Remote Sensing of Environment 342 (2026) 115467

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Cross-platform forest understanding: A multi-platform synergistic training framework for generalized forest point cloud segmentation

Jundi Jiang [a,c,1], Yueqian Shen [a,1,*], Jinhu Wang [b,1,*], W. Daniel Kissling [b], Markus Hollaus [c], Hongjun Su [d], Jinguo Wang [a], Vagner Ferreira [a], Norbert Pfeifer [c]

[a] School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China
[b] Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, P.O. Box 94240, 1090, GE, Amsterdam, the Netherlands
[c] Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna 1040, Austria
[d] College of Geography and Remote Sensing, Hohai University, Nanjing 211100, China

ARTICLE INFO

Editor: Marie Weiss

Keywords:
Forest scene
Semantic segmentation
Instance segmentation
Cross-platform
Negative transfer
Point cloud

ABSTRACT

Precise forest inventories are fundamental for sustainable ecosystem management, biodiversity conservation, and assessment of forest carbon stocks. Light Detection and Ranging (LiDAR) has emerged as a dominant remote sensing technique capable of accurately characterizing the vertical canopy profile over large areas. While the growing availability of LiDAR datasets from diverse acquisition platforms has driven significant advances in forest structural component and individual tree segmentation tasks, inherent heterogeneity in data characteristics across multiple platforms poses a critical challenge for cross-platform generalization. Traditional data-driven algorithms often fail to generalize across multi-platform forest datasets due to substantial heterogeneity in data characteristics from different acquisition platforms. Furthermore, conventional multi-platform mixed training strategies induce negative transfer, ultimately undermining segmentation performance. To address the inherent heterogeneity and negative transfer challenges arising from multi-platform datasets, we introduce the **M**ulti-platform **S**ynergistic **T**raining (MST) framework, a unified, data- and model-driven representation learning framework. This framework initially pretrains with virtual synthetic forest datasets to extract generalized feature representations, followed by platform-specific fine-tuning utilizing real-world datasets. Extensive experiments were conducted to evaluate the proposed MST on semantic and instance segmentation tasks using forest benchmark datasets from multiple platforms. The results reveal that the MST achieves consistently high segmentation performance across multi-platform datasets (from airborne, unmanned aerial vehicle, mobile and terrestrial laser scanning). Additionally, leveraging MST pretraining enables the use of only 20% of labeled real-world data to match the segmentation accuracy achieved by training on fully annotated datasets. The MST framework therefore represents a powerful and effective representation learning framework with the potential to support downstream forest inventory and ecological applications.

1. Introduction

Forests are critical ecosystems that sustain ecological processes, biodiversity conservation, and climate regulation (Maes et al., 2023; Mo et al., 2023). Effective forest resource management necessitates accurate, timely, and comprehensive spatial data acquisition. Traditional field surveys are labor-intensive, time-consuming, and prohibitive costly, and limited in spatial coverage, constraining their effectiveness for large-scale inventories (Panagiotidis et al., 2022). Optical remote sensing has emerged as an efficient technique for extensive forest data acquisition due to its ability to cover large geographic areas rapidly (Lei et al., 2025; Li and Fang, 2025; Meng et al., 2024; Tong and Zhang, 2025; Zeng et al., 2025). However, optical imagery typically lacks the capability to capture the complex three-dimensional (3D) structure of forest environments (Balestra et al., 2024; Coops et al., 2021). This lack of 3D structural information limits the effectiveness of forest resource

* Corresponding authors at: School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China, and Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, P.O. Box 94240, 1090 GE Amsterdam, The Netherlands
E-mail addresses: jdjiang@hhu.edu.cn (J. Jiang), y.shen_lidar@hhu.edu.cn (Y. Shen), jinhu.wang@hotmail.com (J. Wang).
[1] Jundi Jiang, Yueqian Shen and Jinhu Wang contributed equally to this work and should be considered co-first author.

https://doi.org/10.1016/j.rse.2026.115467
Received 17 June 2025; Received in revised form 26 April 2026; Accepted 3 May 2026
Available online 14 May 2026

assessments, particularly in heterogeneous, multilayered forest systems.

In recent decades, laser scanning technology has revolutionized forest structural analysis by providing precise 3D point cloud representations of forests. Laser scanning systems, including airborne laser scanning (ALS), unmanned aerial vehicle-based laser scanning (ULS), mobile laser scanning (MLS), and terrestrial laser scanning (TLS) (Han et al., 2024), are characterized by distinct spatial resolution, coverage extent, and data density. Specifically, MLS and TLS excel in capturing highly detailed and dense point cloud data at the plot or individual tree scale, enabling precise characterization of forest structural attributes (Ali et al., 2025; Balestra et al., 2024; Ghorbani et al., 2024; Kükenbrink et al., 2025). However, the restricted operational scalability limits their practical applicability for large-scale forest inventories. In contrast, ALS and ULS facilitate rapid acquisition of large-area datasets but generally yield lower-density point clouds, potentially compromising fine-scale structural delineation (Bruggisser et al., 2024; Gavilán-Acuna et al., 2024; Lu et al., 2025; Yun et al., 2021). The increasing availability and diversity of forest point cloud datasets collected from multiple platforms in real-world scenes present unprecedented opportunities for advanced forestry analyses (Calders et al., 2022; Liu et al., 2026; Lu et al., 2025; Puliti et al., 2023; Weiser et al., 2022; Wielgosz et al., 2023). Furthermore, synthetic forest point cloud datasets have become an alternative to labor-intensive manual annotation, offering error-free labels and allowing for customizable acquisition parameters (Dobbs et al., 2023; Lin et al., 2020; Liu et al., 2026; Lu et al., 2024; Tang et al., 2024). These virtual datasets enable the simulation of diverse forest typologies (e.g., boreal, temperate) and laser scanning platforms through physics-induced sensor modeling and flexible survey configurations. Synthetic data generation typically integrates detailed tree models with laser scanning simulators (Winiwarter et al., 2022), systematically synthesizing forest scenes while preserving plant morphological attributes and sensor-specific acquisition geometries (González-Quiñones et al., 2024; Wang et al., 2022a). These datasets constitute essential resources for promoting research on large-scale 3D forest scene interpretation and structural parameter estimation.

Semantic and instance segmentation are critical for accurately extracting forest attributes from laser-scanning point cloud data. **Semantic segmentation** (i.e., forest component segmentation) systematically classifies each point within the dataset into meaningful ecological and structural components, such as stems, branches, foliage, and terrain (Liang et al., 2025; Ruoppa et al., 2025). This categorization enables detailed ecological assessments and precise derivation of biophysical attributes, including leaf density (Shen et al., 2024), total wood volume (Rodda et al., 2024), and canopy height (Komárek et al., 2024). These parameters are essential for understanding forest ecosystem functioning, assessing carbon stock potentials and monitoring forest health. **Instance segmentation** (i.e., individual tree segmentation) isolates individual objects within complex forest environments, such as single trees. Instance segmentation provides critical insights for tree height estimation (Sun et al., 2025), tree-specific biomass estimations (Lu and Jiang, 2024; Oehmcke et al., 2024; Wild et al., 2026), and tree species classification (Ma et al., 2024; Wang et al., 2024). Moreover, after individual tree segmentation, quantitative biophysical attributes, including diameter at breast height (DBH), and crown width, can be systematically retrieved at the individual tree level. These allometric attributes constitute variables for aboveground biomass estimation (Beyene et al., 2020; Gaikadi and Selvaraj, 2024).

Traditional semantic and instance segmentation methods for forest scenes typically relied on predefined geometric rules or handcrafted features (Li et al., 2012; Paris et al., 2016). However, these traditional approaches exhibit limited scalability and generalizability across diverse forest environments. Recent advances in deep learning have revolutionized forestry analysis, primarily due to its ability to automatically extract representative features directly from data (Jarahizadeh and Salehi, 2025). Notably, these advancements are further driven by the growing availability of high-quality point cloud datasets collected from various platforms. However, forest point cloud data acquired across different platforms exhibit substantial heterogeneity and complexity, posing significant challenges for generalized segmentation. Specifically, the complexity of forest scenes primarily stems from variations in vegetation types, species diversity, and terrain characteristics (Lu et al., 2025). The overlapping of tree crowns and branching structures further complicates the segmentation process. Moreover, heterogeneity among forest point cloud datasets arises from differences in data acquisition techniques across platforms. As shown in Fig. 1, MLS and TLS, which are ground-based approaches, generate dense point clouds with highly detailed trunk structures. However, occlusion and limited viewing angles often lead to incomplete canopy coverage. In contrast, ULS conducted from elevated platforms, such as drones, offers comprehensive canopy coverage but suffers from missing or sparse trunk structures due to vegetation obstruction. ALS, typically deployed at higher altitudes, produces lower-density point clouds with partial data loss in both the canopy and trunks, primarily due to occlusion and sensor limitations at higher altitudes.

The inherent complexity and heterogeneity result in poor generalizability of models trained on single-platform datasets to other platform datasets. Intuitively, merging multi-platform forest datasets and collaboratively training a single model is a potential solution. However, differences in feature distributions and data characteristics across platforms tend to bias the model toward learning platform-specific features rather than capturing shared structural representations, ultimately leading to performance degradation, a phenomenon known as negative transfer. Negative transfer (Caruana, 1997) refers to the phenomenon in which learning from one dataset can negatively affect performance on another dataset due to differences in data distribution (Wu et al., 2024b). Little effort has been put into addressing this negative transfer across multiple platform datasets in the forest scene. Existing approaches fail to account for inherent domain gaps and negative transfer effects that emerge during joint training on multi-platform LiDAR datasets. This limitation hinders their stability, transferability and scalability in real-world cross-platform applications.

To address these challenges, we developed a cross-platform representation learning paradigm that enables consistent forest point cloud segmentation across different LiDAR acquisition platform data, rather than developing a novel semantic or instance segmentation network. Our aims were to (1) quantify the generalization limitations of single-platform training and characterize the negative transfer effects that arise under mixed training of multi-platform forest point clouds; (2) design a framework that explicitly accounts for data heterogeneity while promoting shared structural representations that can be integrated into both semantic and instance segmentation backbones; and (3) evaluate the proposed method on multiple-platform datasets, and assess how it improves label efficiency in downstream training. We achieved this by.

(1) A systematic empirical analysis that reveals the limitations of single-platform training and that uncovers negative transfer induced by mixed training with multi-platform forest point cloud data.
(2) Introducing a multi-platform synergistic training (MST) framework, which combines learnable Cross-Platform-Aware Tokens (CPATs) with a plug-and-play Contextual Integration Module (CIM) to reconcile platform heterogeneity and encourage the learning of platform-invariant structural features.
(3) Extensive evaluations on nine benchmark datasets, which demonstrate consistent cross-platform generalization. Additional experiments show that using 20% of real-world labeled data can reach the accuracy level achieved with full supervision.

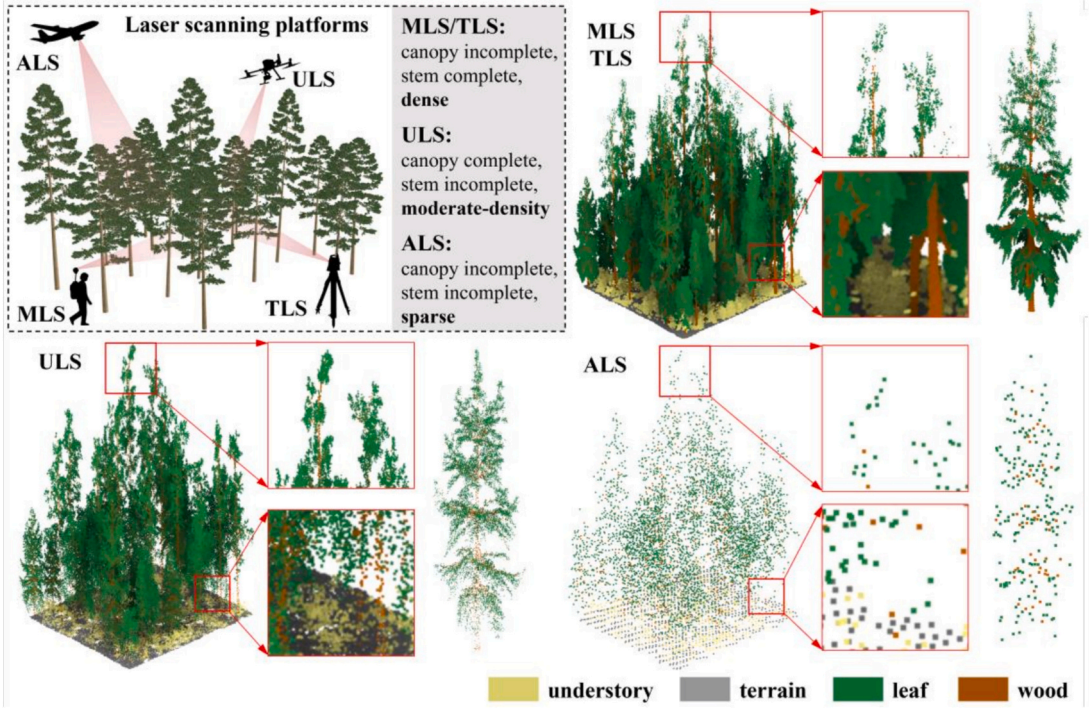

Fig. 1. Schematic of the characteristics of forest point cloud data acquired by different platforms. Blender software (https://www.blender.org/) was used to generate tree models in this figure.

