity levels, it was better equipped to generate effective features for datasets with less immediate and more dispersed point relationships. Despite the potential benefits of employing convolutional kernels with larger aspect ratios to optimize performance by better capturing nuanced spatial relationships, we chose not to modify the core structure of the model during this study. This decision ensured that any observed improvements could be

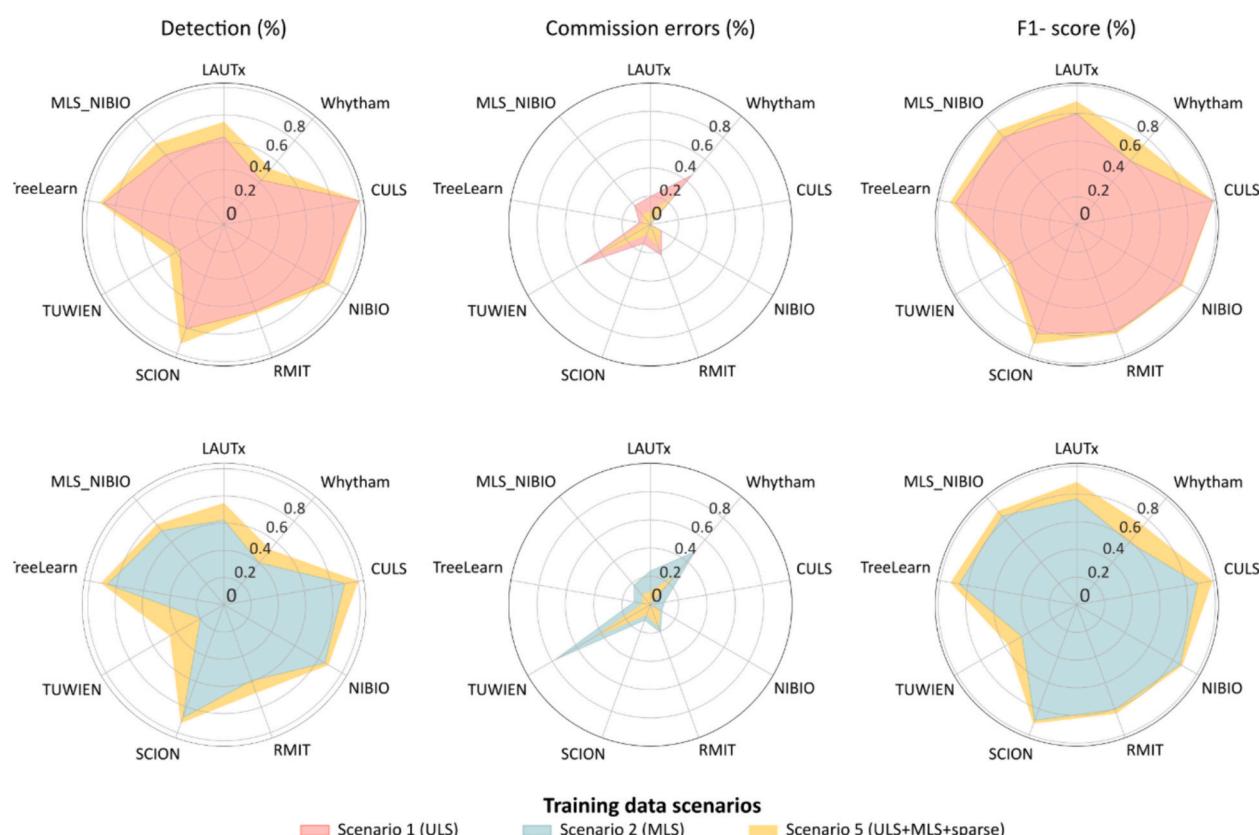


Fig. 4. Radar charts comparing the performance of the model trained on using all data against the models using only the ULS (upper row) or the MLS data (lower row), for comparison we also included the performance of scenario 5 in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Summary of the performance metrics across the different studied scenarios and for each of the used test datasets. The bold numbers indicate the best results for each of the metrics and across the tested augmentation scenarios and datasets.