2. Related work

2.1. Point cloud segmentation in forestry

Point cloud segmentation in forestry comprises semantic and instance segmentation. Semantic segmentation assigns each point to an ecologically meaningful class. Early works on semantic segmentation of forest point clouds relied on explicit rules or handcrafted geometric descriptors. Typical processing pipelines involve removing terrain by ground filtering, e.g., CSF, Cloth Simulation Filter (Zhang et al., 2016), detecting stems through cylinder fitting or vertical density profiles, and discriminating woody components from foliage based on local geometric descriptors (Dong et al., 2022; Hui et al., 2021; Shao et al., 2023; Vicari et al., 2019; Wan et al., 2021; Wang, 2020; Xu et al., 2021). These rule-based methods provided reasonable results in relatively simple environments but proved difficult to generalize (Shao et al., 2023), which required extensive parameter tuning. Deep learning has advanced this field by learning hierarchical representations directly from raw 3D coordinates, thereby removing the need for handcrafted features (Jiang et al., 2023; Lu et al., 2025; Zhang et al., 2022). However, trees and understory vegetation (e.g., shrubs) form spatially continuous structures that require long-range contextual information. Lu et al. (2025) introduced a long-range context enhancement module that explicitly incorporates global forest structure, improving recognition of large-scale elements such as crowns and ground surfaces. Additionally, forest point clouds encompass features ranging from understory vegetation to trunks and canopies, which require multi-scale feature learning (Chen et al., 2022). Zhong et al. (2025) proposed a multi-scale feature learning mechanism to differentiate wood from leaves by a Dual Point Attention module. Furthermore, multi-scale attention in transformer backbones further supports aggregation across variable point densities and tree sizes (Lu et al., 2025).

Instance segmentation in forests aims to identify each tree as a separate instance. This task is generally divided into raster-based and point-based individual tree segmentation method. Raster-based methods typically begin by generating a digital surface model (DSM) or a digital height model (DHM) from the point cloud. Image processing methods, such as watershed segmentation (Li et al., 2023b; Yang et al., 2020; Yun et al., 2021), are then applied to delineate tree boundaries. While these methods can effectively extract tree boundaries, they are inherently limited in detecting suppressed understory trees. To mitigate this limitation, Yao et al. (2012) extended the CHM–watershed pipeline by performing a normalized graph cut on a voxel-based representation of the forest, where stem positions are used as prior knowledge to guide the segmentation. More recently, deep learning methods have also been applied to delineate crown boundaries from CHM/DSM representations (Chang et al., 2022a; Chen et al., 2021; Straker et al., 2023), but they remain limited in detecting suppressed understory trees. In contrast, point-based methods directly assign each point to a specific tree instance. In this category, several physical- and geometry-driven approaches construct explicit structural cues and then delineate tree instances through unsupervised partitioning (Lu et al., 2014; Shendryk et al., 2016). Typical pipelines first extract candidate stems or trunks using geometric primitives or 3D shape descriptors, followed by grouping or clustering surrounding points into individual trees using spatial proximity and continuity constraints (Amiri et al., 2017; de Paula Pires et al., 2022; Ding et al., 2025). In parallel, unsupervised graph-based formulations represent the point cloud (or its voxel/superpoint) as a graph, where nodes encode local geometric elements and edges capture neighborhood relationships (Liu et al., 2025; Wang, 2020). Individual trees are then obtained via graph cutting or energy-minimization procedures (e.g., Normalized Cut) that enforce intra-tree coherence and inter-tree separation (Xi and Hopkinson, 2022; Yao et al., 2012). However, these methods often rely on the sampling geometry and point distribution statistics that define the graph topology and edge weights. Consequently, when the acquisition platform changes, substantial variations in point density and viewing direction can alter neighborhood structures and affinity statistics, inducing pronounced domain shifts in the constructed graphs. Additionally, other common pipelines fuse and co-register multi-platform point clouds (e.g., UAV and ground-based LiDAR) to obtain more complete and informative geometry (Fekry et al., 2022; Polewski et al., 2019; Shao et al.,

2022), then perform individual tree segmentation on the integrated data. However, registration errors can further propagate into downstream segmentation, and the requirement for multi-platform acquisitions over the same area limits scalability and increases operational costs. Recent advances in deep learning have attempted to tackle the above challenges through two paradigms. Grouping-based methods conceptualize instance segmentation as a clustering process, grouping points belonging to the same tree (Wielgosz et al., 2024; Xiang et al., 2024). To resolve the ambiguity of overlapping crowns, these methods often introduce an offset branch that predicts a vector from each point to its corresponding instance center (Jiang et al., 2020; Vu et al., 2022; Zhong et al., 2022). The offset-transformed points are then clustered to form tree instances. In contrast, query-based methods aim to reduce under-segmentation by ensuring comprehensive coverage of all trees in a scene (Xiang et al., 2025a). They employ strategies such as farthest point sampling (Eldar et al., 1997) to generate a set of query points distributed across the scene (Lu et al., 2023; Ngo et al., 2023). Each query serves as a potential tree candidate, allowing the network to progressively identify instances while mitigating the risk of missing suppressed or small trees. However, in dense forests, overlapping and interlaced crowns make it difficult to assign points unambiguously to individual trees. Furthermore, forests often exhibit multilayered structures with considerable variability in crown size. Dominant trees may be over-segmented into multiple instances, whereas suppressed trees are frequently neglected. These structural complexities highlight the difficulty of achieving reliable individual tree segmentation.

In conclusion, learning-based methods have demonstrated strong capabilities in both semantic and instance segmentation of forest point clouds. Nevertheless, their practical application in forestry remains constrained by the challenge of achieving consistent segmentation across heterogeneous conditions. Variations in sensor platforms introduce significant domain shifts, making it difficult for models trained in one setting to generalize effectively to others. Current methods explicitly designed to enhance generalization under data and scene heterogeneity remain limited. Bridging this gap will be essential for advancing the operational use of segmentation models in forestry, ensuring reliable performance across diverse environments and datasets.

2.2. Negative transfer

Negative transfer refers to the phenomenon in which transferring knowledge from a source domain results in reduced performance on the target task (Chang et al., 2022b; El Mendili et al., 2025). It arises from discrepancies between the joint distributions of the source and downstream tasks over both domain and label spaces, as well as from differences in data feature distributions (Gong et al., 2016; Wang et al., 2019). Negative transfer has been observed across a wide range of application domains, including architecture, engineering, and construction (Chen et al., 2024b; Rauch and Braml, 2025), autonomous driving (Wang et al., 2022c), remote sensing image recognition (dos Santos et al., 2025; Wang et al., 2022b), and materials science (Zhang et al., 2019). For optical remote sensing recognition, the domain gap between satellite image sources led to a decline in classification accuracy, indicating adverse knowledge transfer (dos Santos et al., 2025). These observations collectively highlight that negative transfer is most likely when the source and target domains or tasks are insufficiently aligned, leading to imported knowledge that confuses rather than supports the target model.

To alleviate negative transfer, prior studies have explored domain-conditioning strategies that adapt the model's feature distributions using domain identity or target-domain statistics (Chang et al., 2019; Li et al., 2018). A representative work focuses on normalization-based conditioning, where Batch Normalization (BN) is adapted to the target domain by updating the normalization statistics (mean and variance) from target-domain data (AdaBN) or by maintaining domain-specific BN parameters while sharing the remaining network weights (Chang et al., 2019). Related strategies parameterize normalization layers using a conditioning signal (e.g., a domain indicator or domain embedding) (de Paula Pires et al., 2022; Perez et al., 2018). Additionally, adapter-based designs introduce lightweight domain-specific capacity without retraining the entire backbone (Houlsby et al., 2019). Beyond explicit conditioning, domain adaptation commonly combines feature-level alignment with self-training and consistency regularization to reduce cross-domain discrepancy (Ganin et al., 2016; Qin et al., 2019). Additionally, adversarial strategies have also been explored in related 3D point cloud adaptation contexts, further indicating their feasibility for reducing cross-domain discrepancy in geometric data (Li et al., 2023a; Xiao et al., 2022). In point clouds, recent test-time adaptation updates BN statistics and affine parameters directly using unlabeled test-time data, providing a practical mechanism to mitigate domain shifts at inference time (Wang et al., 2025). Furthermore, several weakly supervised frameworks can be interpreted as implicitly conditioning the domain by leveraging unlabeled target-domain data to reshape the model's predictive distribution. They enforce output regularization on unlabeled points (e.g., entropy/consistency with pseudo-label refinement) and impose contextual constraints, optionally combined with active learning to query informative labels under domain shifts (Wang and Yao, 2022; Wang et al., 2023).

Despite these advances, several limitations remain when confronting cross-platform forest point clouds. Normalization-based conditioning often addresses distributional mismatch at the feature-statistics level, but it does not explicitly model platform-specific sampling mechanisms that reshape local neighborhoods and geometric evidence, leaving residual domain gaps that can still trigger negative transfer. These gaps motivate learning strategies that explicitly account for platform heterogeneity while promoting shared structural representations across domains, a focus of the proposed MST framework. Moreover, most domain-conditioning and adaptation methods are primarily designed to align feature distributions for semantic prediction, while they offer limited mechanisms to preserve instance-level grouping consistency (e.g., boundary separability and point-to-instance assignment) under platform-induced changes in sampling geometry, which is critical for reliable individual-tree instance segmentation.

3. Materials and methods

In this study, we introduce a practical and effective training framework: a model- and data-driven representation learning framework designed for generalized forest point cloud segmentation across multiple LiDAR acquisition platforms. In this section, we first identify the phenomenon of negative transfer caused by inherent heterogeneity in multi-platform forest point cloud datasets. To address this challenge, we propose the **MST** framework, which integrates datasets from multiple platforms to enhance model generalization. The framework overview is provided in Fig. 2.

3.1. Materials

3.1.1. Datasets

3.1.1.1. Virtual synthetic dataset. The Boreal3D dataset, synthesized using a simulation-to-reality framework integrated with digital cousins technology, was designed to overcome the scarcity of annotated forest point cloud data. Boreal3D is a large-scale, virtual synthetic forest point cloud dataset, containing 1000 procedurally generated forest plots simulating ALS, ULS, MLS, and TLS acquisitions (Liu et al., 2026). These simulations collectively encompass 35.3 billion points representing 48,403 individual trees across dominant boreal species, including *Pinus sylvestris*, *Picea abies*, and *Betula* spp. Each point is annotated with semantic classes (leaf, wood, understory, and terrain) and with a tree ID for each tree. To ensure ecological fidelity, Boreal3D implements three

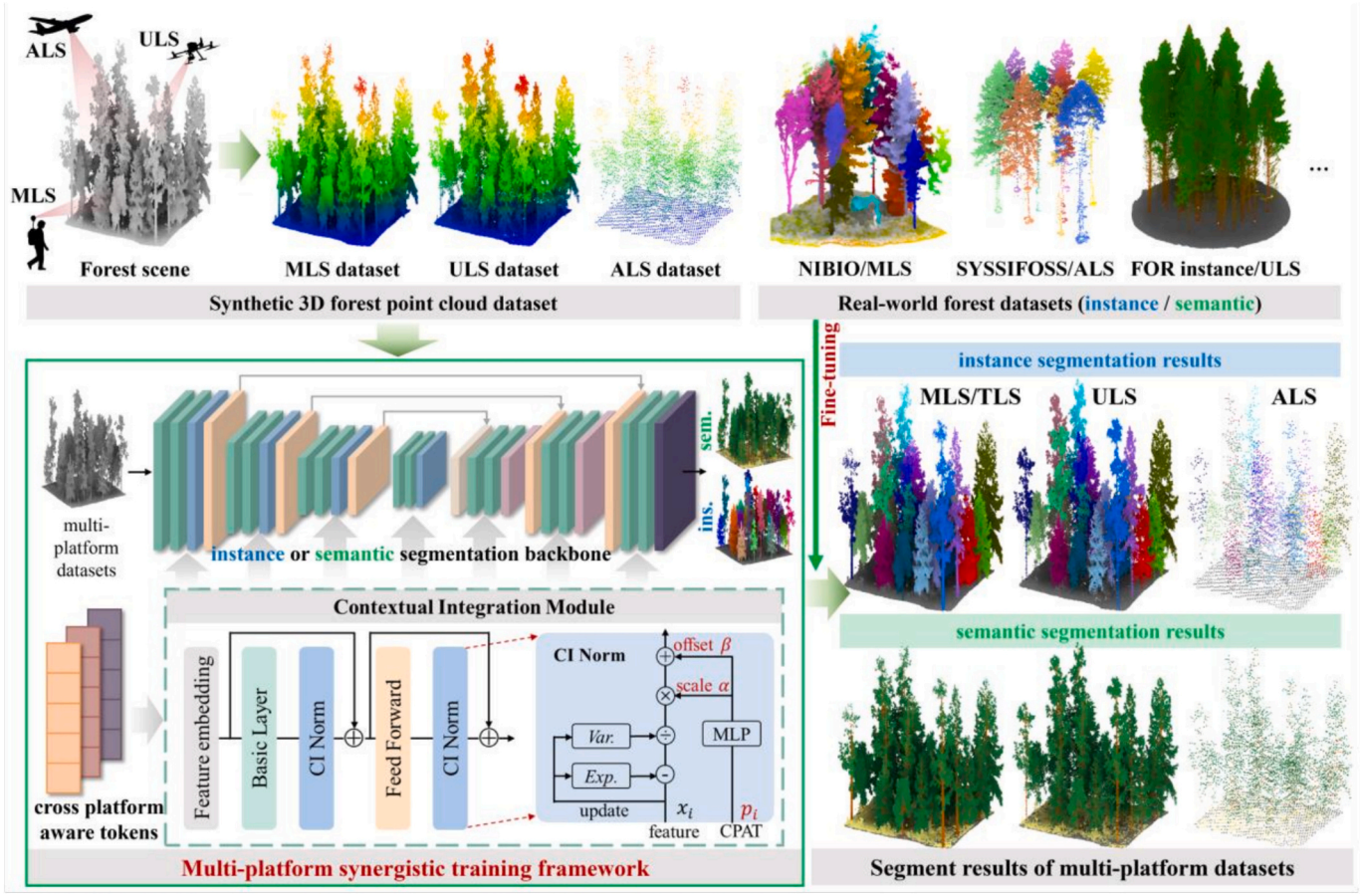

Fig. 2. The overall framework of the Multi-platform Synergistic Training (MST) for semantic and instance segmentation of forest scenes across multi-platform datasets. The MST framework can be applied to existing semantic or instance segmentation backbones.

complexity levels, including easy plots (homogeneous stands with minimal understory occlusion), medium plots (mixed-canopy structures with partial occlusions), and difficult plots (multi-layered vegetation with dense understory and complex terrain). More detailed information is presented in Table 1 and Fig. 3.