Scenario	Test dataset	Detection (%)	Omission (%)	Commission (%)	RMSE H (m)	F1-score (%)
1 (ULS)	Wytham woods	0.42	0.58	0.46	5.11	0.60
	TreeLearn	0.89	0.11	0.08	6.31	0.89
	LAUTx	0.64	0.36	0.20	4.90	0.80
	MLS_NIBIO	0.66	0.34	0.17	3.34	0.82
	FOR-instance CULS	1.00	0.00	0.00	0.15	0.99
	FOR-instance NIBIO	0.84	0.16	0.09	3.37	0.87
	FOR-instance TUWIEN	0.37	0.63	0.57	4.43	0.54
	FOR-instance SCION	0.81	0.19	0.14	4.95	0.84
	FOR-instance RMIT	0.67	0.33	0.23	1.84	0.82
	Wytham woods	0.40	0.60	0.49	6.09	0.59
2 (MLS)	TreeLearn	0.87	0.13	0.12	7.16	0.86
	LAUTx	0.62	0.38	0.24	5.99	0.77
	MLS_NIBIO	0.71	0.30	0.18	3.02	0.84
	FOR-instance CULS	0.90	0.10	0.10	4.94	0.89
	FOR-instance NIBIO	0.86	0.15	0.10	3.77	0.86
	FOR-instance TUWIEN	0.20	0.80	0.78	4.89	0.46
	FOR-instance SCION	0.88	0.12	0.12	2.40	0.89
	FOR-instance RMIT	0.59	0.41	0.21	1.74	0.81
	Wytham woods	0.43	0.57	0.43	4.81	0.62
	TreeLearn	0.86	0.14	0.09	5.99	0.87
3 (ULS + MLS)	LAUTx	0.70	0.30	0.16	4.92	0.82
	MLS_NIBIO	0.71	0.30	0.17	3.63	0.82
	FOR-instance CULS	1.00	0.00	0.00	3.12	1.00
	FOR-instance NIBIO	0.89	0.11	0.07	2.75	0.89
	FOR-instance TUWIEN	0.31	0.69	0.65	4.58	0.51
	FOR-instance SCION	0.86	0.14	0.11	3.28	0.88
	FOR-instance RMIT	0.75	0.25	0.11	1.42	0.87
	Wytham woods	0.48	0.52	0.34	3.88	0.67
	TreeLearn	0.93	0.07	0.04	5.14	0.92
	LAUTx	0.72	0.28	0.15	4.13	0.85
4 (ULS + MLS + sparse 1000)	MLS_NIBIO	0.75	0.26	0.12	3.43	0.86
	FOR-instance CULS	1.00	0.00	0.00	0.15	1.00
	FOR-instance NIBIO	0.91	0.09	0.07	3.47	0.90
	FOR-instance TUWIEN	0.43	0.57	0.48	3.05	0.58
	FOR-instance SCION	0.90	0.10	0.10	2.89	0.93
	FOR-instance RMIT	0.70	0.30	0.17	1.37	0.82
	Wytham woods	0.53	0.47	0.27	4.19	0.75
	TreeLearn	0.92	0.09	0.05	0.09	0.92
	LAUTx	0.75	0.25	0.10	3.11	0.89
	MLS_NIBIO	0.77	0.24	0.09	3.44	0.88
5 (ULS + MLS + sparse 1000, 500, 100, 10)	FOR-instance CULS	1.00	0.00	0.00	0.15	1.00
	FOR-instance NIBIO	0.88	0.12	0.08	3.41	0.88
	FOR-instance TUWIEN	0.46	0.54	0.45	4.87	0.57
	FOR-instance SCION	0.92	0.08	0.08	1.83	0.91
	FOR-instance RMIT	0.69	0.31	0.17	1.29	0.84

attributed solely to data augmentation techniques rather than changes in the model's architecture, preserving the integrity of our comparative analysis.

In line with findings from Xiang et al. (2024), who explored a wider and more complex range of augmentation strategies, our simpler approach also yielded significant improvements, with an overall 5.3% increase in performance from scenario 1 to scenario 5. This compares favorably to the 4% improvement noted by Xiang et al. (2024) between their least and most complex scenarios. These results highlight the importance of data augmentation, particularly where comprehensive labelled training datasets are scarce. While advanced augmentation methods like TreeMix (Xiang et al., 2024) or Helios++ simulation software demand more computational resources, our study suggests that elementary augmentation strategies alone can significantly enhance performance. Future research should thus investigate the optimal combination of augmentation techniques to determine an optimal strategy that balances method effectiveness with computational efficiency.