3.1.1.2. Real-world datasets

3.1.1.2.1. ALS dataset. The SYSSIFOSS dataset provides multi-platform georeferenced point clouds acquired in two mixed temperate forest stands in southwestern Germany, serving as a benchmark for forest structural analysis (Weiser et al., 2022). The ALS component, collected using a RIEGL VQ-780i sensor aboard a Cessna C207 aircraft at 600 m AGL, features an average pulse density of 72.5 points/m^2 across 12 circular forest plots (30 m radius). Further details are shown in Table 1 and Fig. 4.

3.1.1.2.2. ULS datasets. Three publicly available ULS datasets, FOR-Instance (Puliti et al., 2023), Lin3D v0.2 (Lu et al., 2025), and Yuchen (Bai et al., 2023) were utilized. The FOR-Instance dataset comprises UAV-based laser scanning data collected from diverse forest ecosystems using Riegl VUX-1 and MiniVUX-1 UAV sensors. It covers various forest types, including boreal, temperate, alluvial, dry sclerophyll, and coniferous plantation forests. This benchmark provides semantic labels and instance identifiers. The Lin3D v0.2 dataset, collected using Genius 16 and Riegl VUX-1 UAV LiDAR systems, provides semantic labels for *foliage*, *wood* (trunks and branches), *ground*, and *lower vegetation*. The Yuchen dataset contains UAV-based laser scanning data acquired with a Riegl miniVUX-1UAV sensor over a 14,000 m^2 plot at the Paracou Research Station, French Guiana. This species-rich tropical forest on the Guiana Shield has an average canopy height of 27.2 m and maximum crowns reaching 44.8 m. Each point is semantically labeled as *leaf* or *wood*, and every individual tree carries a unique instance ID for tree-level analyses. More information on FOR-Instance, Lin3D v0.2, and Yuchen datasets is detailed in Table 1 and Fig. 4.

3.1.1.2.3. MLS datasets. TreeLearn (Henrich et al., 2024) and NIBIO (Wielgosz et al., 2023) were employed to represent mobile platform data. The TreeLearn dataset comprises MLS point clouds acquired from 19 temperate deciduous forest plots spanning four regions in Germany. Data collection employed a GeoSLAM ZEB Horizon scanner under leaf-off conditions to optimize stem visibility. This dataset provides instance-level annotations, generated through Lidar360 software segmentation with manual refinement. The NIBIO dataset was collected in a managed boreal coniferous forest in southeastern Norway, utilizing a GeoSLAM ZEB-HORIZON MLS system. It comprises 16 plots that represent a variety of tree species compositions, primarily Norway spruce, Scots pine, and birch. The dataset includes semantic annotations categorizing *ground*, *vegetation* (including branches, leaves, and low vegetation), *coarse woody debris*, and *stems*, along with precise instance segmentation labels.

3.1.1.2.4. TLS datasets. Two TLS datasets were incorporated to enable comprehensive 3D forest characterization, including Wytham Woods (Calders et al., 2022) and Lin3D v0.1 (Lu et al., 2025). The Wytham Woods dataset was collected using a RIEGL VZ-400 terrestrial laser scanner at Wytham Woods, a temperate broadleaf forest in the UK. This dataset provides detailed instance annotations for individual trees, making it valuable for precise segmentation tasks. Lin3D v0.1 complements the previously introduced Lin3D dataset, additionally offering TLS-derived point cloud data with semantic labels for *foliage*, *wood*, *ground*, and *lower vegetation*.

Table 1
Detailed information of selected forest point cloud datasets from different laser scanning platforms.

Scenario	Datasets	Platform	Forest type (Tree species)	Country	Label	Number of plots			Point density (pts/m^{-2})
						Train	Val	Test	
Real world	FOR-Instance (Puliti et al., 2023) -CULS	ULS	- Coniferous dominated temperate forest (*Pinus sylvestris*) *	Czech Republic	0: unclassified 1: low-vegetation 2: terrain 3: out-points 4: stem 5: live branches 6: woody branches Tree ID	1	1	1	2585
	FOR-Instance (Puliti et al., 2023) -NIBIO		- Coniferous dominated boreal forest (*Picea abies, Pinus sylvestris, Betula* sp. (few)) *	Norway		8	6	6	9526
	FOR-Instance (Puliti et al., 2023) -RMIT		- Native dry sclerophyll eucalypt forest (*Eucalyptus* sp.) *	Australia		1	0	1	498
	FOR-Instance (Puliti et al., 2023) -SCION		- Non-native pure coniferous temperate forest (*Pinus radiata*) *	New Zealand		2	1	2	4576
	FOR-Instance (Puliti et al., 2023) -TUWIEN		- Deciduous dominated temperate forest (several deciduous species)*	Austria		1	0	1	1717
	Yuchen (Bai et al., 2023)		Tropical forest	French Guiana	1: wood 2: leaf Tree ID	1	1	1	1000
	Lin3D v0.2 (Lu et al., 2025)		Subtropical evergreen broadleaf forest (*Eucalyptus, Magnolia,* Chinese fir, *Castanopsis,* Mongolian Scots pine, Larch)	China	0: ground 1: lower objects 2: wood 3: foliage	3	1	2	944
	Tree Learn (Henrich et al., 2024)		Deciduous Temperate Mixed Forest	Germany	Tree ID	40	28	28	1224
	NIBIO (Wielgosz et al., 2023)	MLS	Coniferous dominated boreal forest (*Picea abies, Pinus sylvestris, Betula* sp.) *	Norway	1: ground 2: vegetation 3: lying deadwood 4: stems Tree ID	8	4	4	20,000
	SYSSIFOSS (Weiser et al., 2022)	ALS	Deciduous Temperate Mixed Forest (*Quercus robur, Robinia pseudoacacia, Acer pseudoplatanus, Tilia cordata, Pseudotsuga menziesii*)	Germany	Tree ID	6	3	3	72
	Wytham woods (Calders et al., 2022)		Deciduous temperate forest (*Fraxinus excelsior, Acer pseudoplatanus, Corylus avellana*) *	UK	Tree ID	16	12	12	9378
	Lin3D v0.1 (Lu et al., 2025)	TLS	Same as Lin3D v0.2	China	0: ground 1: lower objects 2: wood 3: foliage	3	1	2	15,658
Synthetic	Boreal3D (Liu et al., 2026)	ALS ULS MLS TLS	(Pine, Birch, Spruce)	/	0: understory 1: terrain 2: leaf 3: wood Tree ID	70	20	10	ALS: 13 ULS: 908 MLS: 36,500 TLS: 29,726

* Indicates information referenced from Xiang et al. (2025a).

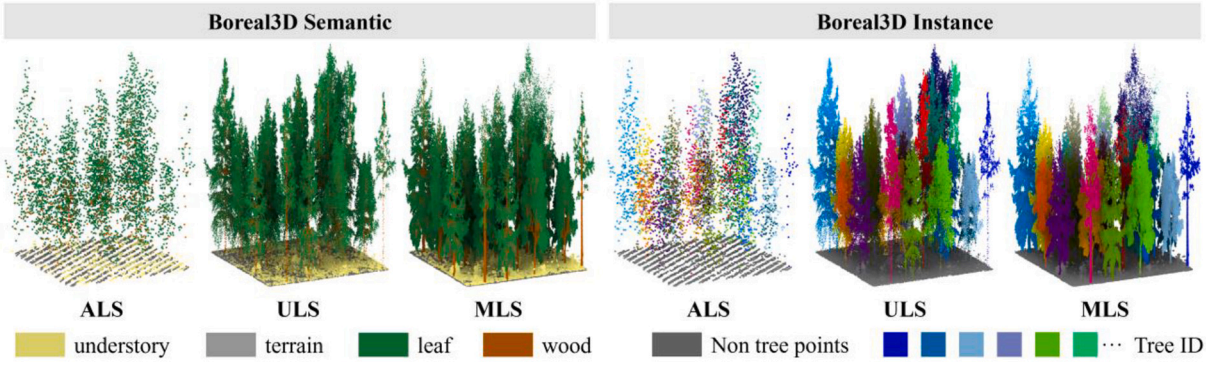

Fig. 3. Visualization of the virtual synthetic dataset (Boreal3D) with its semantic and instance labels.

3.1.2. Dataset processing

Significant differences exist among the aforementioned forest datasets, particularly in their semantic labels. These inconsistencies present challenges when evaluating semantic segmentation performance, potentially biasing model evaluation and limiting generalization across platforms. To address these discrepancies, we implemented a processing pipeline to harmonize the semantic annotations across the different real-world datasets.

Specifically, the FOR-Instance dataset originally includes seven semantic categories (Table 1), of which the *unclassified* and *out-points* categories represent unannotated points or measurement noise and were excluded. The NIBIO dataset contains a semantic category named *lying*

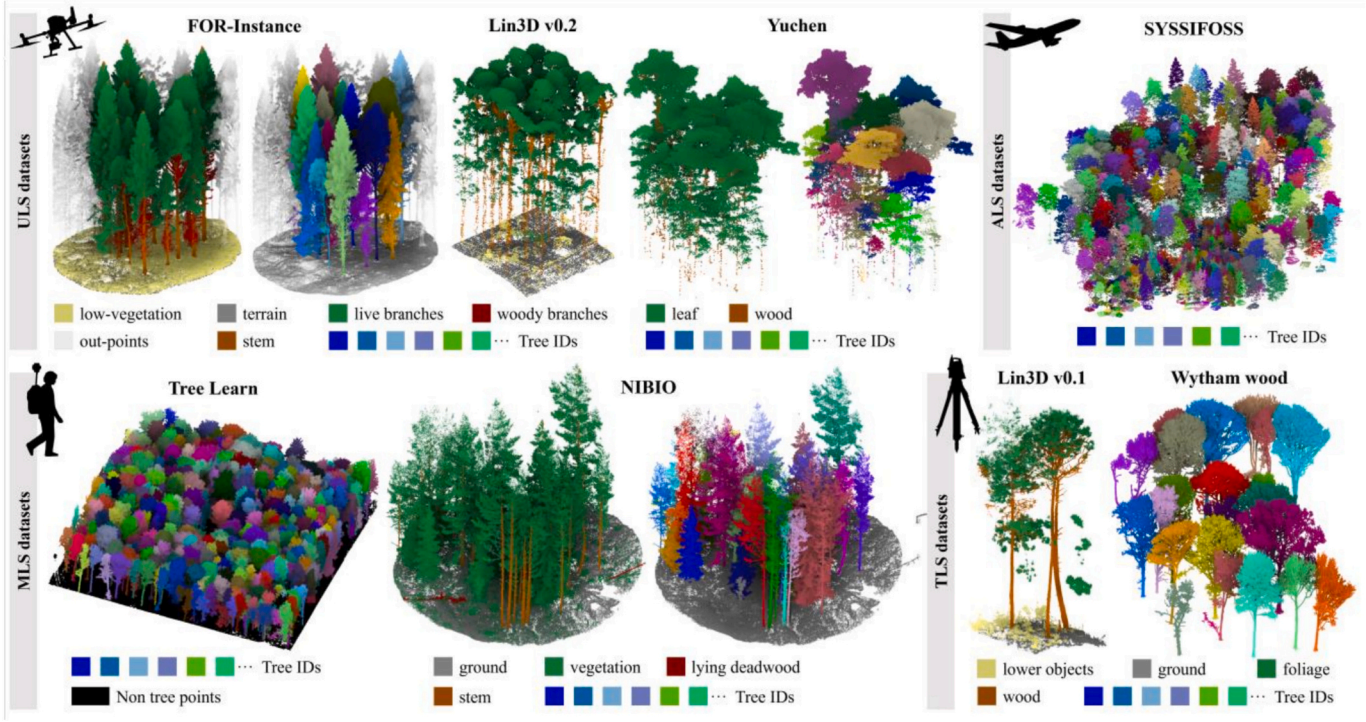

Fig. 4. Visualization of real-world forest point cloud datasets with their semantic or instance labels.

deadwood (Table 1), which is not the focus of our analysis and was therefore excluded. Additionally, following Liu et al. (2026), we merged the categories *stem* and *woody branches* in the FOR-Instance dataset into a unified semantic class labeled as *wood,* as the synthetic dataset does not distinguish between *stem* and *woody branches*. Furthermore, semantic labels across different datasets, including Lin3D, FOR-Instance, and NIBIO, exhibited inconsistent terminologies. The categories *low-vegetation*, *lower object*, and *understory* were semantically overlapping. We systematically standardized these labels to match the Boreal3D semantic categories. Specifically, the *low-vegetation* category in the FOR-Instance dataset and the *lower objects* category in the Lin3D dataset were both mapped to *understory*; the *live branches* category in FOR-Instance and the *foliage* category in Lin3D were both mapped to *leaf*; and the *ground* category in the Lin3D and NIBIO datasets was unified as *terrain*. Additionally, semantic labels 0 to 3 in the synthetic dataset correspond to *understory*, *terrain*, *leaf*, and *wood*, respectively. To ensure consistency across datasets, we adjusted the label mappings in the other datasets to align with this convention.

3.2. Uncovering the negative transfer in multi-platform forest data

In forest scenarios, negative transfer arises prominently from the heterogeneous feature distributions of multi-platform LiDAR acquisition datasets. Point clouds from different acquisition platforms exhibit pronounced structural and density variations: MLS captures dense, detailed tree stem structures due to its proximity to the ground, but lacks canopy coverage. ULS provides comprehensive canopy coverage but struggles to resolve stem details due to nadir-oriented scanning angles. ALS collects data from higher altitudes, suffers from sparse point density, and has an incomplete representation of both stems and canopy. These platform-

Table 2
Instance segmentation and semantic segmentation results on the virtual synthetic dataset (Boreal3D) under different settings. The underlined numbers indicate that the training and testing are on the same platform dataset. The bold numbers indicate the best performance. Green numbers indicate improved performance compared to testing and training on the same platform dataset. Red numbers indicate decreased performance compared to testing and training on the same platform dataset. Unit: %.

Model	Test dataset	Instance segmentation				Semantic segmentation		
		Completeness	Omission	Commission	F1-score	mIoU	mAcc	oAcc
Model 1 (Trained on ALS dataset)	ALS	41.64	58.34	3.56	58.16	57.14	66.55	86.89
	ULS	0.00	100.00	100.00	0.00 (−58.16)	6.35 (−50.79)	22.72	15.57
	MLS	0.00	100.00	100.00	0.00 (−58.16)	7.60 (−49.54)	24.84	28.97
Model 2 (Trained on ULS dataset)	ALS	16.04	89.96	33.10	25.88 (−74.53)	22.06 (−58.57)	43.51	59.05
	ULS	97.45	2.55	2.63	97.41	80.63	84.98	94.12
	MLS	49.07	50.93	87.88	19.44 (−77.97)	23.10 (−57.53)	38.66	42.69
Model 3 (Trained on MLS dataset)	ALS	6.74	93.26	41.91	12.08 (−77.24)	19.71 (−65.52)	27.22	70.46
	ULS	68.42	31.58	28.29	70.03 (−19.29)	28.90 (−53.33)	47.64	74.95
	MLS	84.47	15.53	5.24	89.32	82.23	89.49	91.43
Model 4 (Trained on mixed dataset)	ALS	8.11	91.89	35.81	14.39 (−43.77)	50.46 (−6.68)	60.69	85.35
	ULS	81.15	18.85	25.72	77.57 (−19.84)	71.16 (−9.47)	76.66	91.54
	MLS	68.00	32.00	15.77	75.25 (−14.07)	74.26 (−7.97)	83.58	85.37
Model 5 MST (ours)	ALS	43.86	56.14	0.96	**60.79** (+2.63)	**60.12** (+2.98)	69.58	87.15
	ULS	98.39	1.61	0.43	**98.98** (+1.57)	**83.82** (+3.19)	86.99	95.70
	MLS	97.88	2.12	9.57	**94.01** (+4.69)	**87.13** (+4.90)	91.20	94.49

specific disparities create conflicting feature spaces during multi-source training, resulting in performance degradation across all platforms. As shown in Table 2 (models 1–4), models trained on mixed multi-platform datasets underperform when tested on platform-specific data, particularly in instance segmentation tasks, where accuracy declines sharply.

To investigate the underlying mechanisms of negative transfer in multi-platform forest point cloud analysis, we systematically analyzed the feature representations learned by a deep neural network, following Chen et al. (2024a), Huang et al. (2023), and Lu and Deng (2025). In this study, the synthetic dataset (Boreal 3D) was adopted to demonstrate negative transfer, since it contains point clouds of the same virtual forest scenes acquired under different platform configurations. Employing this dataset eliminates confounding factors unrelated to the platform and controls platform heterogeneity as the sole variable of interest, thereby enabling a more rigorous and interpretable evaluation. Specifically, we extracted feature embeddings from the penultimate layer (the layer preceding the final classification output) during forward inference across varying training and test settings. For semantic segmentation, we extracted point features from the layer preceding the final semantic classifier and grouped them by semantic class. For each class, the features from all points were concatenated and then aggregated by average pooling to obtain a high-dimensional class-level embedding. For the instance segmentation, point features were extracted from the layer immediately preceding the instance prediction head for all tree instances on each acquisition platform dataset. The features were concatenated and aggregated to produce a high-dimensional feature embedding. These feature embeddings effectively encapsulate information captured by the preceding layers. We fed multi-platform forest point cloud data $X \in \{X^{ALS}, X^{ULS}, X^{MLS}\}$ into models trained on individual and mixed platform datasets $\mathscr{F} \in \{\mathscr{F}^{ALS}, \mathscr{F}^{ULS}, \mathscr{F}^{MLS}, \mathscr{F}^{mixed}\}$, extracting feature representations for all data points. These high-dimensional features were subsequently reduced to two dimensions using t-SNE (Van der Maaten and Hinton, 2008) for visualization. The two-dimensional feature embeddings were normalized and presented in a unified feature space. As shown in Fig. 5(a), the semantic segmentation visualization results clearly illustrate distinct patterns of feature distributions across platforms. Under single-platform training and testing conditions, features from different semantic categories formed compact, well-defined clusters in the feature space. Conversely, feature distributions derived from multi-platform mixed training exhibit significant dispersion and overlap among category clusters. This phenomenon demonstrates that platform heterogeneity introduces negative transfer during joint training, as the model fails to learn unified and discriminative feature representations across semantic categories. Additionally, this pattern is further evidenced in instance segmentation tasks, where multi-platform forest point cloud data exhibit significant feature overlap in jointly trained models, as illustrated in Fig. 5(b). Such feature space overlap creates ambiguous decision boundaries, impeding the model capacity to discern unified instance-level patterns across multi-platform datasets.

3.3. Semantic segmentation network

For semantic segmentation of forest point clouds, we utilize Point Transformer V3 (PTv3) (Wu et al., 2024a) as the backbone. Architecturally, PTv3 adopts a U-Net backbone with grid pooling and batch normalization, ensuring reliable feature extraction across vertically stratified forest environments and complex terrains. The PTv3 pipeline is illustrated in Fig. 6.

Traditional semantic segmentation approaches for forest scenes often struggle to capture multi-scale structural features, particularly because of the irregular distribution and varying densities of points representing forest elements. PTv3 addresses these challenges by transforming unstructured forest point clouds into serialized sequences using a novel space-filling curve-based serialization strategy, including *Z*-order and Hilbert curves. This serialization method significantly reduces the computational burden of neighbor-searching operations, such as K-Nearest Neighbors, enabling efficient modeling of spatial relationships

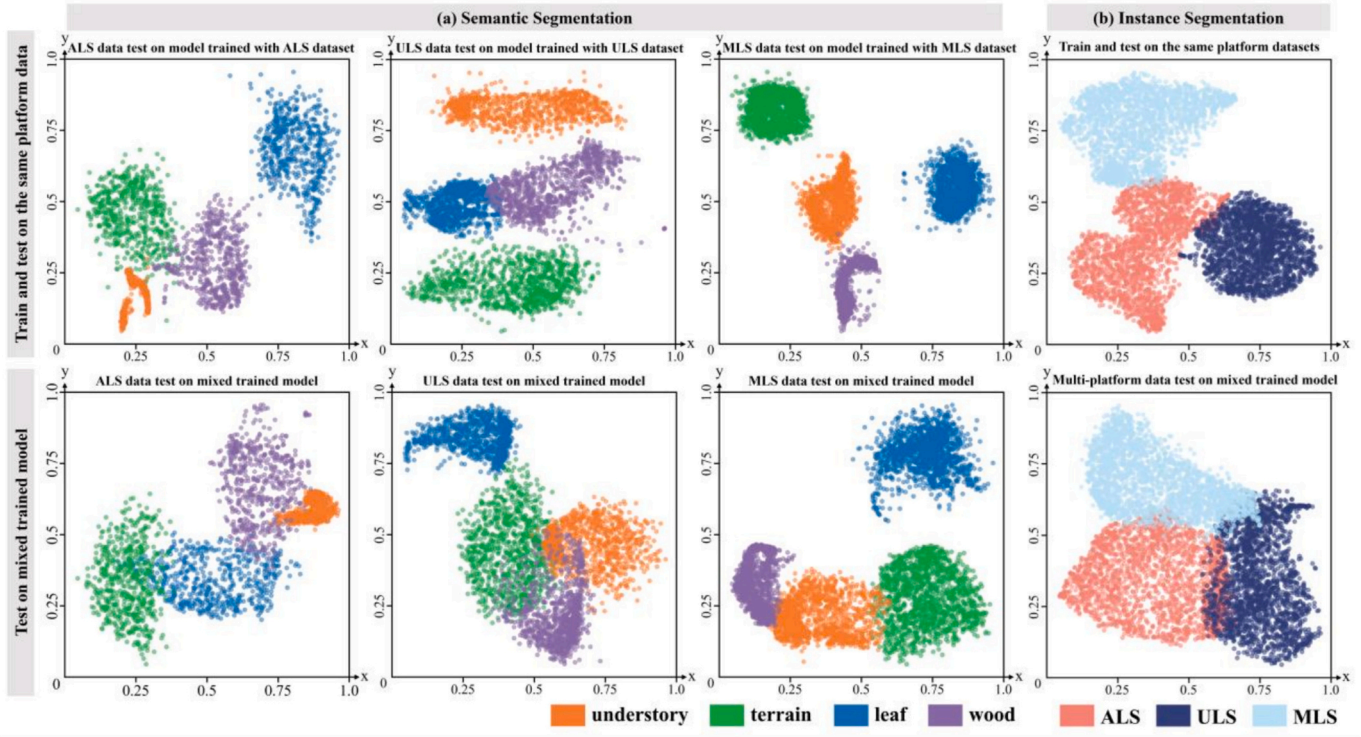

Fig. 5. Feature distribution of semantic and instance segmentation tasks across train-test settings. Each point in the plots represents high-dimensional feature vectors projected into a unified 2D feature space using t-SNE. (a) Distribution of different semantic categories (*understory, terrain, leaf, wood*) in the feature space under different train-test settings. (b) Distribution of different platform point cloud data (ALS, ULS, MLS) in the feature space under different train-test settings.

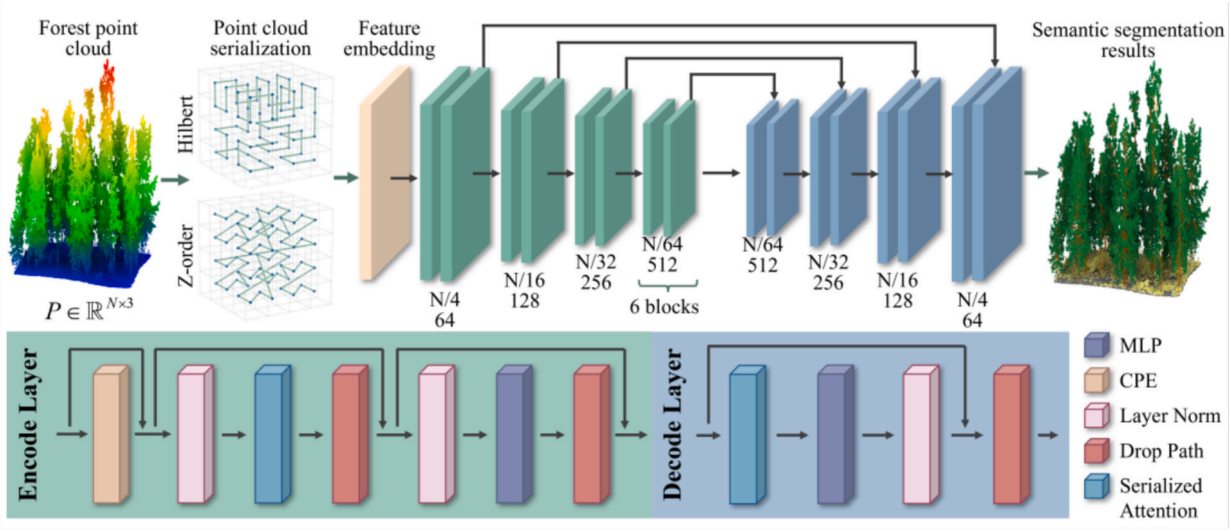

Fig. 6. The overall framework of the Point Transformer V3 for forest scene semantic segmentation.

even in dense, complex forest scenes. Furthermore, serialization helps preserve spatial proximity and multi-scale neighborhood contexts, effectively capturing structural variations within forest canopies and understory layers. Dynamic patch interaction strategies (Shift Order, Shuffle Order) cyclically permute serialization patterns across attention layers, enabling adaptive perception of multi-layered forest structures. This mechanism expands the receptive field, effectively capturing hierarchical relationships between forest elements. Complementing this, the enhanced Conditional Positional Encoding integrates sparse convolutional layers to bypass the inefficiencies of relative positional encoding, thereby further encoding relative spatial information critical for distinguishing semantically similar classes, such as branches and foliage. Additionally, the Lovasz Hinge Loss (Berman et al., 2018) is adopted to optimize semantic segmentation during implementation, specifically addressing the challenges posed by class imbalance and the inherent complexity of forest point clouds. Lovasz Hinge Loss computes a convex surrogate approximation of the IoU metric, thereby providing stable gradients and improved training convergence, particularly beneficial for minority classes commonly observed in forest structures (e.g., stem and understory). Mathematically, the Lovasz Hinge Loss is formulated as Eqs. (1) and (2).

$$\mathscr{L}_{LovaszHinge} = \sum_{c=1}^{C} \mathscr{L}^{c}_{Lovasz}(m(x)), \tag{1}$$

$$\mathscr{L}^{c}_{Lovasz} = \sum_{i=1}^{|m|\downarrow} \Delta\mathscr{J}_{c}\left(M^{c}_{i}\right)\cdot m_{\pi i}, \tag{2}$$

where C represents the total number of semantic classes in the forest environment, and $m(x)$ denotes the vector of errors for each point with respect to class c. $m_{\pi i}$ represents the sorted errors in descending order, $\Delta\mathscr{J}_{c}\left(M^{c}_{i}\right)$ corresponds to the discrete gradient of the Jaccard index associated with submodular set functions, and M^{c}_{i} denotes the set of misclassified points up to index i. More detailed information can be found in (Wu et al., 2024a).