4.2. Performance for different point cloud resolutions

Analyzing the performance scenario 5 model on the FOR-instance point clouds with varying densities (see Fig. 5 and Table A.1 in Appendix 1), the model maintained a stable performance in point clouds

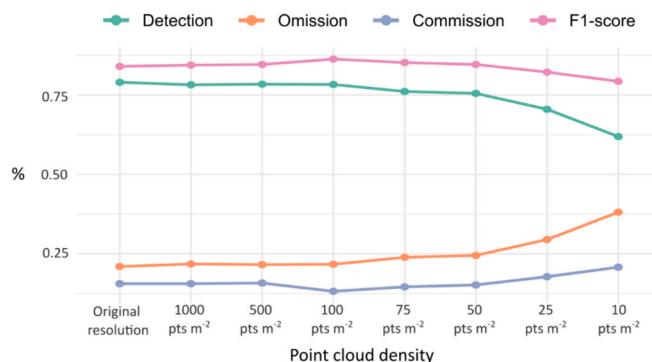


Fig. 5. Average performance across the FOR-instance dataset according to the evaluation metrics and across the increasing augmentation steps tested in this study.

with point densities >50 points m^{-2} . However, the performance dropped for point clouds of 10 points m^{-2} , mainly due to an increase in omission rates (i.e. approximately 20% points). Comparison with metrics obtained using 1000 point m^{-2} Helios++ simulated data (i.e.

detection rate = 76.5%; F1-score = 85.1%) showed that our results (detection rate = 78.3%; F1-score = 84.5%) were of the same order of magnitude and thus as realistic as using complex point cloud simulators.

The results indicate that our method is particularly effective for ALS-HD data captured from low-flying aircrafts, or helicopters with point densities as low as 50 points m^{-2} . While still limited in operational

settings, these data are increasingly being used by researchers and the forest industry in Nordic countries (e.g. Hakula et al., 2023; Hyppä et al., 2022; Persson et al., 2022). To provide a visual understanding of the performance of our model on real data, Fig. A.1 (see Appendix 1) shows the output segmented tree instances on ALS-HD data captured either from a helicopter (920 points m^{-2}), manned aircraft (665 points