3.4. Instance segmentation network

The instance segmentation framework employs Point Group (Jiang et al., 2020), a bottom-up architecture optimized for discerning individual tree instances in complex forest point clouds. The network operates through four synergistic stages: feature extraction, dual-coordinate clustering, probabilistic instance scoring, and duplicate suppression.

Initially, a U-Net backbone processes raw LiDAR point clouds to extract multi-scale geometric features while preserving fine-grained details critical for distinguishing ecologically significant structures. In this study, we utilize PTv3 as the backbone due to its strong feature extraction ability. Parallel semantic and offset prediction branches then generate per-point classifications (tree or non-tree) and offset vectors directing points toward their respective tree centroids. Offset learning is crucial for separating overlapping tree crowns, as neighboring trees are often wrongly grouped due to canopy proximity. Subsequently, the dual-coordinate clustering algorithm operates on both raw and offset-shifted point sets. Clustering on raw coordinates groups proximate points with matching semantic labels, effectively isolating understory vegetation and coarse woody debris as separate instances. Simultaneously, offset-based clustering capitalizes on predicted centroid convergence to resolve ambiguities in dense stands, particularly critical for trees with overlapping crowns. Following clustering, a dedicated scoring module, ScoreNet, evaluates each candidate cluster to determine its quality, thereby facilitating more reliable separation of neighboring trees that share similar structures or heights. Finally, non-maximum suppression (NMS) eliminates redundant proposals, ensuring each tree instance corresponds to a single cluster. The overall framework of the Point Group is illustrated in Fig. 7.

During implementation, a composite loss function is employed to optimize the network, integrating multiple individual losses designed to address specific segmentation challenges. The total loss function is formulated as Eq. (3):

$$\mathscr{L} = \mathscr{L}_{sem} + \mathscr{L}_{o\_reg} + \mathscr{L}_{o\_dir} + \mathscr{L}_{c\_score} \tag{3}$$

Specifically, the semantic segmentation loss $\mathscr{L}_{sem}$ utilizes cross-entropy loss to ensure accurate point-level semantic classification, as illustrated in Eq. (4).

$$\mathscr{L}_{sem} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(\hat{y}_{i,c}\right), \tag{4}$$

where $y_{i,c}$ and $\hat{y}_{i,c}$ represent the ground-truth and predicted class probabilities for point i, C and N denote the total number of points and classes, respectively.

The offset regression loss $\mathscr{L}_{o\_reg}$ ensures predictions converge accurately toward the true instance centroids, formulated as an L1 regression loss:

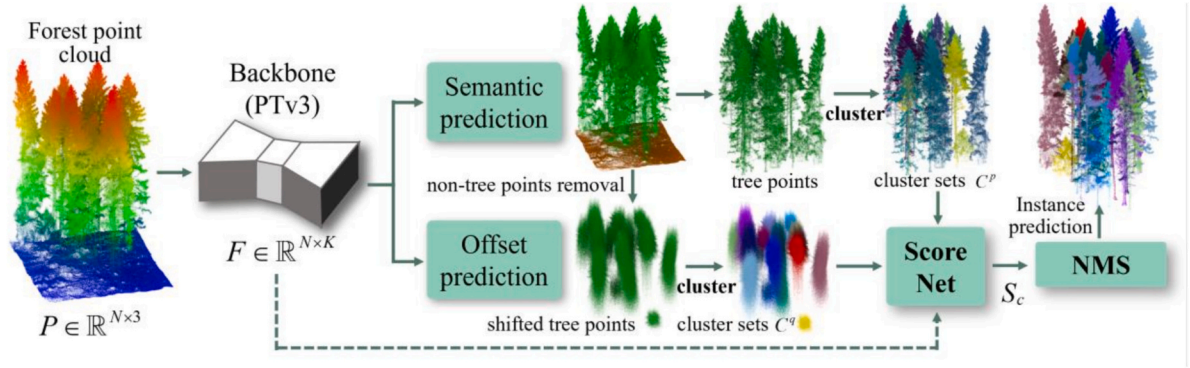

Fig. 7. The overall framework of the PointGroup for forest scene instance segmentation.

$$\mathscr{L}_{o\_reg} = \frac{1}{\sum_i m_i}\sum_i \|o_i - (\hat{c}_i - p_i)\| \cdot m_i, \tag{5}$$

where o_i is the predicted offset vector, p_i denotes the original point coordinates, $\hat{c}_i$ is the centroid coordinates of the corresponding ground-truth instance, and m_i represents a binary mask indicating instance points.

The offset direction loss $\mathscr{L}_{o\_dir}$ (Lahoud et al., 2019) further stabilizes centroid prediction by constraining the directionality of offset vectors, which is formalized as Eq. (6).

$$\mathscr{L}_{o\_dir} = -\frac{1}{\sum_i m_i}\sum_i \frac{o_i}{\|o_i\|}\cdot\frac{\hat{c}_i - p_i}{\|\hat{c}_i - p_i\|}\cdot m_i. \tag{6}$$

Furthermore, the ScoreNet loss $\mathscr{L}_{c\_score}$ adopts binary cross-entropy to evaluate and refine the quality of instance cluster proposals:

$$\mathscr{L}_{c\_score} = -\frac{1}{M}\sum_{i=1}^{M}[\hat{s}_i \log(s_i) + (1-\hat{s}_i)\log(1-s_i)], \tag{7}$$

where $\hat{s}_i$ and s_i represent the ground-truth and predicted scores for cluster i, respectively, and M is the total number of candidate clusters.

3.5. *Multi-platform synergistic training framework*

Traditional data-driven algorithms often fail to generalize across multi-platform forest datasets due to discrepancies in platform-specific feature distributions. To address the inherent heterogeneity and negative transfer challenges arising from multi-platform forest point cloud datasets, we introduce a model and data-driven representation learning framework, termed MST. Specifically, a set of learnable CPATs and a plug-and-play CIM are introduced in MST, enabling adaptive feature alignment while preserving platform-invariant tree representations (Wu et al., 2024b), as shown in Fig. 8. In MST, CPATs are implemented as platform-specific conditional vectors, while CIM injects such platform context into the shared backbone through conditional normalization and feature modulation.

For each platform-specific dataset $D_i \in \{D_{ALS}, D_{ULS}, D_{MLS}\}$, a set of learnable d-dimension CPATs are embedded as conditional vectors to encode platform-specific distributional shifts, where CPAT embeddings are denoted as $P = \{p_i \in \mathbb{R}^d | 1 \le i \le n\}$. These embeddings are jointly optimized end-to-end to model distributional differences across platforms explicitly. By dynamically injecting platform-specific contextual information into the backbone network, CPAT facilitates adaptive differentiation between inherent geometric structures, topological relationships intrinsic to individual trees, and platform-dependent attributes. Each platform domain is associated with its own CPAT, and the corresponding conditional vector is selected according to the platform identity of the input sample. The optimization objective minimizes cumulative loss across datasets by jointly training CPATs and backbone parameters θ, formulated as Eq. (8):

$$\underset{\theta,P}{\operatorname{argmin}} \sum_{i=1}^{n}\frac{1}{D_i}\sum_{(x_i,y_i)\in D_i}\mathscr{L}(f(x_i,p_i;\theta),y_i), \tag{8}$$

where p_i conditions the model on platform-specific contexts, $f(\cdot)$ denotes model outputs, and y_i denotes the corresponding ground truth of the dataset D_i.

To further alleviate disparities in feature distribution across platforms, MST incorporates a CIM to construct a coherent, unified representation space that generalizes across heterogeneous datasets. CIM approaches the geometric discrepancies across datasets as a domain shift challenge and employs a dynamic feature normalization mechanism to implicitly align and enhance features from multiple platforms. Specifically, CIM generates platform-adaptive scaling $\alpha(p)$ and shifting $\beta(p)$ factors derived from CPAT embeddings to perform an affine transformation on normalized feature vectors, expressed as Eq. (9):

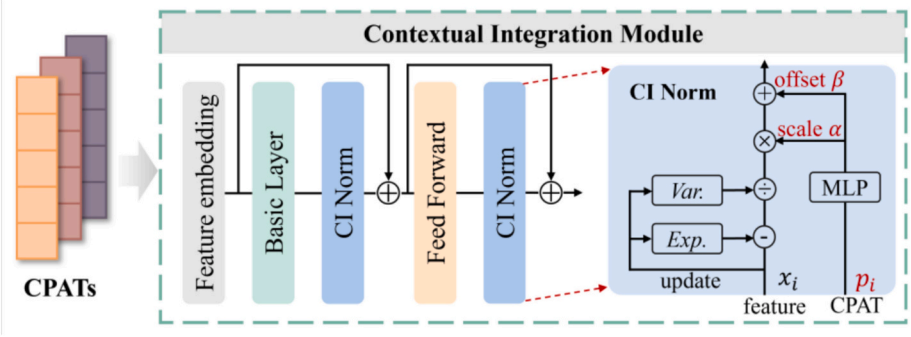

Fig. 8. Diagram of contextual integration module (CIM).

$$CIM(x,p) = \frac{x - E[\overline{x}]}{\sqrt{Var[\overline{x}] + \eta}} \cdot \alpha(p) + \beta(p), \tag{9}$$

where $\alpha(p), \beta(p)$ are linear projections of CPATs, $E[\overline{x}]$ and $Var[\overline{x}]$ are the mean and variance of the input feature x, η is a smoothing term used to avoid divide-by-zero errors. Under this formulation, the conditioning is applied to the normalization layers and the normalized intermediate features within the shared backbone. Feature statistics are first standardized through normalization, after which platform-dependent affine parameters derived from CPATs modulate the normalized responses. This adaptive mechanism explicitly modulates feature distributions without incurring substantial additional parameter overhead, thus effectively facilitating unified representation learning within the shared backbone network. Unlike conventional methods that require extensive parameter tuning, CIM seamlessly integrates with existing normalization layers in the backbone, dynamically embedding cross-platform contextual information to enhance model generalizability. The resulting cross-platform alignment is formed implicitly through shared backbone learning and conditional normalization, without introducing an additional explicit alignment objective.

Accurate classification of forest point clouds requires synergistic modeling of both local geometric details (e.g., individual tree branching structures) and global contextual patterns (e.g., stand-level density distributions). To address this, the MST framework strategically integrates platform-aware context at each stage of the backbone network. Shallow layers focus on platform-invariant geometric primitives (e.g., trunk curvature, branch orientation) through localized feature alignment, establishing stable cross-domain baselines. Deeper layers progressively aggregate global contextual patterns (e.g., stand density, crown morphology) by fusing CPAT-modulated high-order semantics, thereby constructing discriminative cross-platform representations. This staged fusion mechanism ensures granular adaptation to platform-specific sensing characteristics while preserving ecological consistency. Furthermore, MST adopts an initialization strategy where CIM scaling and shifting parameters are initially set to zero to prevent early-stage randomness from perturbing established representation learning patterns. This approach ensures a stable initiation of universal cross-platform feature learning. Specifically, CPATs are trained jointly with the backbone network from scratch. If the CPATs were randomly initialized and injected directly into the backbone, they would strongly disturb the initial feature distribution, leading to unstable optimization and oscillation in the loss during early training. In contrast, a zero-initialization strategy stabilizes the optimization process, resulting in better convergence and final performance. Furthermore, zero initialization ensures the CPATs initially contribute negligible scaling and shifting signals. This allows the backbone to first concentrate on learning generalized forest structural representation before CPATs progressively inject platform-specific adaptations. This approach enhances the stability of the learned features. Moreover, the CIM normalizes feature statistics independently for each platform and then modulates them through the corresponding CPATs. This design effectively absorbs platform-specific variations within the normalization layer, allowing the backbone to focus on capturing platform-invariant structural content. Consequently, MST achieves stable cross-platform training and delivers consistent segmentation performance across heterogeneous forest datasets. Furthermore, during the initial training phases, CPAT and CIM parameters employ relatively lower learning rates to maintain a primary focus on backbone network optimization. Learning rates for these modules gradually increase as training progresses, progressively guiding the model to incorporate subtle platform-specific adjustments without compromising previously acquired generalized representations. As MST operates at the backbone level, the same training mechanism can be directly coupled with both semantic and instance segmentation networks without redesigning their task-specific prediction heads. In the instance segmentation setting, the original PointGroup pipeline, including the semantic branch, offset branch, dual-coordinate clustering, ScoreNet, and NMS, remains unchanged, while MST strengthens the shared backbone representation used by these components.

Additionally, Liu et al. (2026) demonstrated performance improvements by pre-training models on synthetic datasets, followed by fine-tuning on real-world data. Inspired by this, we adopt a similar strategy in our framework by pre-training the MST framework on synthetic multi-platform forest point cloud datasets and subsequently fine-tuning on real-world datasets.

3.6. Evaluation metrics

For semantic segmentation, performance was assessed using three widely adopted class-aggregated metrics, including Mean intersection over union (mIoU), Mean accuracy (mAcc), and overall accuracy (oACC). These metrics are mathematically defined as Eqs. (10)–(12):

$$mIoU = \frac{1}{n}\sum_{i=0}^{n-1}\frac{TP_i}{TP_i + FP_i + FN_i}, \tag{10}$$

$$mACC = \frac{1}{n}\sum_{i=0}^{n-1}\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}, \tag{11}$$

$$oACC = \frac{\sum_{i=0}^{n-1} TP_i + TN_i}{\sum_{i=0}^{n-1} TP_i + TN_i + FP_i + FN_i}, \tag{12}$$

where TP_i, TN_i, FP_i, and FN_i denote true positive, true negative, false positive, and false negative of class i, respectively. n represents the number of total semantic categories.

For instance segmentation, we followed Wielgosz et al. (2024) and Xiang et al. (2024), employed the completeness, omission error, commission error, and F1-score for instance segmentation evaluation, as formalized in Eqs. (13)-(18), where N denotes the number of instance trees in the test data, R and P represent recall and precision. Specifically, a confusion matrix was constructed to determine the counts of TP, FP, and FN, where predicted instances were classified as TP if their IoU with the corresponding ground truth instances exceeded a threshold of 0.5 (Wielgosz et al., 2024; Xi and Hopkinson, 2022; Xiang et al., 2024; Xiang et al., 2025b). The completeness denotes the proportion of correctly identified ground truth instances, while the omission rate represents the fraction of undetected ground truth instances. The commission Error is the ratio of spurious predictions to the total number of predicted instances, and the F1-score is the harmonic mean of precision and recall, quantifying the balance between detection accuracy and sensitivity.

$$Completeness = \frac{TP}{N}, \tag{13}$$

$$Omissionerror = \frac{FN}{N}, \tag{14}$$

$$Commissionerror = \frac{FP}{TP + FP}, \tag{15}$$

$$R = \frac{TP}{TP + FN}, \tag{16}$$

$$P = \frac{TP}{TP + FP}, \tag{17}$$

$$F1 - score = \frac{2R \cdot P}{R + P}. \tag{18}$$

3.7. Implementation details

The model was implemented on a workstation equipped with a vCPU Intel(R) Xeon(R) Platinum 8457C with 20 cores and an NVIDIA L20 (48 GB RAM), operating under Ubuntu 22.04. The model computations were performed using PyTorch 2.4.1 in Python. All programming tasks were conducted within a virtual environment established with Miniconda, which included CUDA 12.4 and cuDNN 9.1.0, and Python 3.10. More detailed configuration settings are available in Table S1 of Appendix S1. The code for the proposed representation learning framework is available at: https://github.com/jdjiang312/MST.

4. Experimental results

4.1. Semantic and instance segmentation results on the virtual synthetic dataset

To systematically validate the effectiveness of the MST framework, extensive experiments were conducted on the virtual synthetic dataset (Boreal3D). Specifically, five models were trained under distinct configurations: Models 1, 2, and 3 (platform-specific models) were trained exclusively on ALS, ULS, and MLS forest datasets, respectively; Model 4 employed a conventional mixed training framework combining all platform datasets; and Model 5 employed the MST framework. All models were optimized with identical hyperparameter settings to ensure a fair comparison, and performance was evaluated using optimal checkpoints.

Quantitative and qualitative results of the instance segmentation task are summarized in Table 2 and Fig. 9. Models trained and evaluated on the same platform consistently demonstrated satisfactory performance. Specifically, the model trained and evaluated on ULS point clouds achieved the highest detection rate of 97.45%. In contrast, the MLS-trained model exhibited a marginally lower detection rate, which can be attributed to incomplete canopy points. The lowest performance was observed with the ALS-trained model (about 0.6 in F1-score), a consequence of extreme sparsity and structural incompleteness in both the canopy and stem regions, which impede spatial feature discrimination. Cross-platform generalization tests revealed substantial performance degradation. Particularly, the model trained on the ULS dataset exhibited an over 70% reduction in F1-score when evaluated on MLS and ALS datasets. Notably, the model trained on ALS data (Model 1) exhibited markedly poor segmentation performance when evaluated on ULS and MLS datasets. This decline underscores the pronounced heterogeneity across datasets from different platforms, severely limiting the generalization capability of platform-specific models. Furthermore, results from the Model 4 trained using the mixed platform datasets validated the negative transfer across different platform forest data. Compared to platform-specific models, the mixed-training approach experienced varying degrees of accuracy deterioration. This degradation stems from implicit feature distribution discrepancies among platforms, which encourage models to learn platform-biased representations rather than domain-invariant features. In contrast, the MST framework (Model 5) consistently improved performance across all three platforms. We note that some commission errors in Table 2 are close to zero, particularly for the ULS data. This mainly reflects the conservative proposal generation and filtering strategy of Point Group, together with the relatively clean and structurally separable characteristics of some plots, which suppress false instance predictions.

For the semantic segmentation task, the quantitative and qualitative results are illustrated in Fig. 10 and Table 2. Specifically, platform-specific models (Models 1–3) demonstrate competitive performance within their respective datasets but exhibit limited generalization when evaluated on other platform datasets. Cross-platform evaluations revealed up to 65% declines in mIoU, indicating these models predominantly learned platform-dependent features rather than domain-invariant representations. The conventional mixed-training framework (Model 4) partially mitigated domain gaps, improving generalization capability modestly over platform-specific models. However, its performance remained substantially inferior to platform-specific baselines with mIoU reductions of 6.68 (ALS), 9.47 (ULS), and 7.97 (MLS). This

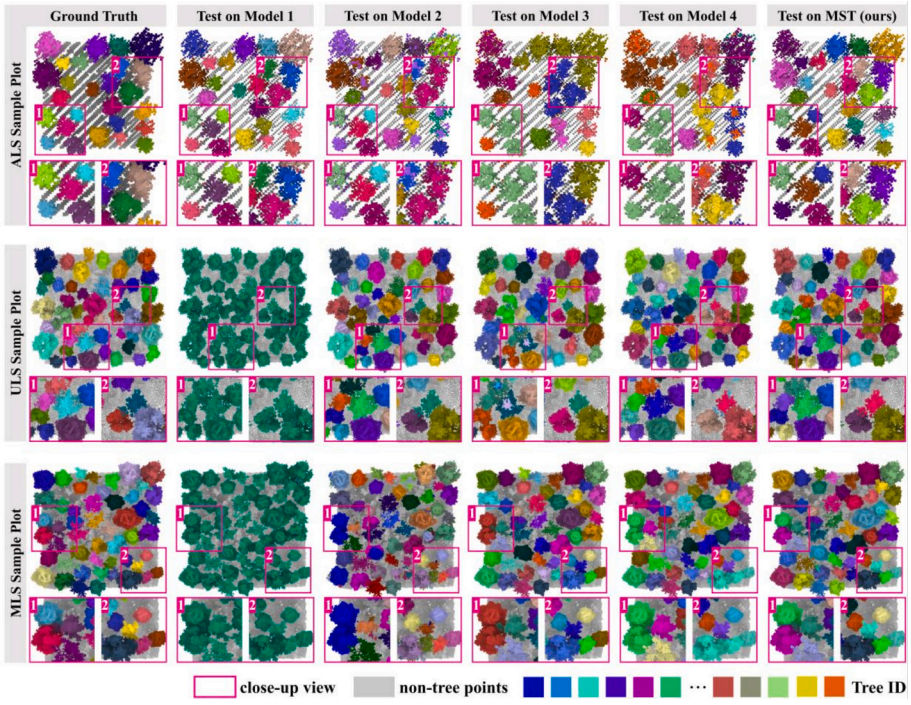

Fig. 9. Visualization results of instance segmentation for the virtual synthetic dataset (Boreal3D). The forest point cloud data from each platform were evaluated using five models, with results presented from two perspectives: a top-down view (above) and a close-up view (below). Model 1, Model 2, and Model 3 represent models trained on ALS, ULS, and MLS datasets, respectively, while Model 4 was trained on the mixed dataset (multi-platform dataset). Model 5 represents the MST framework. Each color represents an individual tree.

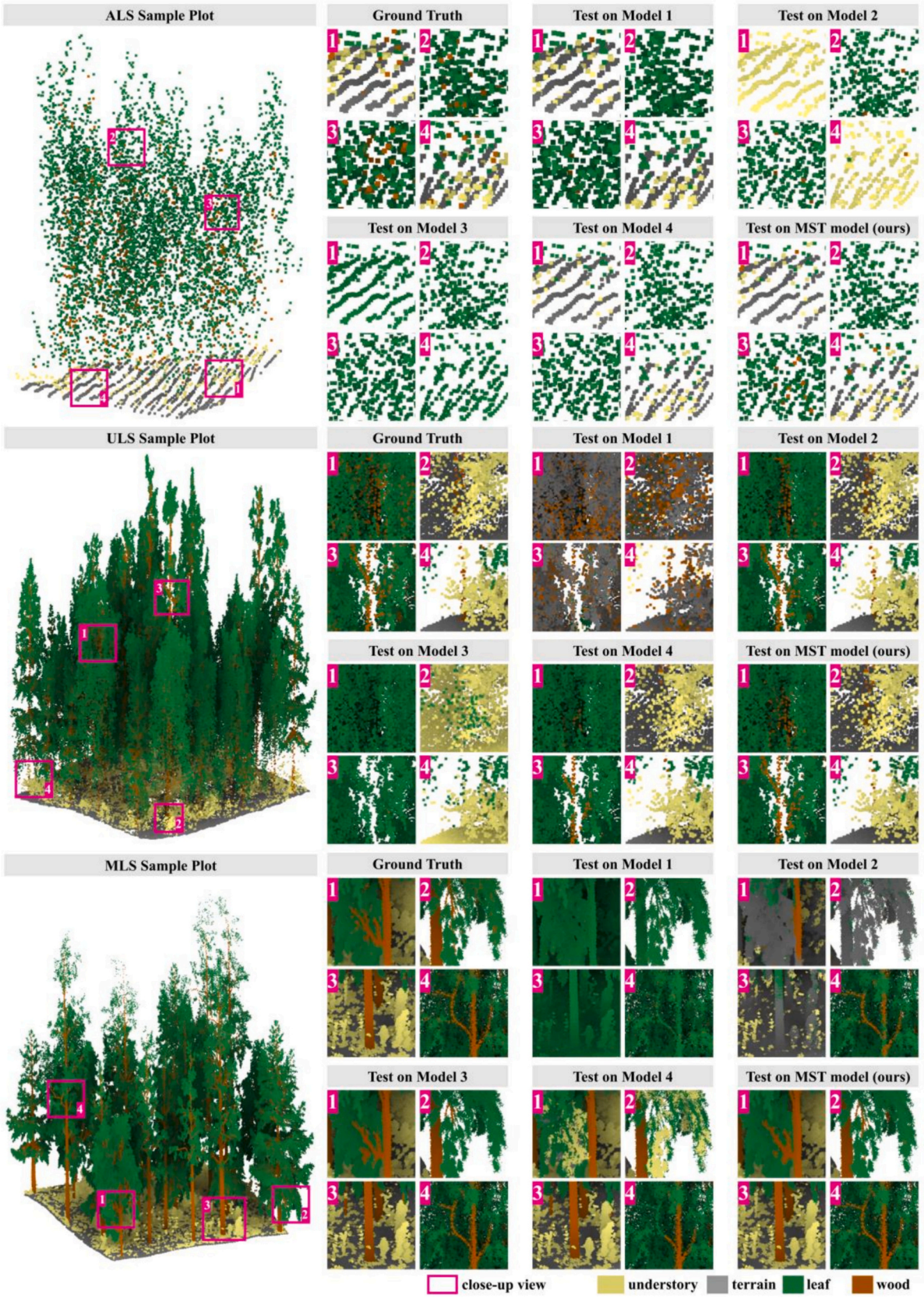

Fig. 10. Visualization results of semantic segmentation for the virtual synthetic dataset (Boreal3D). The forest point cloud data from each platform was evaluated using five models. Model 1, Model 2, and Model 3 represent models trained on ALS, ULS, and MLS datasets, respectively, while Model 4 was trained on the mixed dataset (multi-platform dataset). Model 5 represents the MST framework.

degradation confirms that negative transfer persists, as heterogeneous feature distributions across platforms induce conflicting gradient signals during joint optimization. Despite sharing identical sample plots during simulating point cloud data by different acquisition platforms, the implicit feature distribution gap arising from varying sensor configurations and occlusion patterns prevents mixed-training models from learning unified feature representations. Conversely, employing the MST training framework achieved consistent cross-platform improvements, surpassing platform-specific models by 3–5% in mIoU while maintaining superior generalization. This enhancement stems from the capacity to disentangle cross-platform semantic features from domain-specific features. Through the contextual integration mechanism, the framework establishes a shared feature space in which ecological semantics are preserved while platform-induced variations are explicitly suppressed.

4.2. Semantic and instance segmentation results on real-world dataset

We conducted a comprehensive evaluation of the MST framework on multiple platform forest point cloud benchmarks to assess its performance in both semantic and instance segmentation tasks. For semantic segmentation, FOR-Instance, Lin3D, Yuchen, and NIBIO were selected due to their provision of semantic annotations. Notably, the MST framework defines four semantic categories, including *understory*, *terrain*, *leaf*, and *wood*, whereas FOR-Instance further partitions *wood* into *woody-branches* and *stem*. To ensure consistency, we merged the subdivided labels to align with the four-class semantic categories during evaluation. The quantitative results, as shown in Table 3, reveal that MST exhibited consistent generalization across both ULS and MLS acquisition platform datasets. More than 60% of the test plots achieved mIoU values exceeding 85%, indicating that MST effectively captures cross-platform feature representations. At the per-class level, the *leaf* category attained the highest IoU metrics, reflecting the model proficiency in delineating fine-scale foliage geometry. In contrast, the *wood* category demonstrated comparatively lower performance, which can be attributed to the intrinsic morphological heterogeneity and diffuse spatial distribution of woody elements, as illustrated in Fig. 11. A limited portion of the wood points was misclassified as leaf, primarily attributable to the inherent spatial intertwining between branches and surrounding foliage. Moreover, notable confusion between understory and terrain classes was observed, reflecting the inherent challenges of characterizing structurally diverse understory vegetation against the topographic complexity of terrain surfaces. This issue arises from ecological factors, as understory vegetation frequently manifests heterogeneous morphological attributes influenced by varying levels of illumination, moisture conditions, and local microhabitats, complicating accurate feature extraction and class delineation.