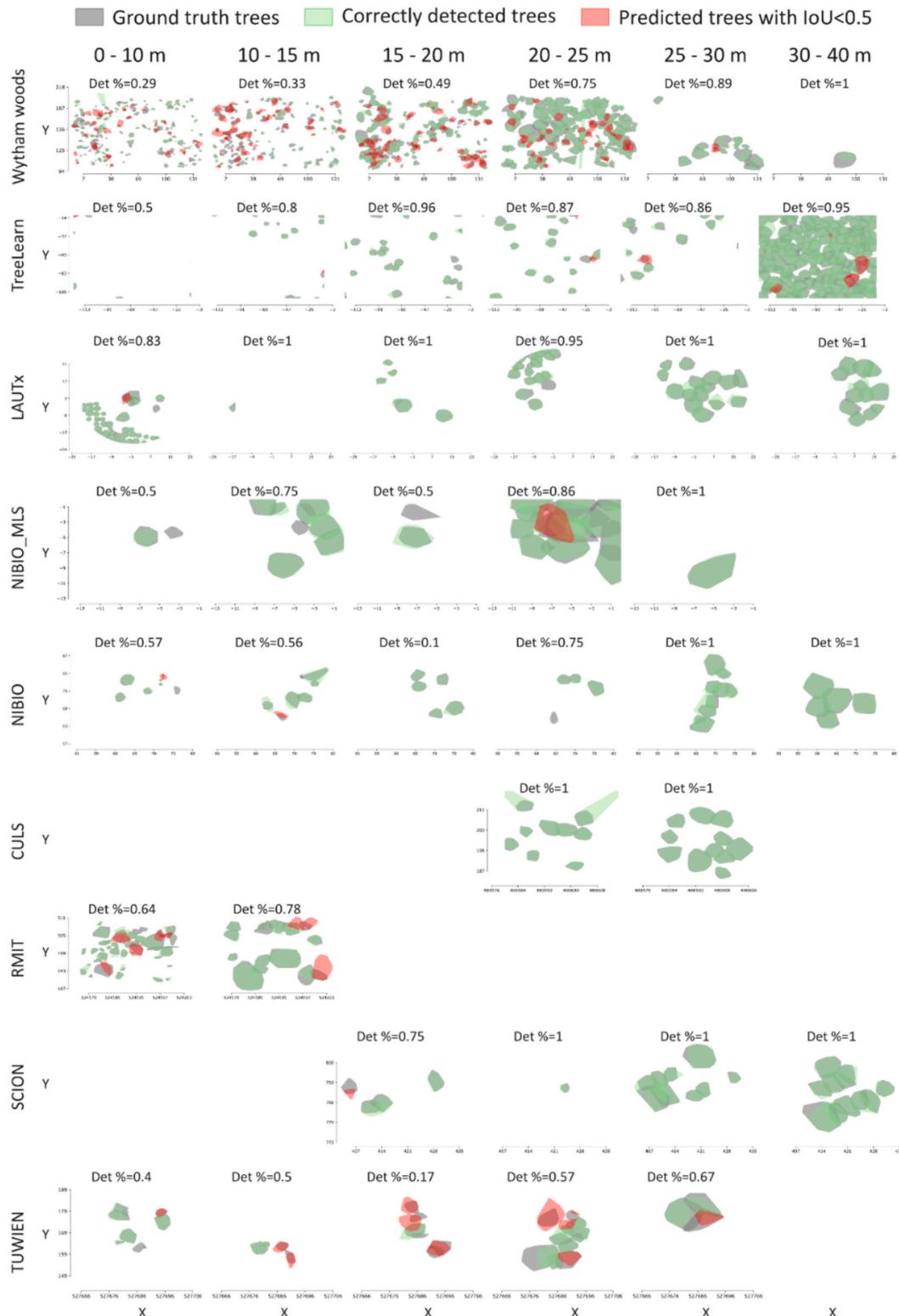


Fig. 6. Visual comparison of the detection rate (Det %) performance for different canopy layers (5 m high bins). To ensure clarity, only the detection rate for each slice is reported.

m^{-2}), and consumer-grade drone laser scanning data (661 points m^{-2}). Given that these datasets have not been manually annotated or have field-based ground data, it is only possible to do a visual assessment of the segmentation that shows promising results across all three datasets.

4.3. Performance across canopy profile

The performance analysis of the scenario 5 model across different canopy layers (see Fig. 6) showed variability based on the forest type (deciduous versus coniferous forests and complex multilayered forests versus single layered forests) and canopy structure (well-separated