Additionally, a more detailed analysis of the FOR-Instance dataset (Appendix S2 Fig. A1 and Table 3) shows an overall mIoU of 78.37%. The CULS and SCION test plots achieved the highest accuracies, attributable to their relatively homogeneous species composition and higher point densities, which reduce ambiguity at branch–leaf interfaces. In contrast, the NIBIO and RMIT plots exhibited reduced performance. NIBIO suffered *terrain-understory* confusion owing to the spatial proximity and feature similarity of these classes, and the sparse point density in the RMIT plot hindered accurate *stem* detection. The segmentation results for the Lin3D dataset are illustrated in Appendix S2 Fig. A2 and Table 3. The results reveal that MST achieved consistently high segmentation accuracy across all tree components, with only minor misclassification occurring between *leaf* and *wood* points. A more challenging site is the Yuchen dataset, collected in a tropical forest characterized by a dense canopy and pronounced morphological heterogeneity across species. Despite these challenges, the MST framework demonstrated effective semantic segmentation performance, achieving an IoU of 98% for the leaf class (Appendix S2 Fig. A3 and Table 3). However, the IoU for the wood class was limited to 60%. This is likely due to the favorable growth conditions of tropical forests, which promote high canopy closure. This leads to occlusion and a scarcity of trunk and branch points in the mid- and lower canopy layers. These data characteristics hinder a complete geometric representation of the wood class, leading to semantic boundary confusion with the leaf class. Additionally, in the NIBIO dataset (Appendix S2 Fig. A4 and Table 3), the lack of detailed annotations between *understory* and *leaf* classes allowed the model to learn generalized features for the merged *vegetation* category, resulting in a high mIoU of 90.42%.

For the individual tree segmentation task, we evaluated the generalized cross-platform segmentation capability of the MST on the FOR-Instance, Yuchen, TreeLearn, NIBIO, and SYSSIFOSS benchmark datasets. The quantitative results, detailed in Table 4, demonstrate that MST consistently delivers consistent segmentation performance across diverse forest point cloud datasets acquired from various platforms. Notably, despite generally accurate delineation at the individual-tree level, a minor occurrence of over-segmentation was observed. This phenomenon prominently emerged in datasets derived from ULS and ALS, as exemplified by the RMIT and TU Wien plots in the FOR-Instance dataset (Appendix S3 Fig. A5) and the SYSSIFOSS dataset (Appendix S3 Fig. A9). This over-segmentation primarily arises from inherently lower point densities and noise associated with occlusions characteristic of ULS and ALS acquisition techniques.

Specifically, although most sample plots within the FOR-Instance dataset yielded satisfactory segmentation outcomes (Appendix S3 Fig. A5), the TU Wien plot constituted a pronounced deviation from this overall trend. The TU Wien plot exhibited significant segmentation challenges due to its complex, unmanaged, multi-layered, mixed-deciduous forest structure, with substantial canopy overlap. Such complex canopy interactions make it difficult to distinguish among adjacent tree crowns, thereby adversely affecting segmentation fidelity. Additionally, the MST achieved a high completeness of 82.5% in the Yuchen dataset. However, visual inspection revealed instances of under-segmentation, where multiple large trees were grouped into a single instance. This is likely due to the characteristics of the tropical forest environment, which features a continuous, closed, multi-layered canopy structure. These conditions make it challenging for the model to delineate crown

Table 3
Semantic segmentation results on real-world datasets. The last four columns represent the IoU values for different semantic categories. The symbol '-' denotes the absence of that semantic category in the sample plot. Unit: %.

	Platform	mIoU	mACC	oACC	understory	leaf	terrain	wood
FOR-Instance	ULS	78.37	85.92	94.96	70.24	94.54	82.07	66.65
- CULS		93.38	96.20	98.95	–	97.63	99.91	82.59
- NIBIO		72.54	78.11	94.80	94.72	94.48	31.28	69.65
- RMIT		70.95	78.79	90.55	43.48	91.56	88.49	60.29
- SCION		83.23	88.39	95.02	83.02	94.68	92.72	62.52
- TU Wien		77.56	88.12	95.48	59.73	94.36	97.96	58.18
Lin3D	ULS	87.38	91.05	98.28	73.02	98.33	95.08	83.09
Yuchen	ULS	79.36	86.81	98.30	–	98.27	–	60.45
NIBIO	MLS	90.42	95.15	96.80	88.79 **(vegetation)**		89.21	93.27

Note: In the NIBIO dataset, the semantic labels of understory and leaf are merged as vegetation.

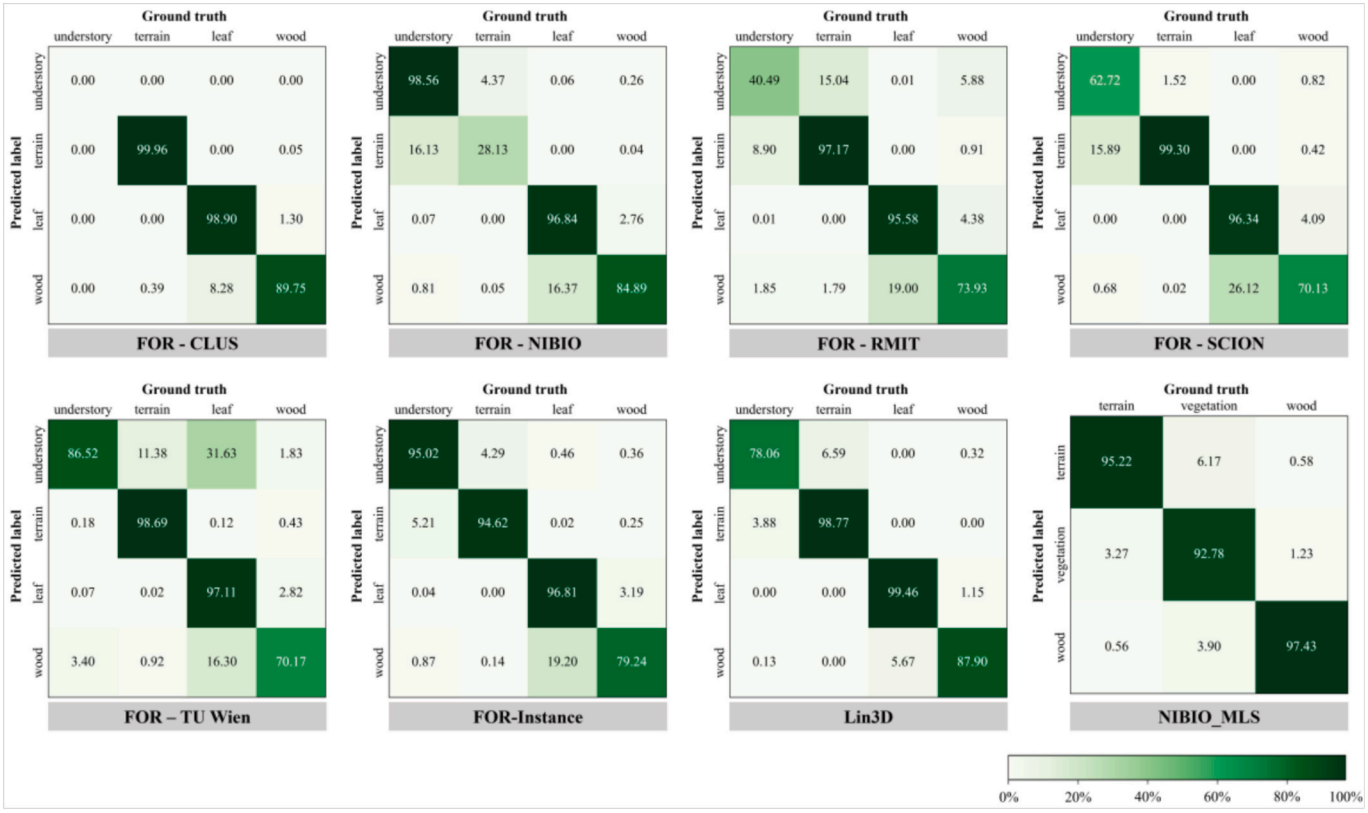

Fig. 11. Confusion matrix for the semantic categories of different real-world datasets.

Table 4
Instance segmentation results on real-world datasets.

	Platform	Completeness (%)	Omission (%)	Commission (%)	F1-score (%)
FOR-Instance		81.92	18.08	12.66	84.25
- CULS		100.00	0.00	9.09	95.24
- NIBIO	ULS	90.24	9.76	6.67	91.76
- RMIT		70.31	29.69	29.69	70.31
- SCION		83.33	16.67	0.00	90.91
- TU Wien		65.71	34.29	17.86	73.02
Yuchen	ULS	82.50	17.50	10.81	85.71
TreeLearn	MLS	93.68	6.32	3.94	94.84
NIBIO	MLS	77.92	22.08	3.23	86.33
SYSSIFOSS	ALS	85.61	14.39	22.09	81.58

boundaries between adjacent large trees accurately. Additionally, MST demonstrated notably strong performance on MLS-derived datasets, particularly in the TreeLearn dataset, which achieved an impressive completeness score of 93.68% alongside a low commission error rate of 3.94%. This exceptional result underscores the powerful cross-platform feature learning capabilities of MST, effectively capturing distinct canopy structures and tree crown morphologies unique to MLS-acquired data. Representative segmentation outcomes from selected plots are illustrated in Appendix S3 Fig. A7 and Appendix S3 Fig. A8, underscoring the precision of MST in delineating trees even in densely structured stands. Similarly, commendable results were observed within ALS-acquired datasets (Appendix S3 Fig. A9 and Table 4). While ALS technology inherently captures data at elevated altitudes, efficient for landscape-scale inventories, it inevitably results in relatively sparse point densities, potentially impairing the structural integrity of forest representations and complicating model feature learning. However, the strong capacity of MST for extracting generalized features from diverse datasets facilitates stable performance, effectively accommodating the challenges posed by lower point density and structural fragmentation in ALS data. Consequently, MST successfully delineated individual trees even in densely forested environments, as exemplified in Appendix S3 Fig. A9.

The following demonstrates the exploitation of individual tree segmentation results. Segmentation enables the derivation of tree-level structural attributes, which are fundamental to forest inventories, biomass and carbon assessments. Therefore, following common practice (Kato et al., 2009; Weiser et al., 2022; Yao et al., 2012), we evaluate the accuracy of the extracted tree height (H), crown diameter (CD), crown volume (CV), and diameter at breast height (DBH) using scatterplots of matched individual trees. Notably, DBH is not reported for the Yuchen dataset, the RMIT plot in the FOR-Instance dataset, or the SYSSIFOSS dataset, as these datasets were acquired by ALS or ULS and contain sparse stem observations, which preclude reliable DBH estimation. As shown in Fig. 12, the estimated H correlates well with the reference values, with R^2 generally exceeding 0.9 across datasets. In contrast, crown-related attributes (i.e., CD and CV) exhibit larger discrepancies. The RMSE of CD is 0.74 ± 0.10 m. In contrast, CV shows an even larger deviation, primarily attributable to occasional under-segmentation that merges neighboring crowns and inflates the estimated crown volume to reflect multiple trees. For DBH, performance is relatively weaker on FOR-Instance, consistent with the ULS acquisition setting. Although some plots exhibit comparatively high point density, stem sampling remains substantially sparser than in point clouds acquired by ground-based LiDAR systems, thereby limiting the robustness of DBH estimation.

4.3. *Transferability to TLS forest data*

To further evaluate the generalization of the MST framework to TLS

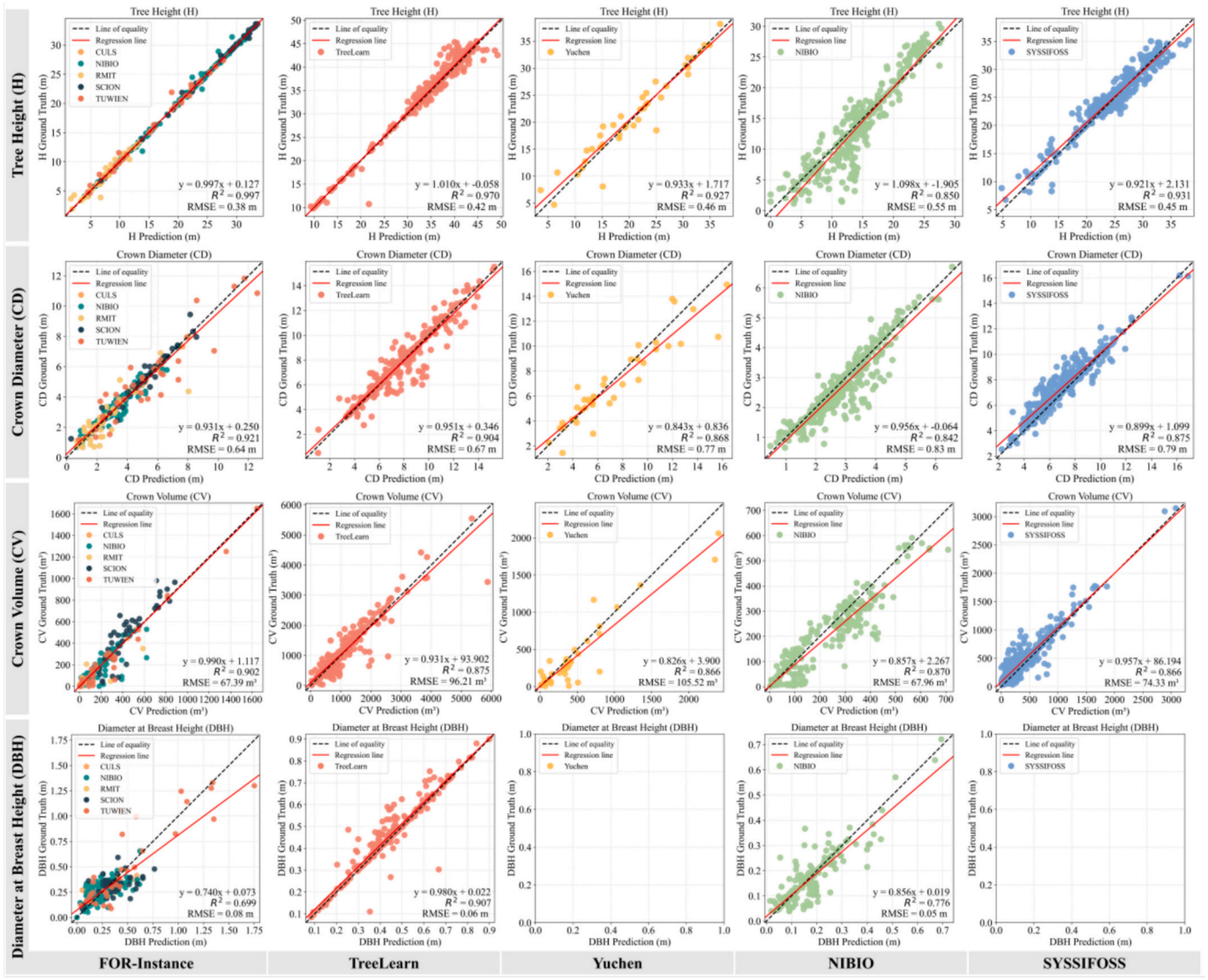

Fig. 12. Scatterplots between the predicted and reference values for individual tree height (H), crown diameter (CD), crown volume (CV), and diameter at breast height (DBH) of real-world datasets. DBH is not reported for the Yuchen dataset, nor for the RMIT plot in the FOR-Instance dataset or the SYSSIFOSS dataset, as these datasets were acquired by ALS or ULS and contain sparse stem observations, which preclude reliable DBH estimation.

forest data, we conducted additional experiments on Lin3D_v0.1 and Wytham Woods datasets, for the semantic and instance segmentation tasks, respectively. Given the similarity in data characteristics between MLS and TLS, which are both near-ground acquisition modalities, our framework excluded TLS data from the initial training set. Despite MST being trained solely on synthetic ALS, ULS, and MLS datasets without any TLS-specific data, results demonstrated satisfactory segmentation performance on both tasks.