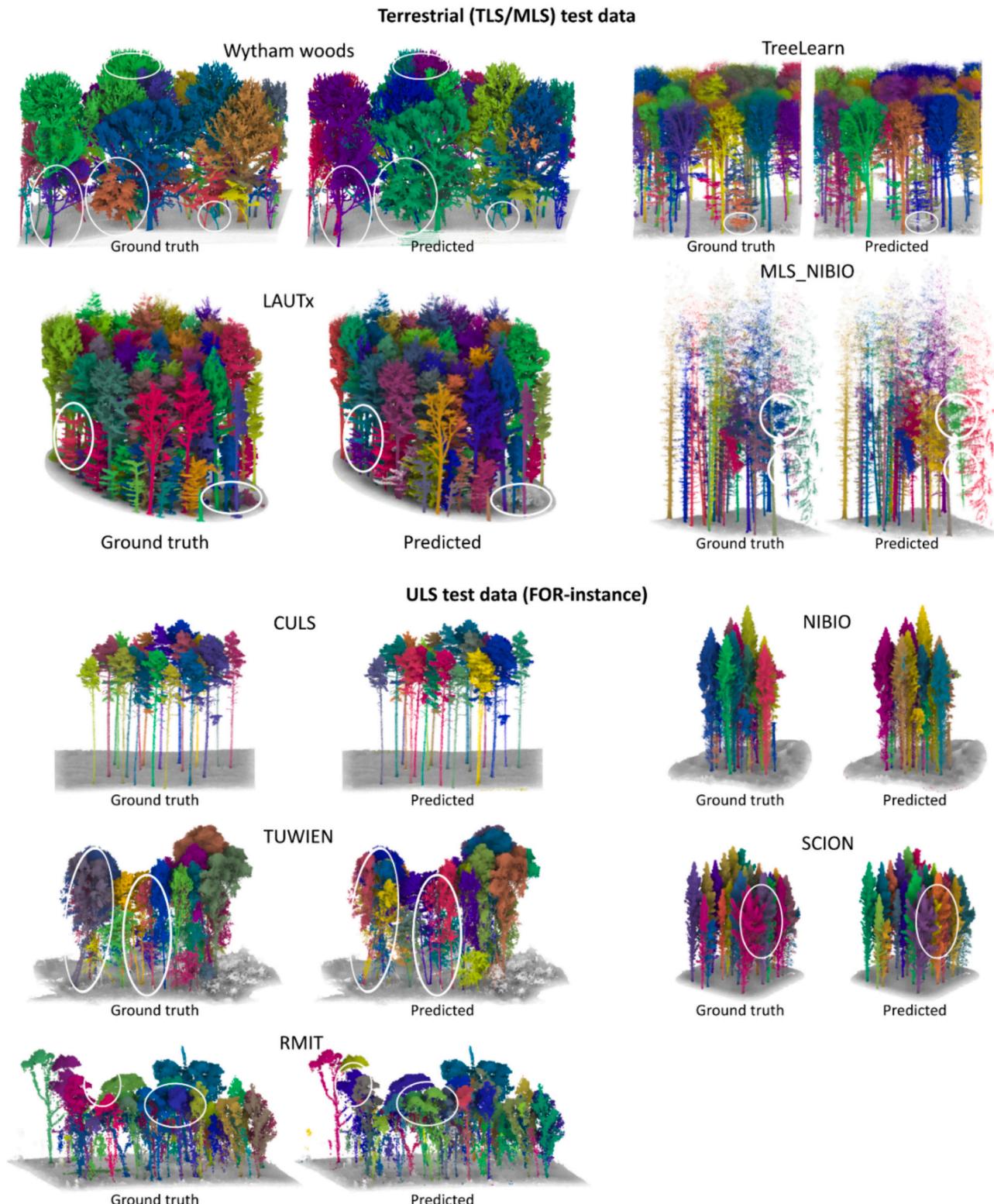


Fig. 7. Example point clouds from the test dataset and the respective predictions using scenario 5. The white ellipses highlight some main issues in the output segmented instances.

layers versus vertically connected layers). The model demonstrated transferability to both terrestrial and ULS data without a clear pattern of performance decline tied to specific data sources. Using tree height as a proxy for the social tree status, we found (see Fig. 6) that the model was able to segment trees throughout the canopy vertical profile, including understory suppressed trees (trees shorter than 10 m; detection rates in the range 0.2–0.8), co-dominant, and dominant trees.

Performance was lower in multilayered broadleaved forests (see Fig. 6), likely due to the predominance of managed boreal forests in the training data. The Wytham Woods dataset, characterized by a dense understory with 53% of the trees under 15 m tall, composed of young saplings and multi-stem trees, highlighted our model's limitation (detection rate is on average 37%) The underrepresentation of such forest structures in the training data likely contributed to the model's poor performance on this dataset.

The 3D visual evaluation of the model's predictions from scenario 5 (see Fig. 7) provides further evidence of the segmentation quality for most datasets, including the ability to detect dominated trees (Fig. 6, LAUTx and TreeLearn), trees very close to each other (Fig. 7, MLS_NIBIO). Fig. 6 also shows some challenges in accurately segmenting mixed temperate forests (Fig. 6, Wytham woods and TUWIEN), including the challenge in separating very plastic, intermingled, and asymmetric crowns.

4.4. Computational efficiency

The evaluation of the CE showed (See Table 4) that the method presented in this paper performs 80% more efficiently than what reported for an alternative method such as TLS2Trees (Wilkes et al., 2023). Translating the CE into processing time shows that our method represents a substantial improvement compared to TLS2Trees, and as an example, given a constant compute shape of a single GPU and CPU machine and a point cloud of 4 million points (127 MB), our method would require 5.6 min to run against 62 min required for TLS2Trees. While many studies do not report the computational efficiency of their methods, which, along with the availability of the code it is a very important aspect for the uptake of new methods.

4.5. Advancing the state-of-the-art

Our study evaluated our proposed method against the current SOTA by benchmarking it using well-established datasets, providing an objective and efficient approach towards tracking advancements in model development (Lines et al., 2022). Specifically, we compared our results to those from earlier studies using the benchmark datasets in Table 5.

Our study established the first comprehensive baseline performance for the entire 1.4 ha of Wytham Woods (Calders et al., 2022), where previous efforts like those by Wilkes et al. (2023) and Xu et al. (2023) focused on a smaller area. Similarly to the results by Wilkes et al. (2023), the complexity of this forest (i.e. multiple species, canopy layers, and

Table 4
Computational efficiency of our method against TLS2trees.

Method	Tested dataset	Time (min)	Size (MB)	CE (MB/Core/Min)
This study	LAUTx	403	11,264	0.44
	Wytham woods	309	5120	0.26
	FOR-instance	49	1638	0.52
	TreeLearn	57	480	0.13
	NIBIO_MLS	49	1331	0.42
	average	173	3966	0.354
	RUSH	101	300	0.015
	NOU	1245	8212	0.033
	MLA	1370	13,487	0.05
	TLS2trees*	905	7333	0.032

*values reported by Wilkes et al. (2023)

intermingled crowns) resulted in our lowest performance metrics with detection rates as low as 0.53. For the TreeLearn dataset, our method achieved a high detection rate (0.93) and F1-score (0.92), demonstrating a strong performance in simpler broadleaved forest structures characterized by single layered canopy. In the LAUTx dataset, our method largely outperformed the baseline methods (Point2Tree and TLS2trees) by Wielgosz et al. (2023), demonstrating improvements in all measured metrics by about 20–30%. Our method also showed substantial improvements in the NIBIO_MLS dataset compared to Point2Tree (Wielgosz et al., 2023), with detection rates and F1-scores rising by 17–27% and 26–27% respectively, indicating its suitability for managed coniferous forests.

For the FOR-instance data, our model performed similarly to the existing SOTA by Xiang et al. (2024), but significantly underperformed in the TUWIEN dataset. This indicates that while simple augmentation strategies work well in uniform forest structures, incorporating deeper semantic information and advanced synthetic data augmentation might be necessary for more complex forest environments.

4.6. Lessons learnt and future directions

The results demonstrated the feasibility of developing sensor- and platform-agnostic instance segmentation networks that deliver robust performance across a diverse range of lidar datasets and forest conditions. We have particularly emphasized the importance of data augmentation, which significantly enhanced segmentation quality. We found that our model maintained a high accuracy even at low point densities without the need for extensive hyperparameter adjustments, positioning them as a promising alternative to traditional ITC methods for ALS data.

Despite these achievements, there is substantial room for improvement to the SegmentAnyTree model. Currently, as the training data primarily consisted of coniferous managed forests, the model performs optimally in similar environments. It also shows promise in forest with a higher proportion of broadleaved species (i.e. LAUTx or TreeLearn datasets), although the performance drops when applying our models to complex forest types like TUWIEN and Wytham Woods. In these forests we find a more complex species composition and canopy structure. Traditional ITC segmentation approaches are particularly limited by presence of multiple layers, with an understory layer composed of saplings and multi-stem trees, and where highly plastic crowns of the different layers are intertwined and overlapping. With the increasing demand of accurate data to ensure the protection of forest ecosystems of high complexity and biodiversity value it is important to work towards extending the capability of end-to-end deep learning models like SegmentAnyTree to these forest types. We believe that the current performance of SegmentAnyTree in complex forest currently is more limited by training data than by the overall model architecture. However, high-quality labelled training data from complex forests remain difficult to manually label and thus these data are intrinsically scarce. One possibility to overcome such labelled data gap would be to explore the use of in self-supervised point cloud methods that might offer a way to learn from larger amounts of available unlabeled data and thanks to that improve the model's transferability.

Looking forward, further potential developments of SegmentAnyTree could result by leveraging its flexible architecture and include additional model head branches relying on the same backbone as the current instance and semantic segmentation heads. Amongst these, would be obvious to include additional model heads for tree species classification and for regression tasks (e.g. tree volume or wood quality). Further, the SegmentAnyTree model could be extended to perform a finer level of segmentation, where for example individual lying or standing deadwood logs are instance segmented. Such developments would not only increase the ecological value of SegmentAnyTree but also enhance the granularity of segmentation results.

Table 5

Summary of the performance metrics for existing studies against our best results for the tested datasets. The bold font indicates the best performances on each of the evaluated metrics.

Test dataset	Method	Detection (%)	Omission (%)	Commission (%)	RMSE H (m)	F1-score (%)
Wytham woods	Our method	0.53	0.47	0.27	4.2	0.88
TreeLearn	TreeLearn (Henrich et al., 2023)	–	–	–		0.98
	Our method	0.93	0.07	0.03	5.1	0.92
LAUTx	Point2tree (Wielgosz et al., 2023)	0.55	0.45	0.15	2.7	0.67*
	TLS2trees (Wielgosz et al., 2023)	0.405	0.6	0.3	3.6	0.63*
	Our method	0.75	0.25	0.09	3.1	0.88*
NIBIO MLS	Point2tree (Wielgosz et al., 2023)	0.57	0.43	0.07	3.47	0.61*
	TLS2trees (Wielgosz et al., 2023)	0.59	0.41	0.14	3.6	0.62*
	Our method	0.77	0.23	0.09	3.4	0.88*
FOR-instance	Straker et al. (2023)	0.67	0.33	–		–
NIBIO	Xiang et al. (2024)	0.88	0.12	0.03		0.92**
	Our method	0.88	0.12	0.09	3.4	0.88*
FOR-instance	Straker et al. (2023)	1	0	–		–
CULS	Xiang et al. (2024)	1	0	0.13		0.93**
	Our method	1	0	0	0.14	0.99
FOR-instance	Straker et al. (2023)	0.86	0.14	–		–
SCION	Xiang et al. (2024)	0.87	0.13	0.04		0.91**
	Our method	0.92	0.08	0.07	1.7	0.91
FOR-instance	Straker et al. (2023)	0.58	0.42	–		–
RMIT	Xiang et al. (2024)	0.64	0.36	0.24		0.7**
	Our method	0.69	0.31	0.17	1.3	0.83
FOR-instance	Straker et al. (2023)	0.2	0.8	–		–
TUWIEN	Xiang et al. (2024)	0.71	0.29	0.32		0.69**
	Our method	0.46	0.54	0.45	4.8	0.57

* local computation of metric (i.e. F1-score is computed only using correctly detected trees)

** global computation of metric (i.e. F1-score is computed for all trees)

5. Conclusion

The possibility of training agnostic models carries significant importance in streamlining the use of dense lidar data in forest inventory and is clearly an important research avenue. The transferability of instance segmentation model to airborne datasets has the potential to improve large-scale ITC inventories. Such model transferability is crucial in an era where dense lidar data is increasingly accessible through an increased number of sensors and platforms. The “one model fits all” approach simplifies individual tree crown methods, offering flexibility in data input and capture solutions, whether from the air or ground. In this context, our study shows the synergistic potential amongst various lidar data types, suggesting a convergence in segmentation techniques for both terrestrial and airborne point clouds—areas that have historically operated independently within the forest remote sensing community.

The potential of fully agnostic models to refine and streamline the characterization of forest environments using dense laser scanning data is significant. With the growing availability of open, machine learning-ready forest lidar datasets, improvements in our proposed model are anticipated. Moreover, advancements in vision transformers and self-supervised training techniques are expected to yield more efficient and effective architectures. While these methods will continue to evolve, the concept of training networks using broad, diverse datasets remains a promising direction for future research in forest remote sensing.

Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used OpenAI’s ChatGPT v4.0 to improve readability. After using this tool the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Maciej Wielgosz: Writing – original draft, Visualization, Validation,

Software, Methodology. Stefano Puliti: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Binbin Xiang: Writing – review & editing, Software, Resources, Conceptualization. Konrad Schindler: Writing – review & editing, Supervision, Project administration, Formal analysis. Rasmus Astrup: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

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.

Data availability

The data used in this study is publicly available at the following links: FOR-instance dataset: <https://zenodo.org/records/8287792>; NIBIO_MLS dataset: is available at: <https://zenodo.org/records/12754726>; Wytham woods dataset: <https://zenodo.org/records/7307956>; TreeLearn dataset: <https://data.goettingen-research-online.de/dataset.xhtml?persistentId=doi:10.25625/VPMPIID>; LAUTx dataset: <https://zenodo.org/records/6560112>; The code for training and prediction using the proposed deep learning model is available at: <https://github.com/SmartForest-no/SegmentAnyTree>.

Acknowledgements

This work is part of the Center for Research-based Innovation SmartForest: Bringing Industry 4.0 to the Norwegian forest sector (NFR SFI project no. 309671, smartforest.no).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2024.114367>.

References

- Ayrey, E., et al., 2017. Layer stacking: a novel algorithm for individual Forest tree segmentation from LiDAR point clouds. *Can. J. Remote. Sens.* 43, 16–27.
- Burt, A., et al., 2019. Extracting individual trees from lidar point clouds using treeseg. *Methods Ecol. Evol.* 10, 438–445.
- Calders, K., et al., 2022. Terrestrial laser scanning data Wytham Woods: individual trees and quantitative structure models (QSMs) (Version 1.0). Available at: <https://doi.org/10.5281/zenodo.7307956>.
- Cao, Y., et al., 2023. Benchmarking airborne laser scanning tree segmentation algorithms in broadleaf forests shows high accuracy only for canopy trees. *Int. J. Appl. Earth Obs. Geoinf.* 123, 103490.
- Commission, E., 2020. EU Biodiversity Strategy for 2030. Brussels. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52020DC0380>.
- Commission, E., 2021. New EU Forest Strategy for 2030. Brussels. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021DC0572>.
- Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods Ecol. Evol.* 7, 1236–1245.
- Ferraz, A., et al., 2012. 3-D mapping of a multi-layered Mediterranean forest using ALS data. *Remote Sens. Environ.* 121, 210–223.
- GeoSLAM, 2020. ZEB-HORIZON™ User's Manual. Available at: <https://geoslam.com/wp-content/uploads/2021/02/ZEB-Horizon-User-Manual-v1.3.pdf>.
- Hakula, A., et al., 2023. Individual tree segmentation and species classification using high-density close-range multispectral laser scanning data. *ISPRS Open J. Photogram. Remote Sens.* 9, 100039.
- Henrich, J., et al., 2023. TreeLearn: a comprehensive deep learning method for segmenting individual Trees from Forest point clouds. arXiv preprint arXiv: 2309.08471.
- Hyypä, E., et al., 2022. Direct and automatic measurements of stem curve and volume using a high-resolution airborne laser scanning system. *Sci. Remote Sens.* 5, 100050.
- Hyypä, J., et al., 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Trans. Geosci. Remote Sens.* 39, 969–975.
- Jiang, L., et al., 2020. Pointgroup: dual-set point grouping for 3d instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4867–4876). Available at:
- Kaartinen, H., et al., 2012. An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sens.* 4, 950–974.
- Krisanski, S., et al., 2021. Sensor agnostic semantic segmentation of structurally diverse and complex Forest point clouds using deep learning. *Remote Sens.* 13, 1413.
- Li, W., et al., 2012. A new method for segmenting individual trees from the Lidar point cloud. *Photogramm. Eng. Remote. Sens.* 78, 75–84.
- Liang, X., et al., 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS J. Photogramm. Remote Sens.* 144, 137–179.
- Lines, E.R., et al., 2022. AI applications in forest monitoring need remote sensing benchmark datasets. In: 2022 IEEE International Conference on Big Data (Big Data) (pp. 4528–4533). IEEE. Available at:
- Persson, H.J., et al., 2022. Two-phase forest inventory using very-high-resolution laser scanning. *Remote Sens. Environ.* 271, 112909.
- Popescu, S.C., et al., 2002. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Comput. Electron. Agric.* 37, 71–95.
- Puliti, S., et al., 2023a. FOR-instance: a UAV laser scanning benchmark dataset for semantic and instance segmentation of individual trees. arXiv preprint arXiv: 309.01279.
- Puliti, S., et al., 2023b. FOR-instance: a UAV laser scanning benchmark dataset for semantic and instance segmentation of individual trees (Version 1). Available at: <https://zenodo.org/records/8287792>.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *J. Big Data* 6, 60.
- Straker, A., et al., 2023. Instance segmentation of individual tree crowns with YOLOv5: a comparison of approaches using the ForInstance benchmark LiDAR dataset. *ISPRS Open J. Photogram. Remote Sens.* 9, 100045.
- Tao, S., et al., 2015. Segmenting tree crowns from terrestrial and mobile LiDAR data by exploring ecological theories. *ISPRS J. Photogramm. Remote Sens.* 110, 66–76.
- Tockner, A., et al., 2022. Lautx-individual tree point clouds from Austrian Forest Inventory Plots (p. 5281). Available at: <https://zenodo.org/records/6560112>.
- Weiser, H., et al., 2022. Individual tree point clouds and tree measurements from multiplatform laser scanning in German forests. *Earth Syst. Sci. Data* 14, 2989–3012.
- Wielgosz, M., et al., 2023. Point2Tree(P2T): framework for parameter tuning of semantic and instance segmentation used with mobile laser scanning data in coniferous Forest. *Remote Sens.* 15, 3737.
- Wilkes, P., et al., 2023. TLS2trees: A scalable tree segmentation pipeline for TLS data. *Methods Ecol. Evol.* 00, 1–17.
- Windrim, L., Bryson, M., 2020. Detection, segmentation, and model fitting of individual tree stems from airborne laser scanning of forests using deep learning. *Remote Sens.* 12, 1469.
- Winiwarter, L., et al., 2022. Virtual laser scanning with HELIOS++: a novel take on ray tracing-based simulation of topographic full-waveform 3D laser scanning. *Remote Sens. Environ.* 269, 112772.
- Xiang, B., et al., 2023. Towards accurate instance segmentation in large-scale LiDAR point clouds. arXiv. <https://doi.org/10.48550/arXiv.2307.02877> preprint arXiv:2307.02877. Available at:
- Xiang, B., et al., 2024. Automated forest inventory: analysis of high-density airborne LiDAR point clouds with 3D deep learning. *Remote Sens. Environ.* 305, 114078.
- Xu, X., et al., 2023. Topology-based individual tree segmentation for automated processing of terrestrial laser scanning point clouds. *Int. J. Appl. Earth Obs. Geoinf.* 116, 103145.