Specifically, the MST model achieved an overall mIoU of 79.48% on the Lin3D dataset, with apparent segmentation inaccuracies primarily related to understory vegetation, as illustrated in Table 5 and Fig. 13. Such inaccuracies can be attributed to the inherent characteristics of TLS data, which provide highly detailed and fine-grained representations of understory vegetation structures, including complete grass and shrub architectures. In contrast, the understory vegetation in the synthetic dataset used for MST pre-training was generated through graph-path algorithms applied to benchmark plot data and augmented by random scaling and rotation (Liu et al., 2026). Although this strategy aims to approximate realistic vegetation conditions, it inevitably lacks the precision and granularity of understory vegetation captured by TLS, leading to comparatively ambiguous feature representations learned by MST. Despite this suboptimal performance, MST achieved improved segmentation results across other semantic categories in the Lin3D dataset. In the Wytham Woods dataset, collected under leaf-off conditions in

Table 5
Semantic and instance segmentation results on TLS acquisition datasets. Unit: %.

Lin3D v0.1	mIoU	mACC	oACC	Lower objects	Foliage	Ground	Wood
	79.48	85.59	94.76	49.70	88.02	95.91	84.28
Wytham woods	Completeness	Omission	Commission	F1-score			
	78.57	21.43	14.29	81.99			

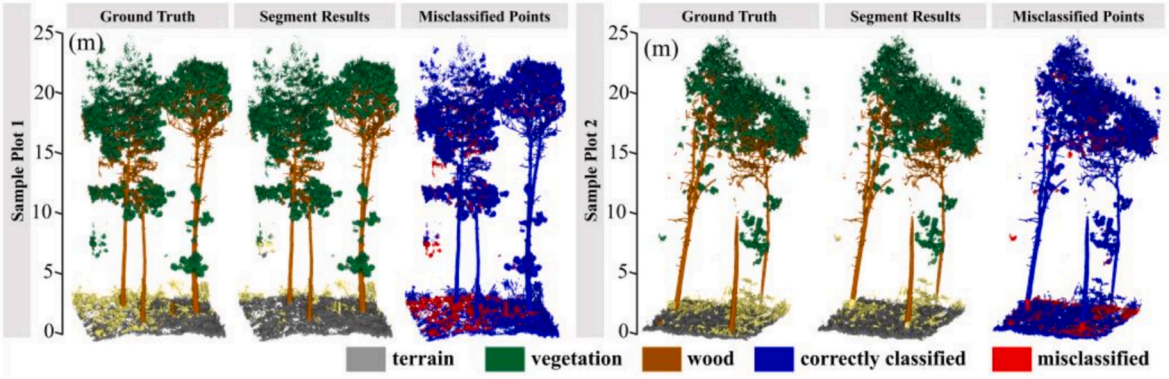

Fig. 13. Visualization results of semantic segmentation for the Lin3d_v0.1 dataset.

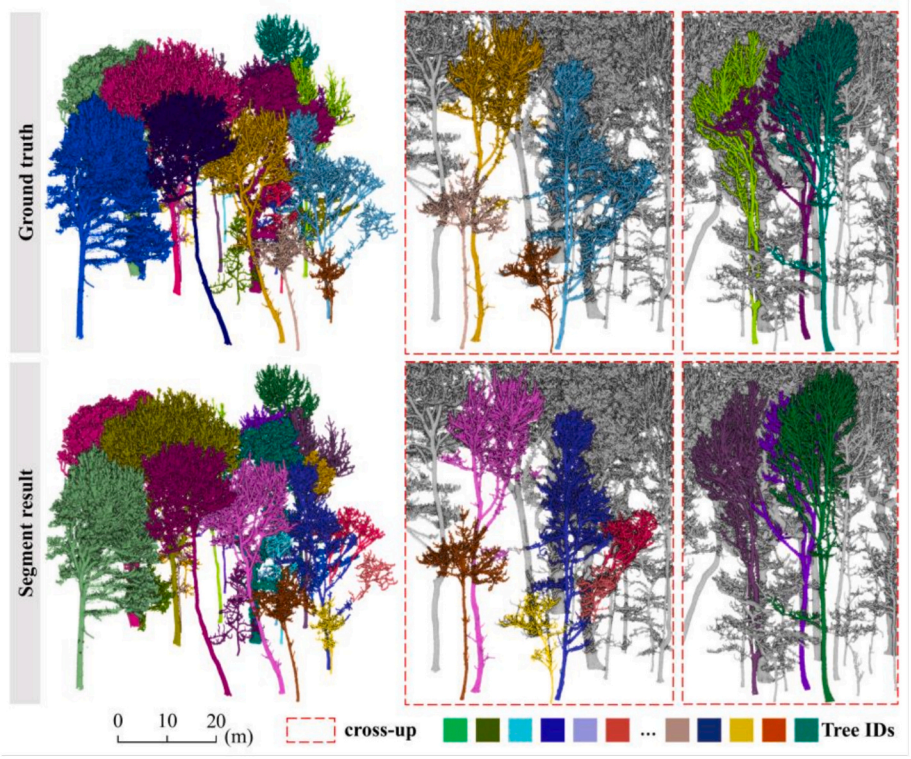

Fig. 14. Visualization results of instance segmentation for the Wytham Woods dataset. Each color represents an individual tree.

winter, MST achieved a completeness of 78.57%. As shown in Fig. 14, MST effectively delineated most individual trees, demonstrating reliable performance even in scenarios with intersecting branches. However, some over-segmentation occurred in bifurcated branching zones. This was primarily due to the absence of leaves, which exposed only trunks and branches, increased structural complexity, and made intricate branch junctions more visible, thereby complicating instance-level differentiation.

Overall, despite the absence of TLS data during training, MST demonstrated commendable transferability and segmentation performance on TLS-acquired datasets. This effectiveness underscores the adeptness of MST in learning generalized, transferable representations across diverse platform-derived point clouds, reinforcing its applicability in detailed ecological studies and forest management practices requiring precise individual tree delineation across varying data acquisition techniques.

5. Discussion

5.1. Comparison of MST with other models and pipelines

To comprehensively evaluate the effectiveness of the MST framework in understanding forest scenes, we conducted comparative analyses with state-of-the-art methods across semantic and instance segmentation tasks. For semantic segmentation, the FOR-Instance and Lin3D datasets were employed as benchmarks. On the FOR-Instance dataset, two deep learning networks dedicated to forest analysis were deployed, including Sen Net (Lu et al., 2025) and ForAINet (Xiang et al., 2024). Sen Net is specifically constructed for semantic component segmentation of forest scene point clouds. ForAINet is a recent panoptic framework optimized for holistic forest interpretation. Notably, MST adopts a single *wood* category, whereas FOR-Instance subdivides this class into *stem* and *wood branches*. Although a rigorous comparison is not

Table 6
Quantitative results of semantic segmentation comparing MST with other forest components segmentation methods. The bold numbers indicate the best performance. The last four columns represent the IoU values for different semantic categories. The symbol '-' denotes the absence of that semantic category in the dataset, while '/' indicates that the evaluation metric value was not provided by the method. Unit: %.

	Model	mIoU	mACC	oACC	Low-vegetation	Live branches	Terrain	Stem	Woody branches
	Sen Net*	78.2	/	84.5	**90.0**	**95.5**	78.0	43.0	77.3
FOR-Instance	ForAINet*	75.7	/	93.6	89.1	94.3	78.7	55.6	60.7
	MST (ours)	**78.4**	**85.9**	**95.0**	70.2	94.5	**82.1**	**66.7**	
	Model	mIoU	mACC	oACC	Lower objects	Foliage	Ground	Wood	
	WLSeg*	76.2	/	/	/	90.2	/	62.1	
Lin3D	LeWoS*	77.6	/	/	/	91.0	/	64.2	
(ULS)	Sen Net*	85.2	/	97.6	66.8	**98.9**	93.1	81.8	
	MST (ours)	**87.4**	**91.0**	**98.3**	**73.0**	98.3	**95.1**	**83.1**	

* values reported by Lu et al. (2025).

attainable, the reported outcomes still provide meaningful reference. Quantitative results summarized in Table 6 reveal that MST slightly surpasses Sen Net in mIoU even though no forest-specific architecture was designed in MST. However, MST underperforms Sen Net on the *understory* class, potentially due to limitations in transferring representations from the virtual synthetic dataset (Boreal3D), which fails to fully encapsulate the high morphological variability of understory vegetation. Conversely, MST demonstrates effective segmentation of terrain and wood components, reflecting the reliability of its learned feature representations. MST also exhibits enhanced performance compared to ForAINet. This performance advantage mainly stems from MST in semantic segmentation specially focuses on component-level segmentation, whereas the broader panoptic objectives of ForAINet may dilute feature learning for individual classes.

Further validation was conducted on the Lin3D dataset against three representative approaches, including WLSeg (Wan et al., 2021), LeWoS (Wang et al., 2020), and Sen Net (Lu et al., 2025). The machine learning-based WLSeg and LeWoS leverage botanical growth patterns of tree structures to facilitate wood-leaf separation. The quantitative segmentation results are presented in Table 6. Notably, the values were reported

Table 7
Quantitative results of instance segmentation comparing MST with other individual tree segmentation methods. The bold numbers indicate the best performance compared with other methods. The underlined numbers indicate suboptimal performance. The symbol '/' indicates that the evaluation metric value was not provided by the method.

	Model	Completeness	Omission	Commission	F1-score
	YOLOv5*	**1**	**0**	/	/
	ForAINet*	**1**	**0**	0.13	0.93
FOR-Instance	SegmentAnyTree*	**1**	**0**	**0**	**0.99**
- CULS	Treeios	0.95	0.05	0.09	0.93
	SSSC	0.67	0.33	0.18	0.74
	MST (ours)	**1**	**0**	0.09	0.95
	YOLOv5*	0.67	0.33	/	/
	ForAINet*	0.88	0.12	**0.03**	**0.92**
FOR-Instance	SegmentAnyTree*	0.88	0.12	0.09	0.88
- NIBIO	Treeios	0.81	0.19	0.04	0.88
	SSSC	0.41	0.59	0.54	0.43
	MST (ours)	**0.90**	**0.10**	0.07	**0.92**
	YOLOv5*	0.58	0.42	/	/
	ForAINet*	0.64	0.36	0.24	0.70
FOR-Instance	SegmentAnyTree*	0.69	0.31	**0.17**	**0.83**
- RMIT	Treeios	0.59	0.41	0.39	0.60
	SSSC	0.02	0.98	0.86	0.03
	MST (ours)	**0.70**	**0.30**	0.30	0.70
	YOLOv5*	0.86	0.14	/	/
	ForAINet*	0.87	0.13	0.04	**0.91**
FOR-Instance	SegmentAnyTree*	**0.92**	**0.08**	0.07	**0.91**
- SCION	Treeios	0.83	0.17	0.21	0.81
	SSSC	0.17	0.83	0.67	0.22
	MST (ours)	0.83	0.17	**0.00**	**0.91**
	YOLOv5*	0.20	0.80	/	/
	ForAINet*	**0.71**	**0.29**	0.32	0.69
FOR-Instance	SegmentAnyTree*	0.46	0.54	0.45	0.57
- TU Wien	Treeios	0.49	0.51	0.41	0.53
	SSSC	0.11	0.89	0.50	0.19
	MST (ours)	0.66	0.34	**0.18**	**0.73**
	TreeLearn*	/	/	/	**0.98**
	SegmentAnyTree*	0.93	0.07	**0.03**	0.92
TreeLearn	Treeios	0.91	0.09	0.37	0.74
	SSSC	0.91	0.09	0.10	0.90
	MST (ours)	**0.94**	**0.06**	0.04	0.95
	Point2tree*	0.57	0.43	0.07	0.61
	TLS2trees*	0.59	0.41	0.14	0.62
NIBIO	SegmentAnyTree*	0.77	0.23	0.09	**0.88**
	Treeios	0.59	0.41	0.19	0.68
	SSSC	0.56	0.44	0.24	0.64
	MST (ours)	**0.78**	**0.22**	**0.03**	0.86

* values reported by Wielgosz et al. (2024).

by Lu et al. (2025) and were derived from evaluations conducted on the full Lin3D test dataset. However, the segmentation results of MST were tested exclusively on a subset of plots currently publicly available in the Lin3D dataset. Results indicate that MST achieves an mIoU of 87.4%, outperforming WLSeg and LeWoS by 10%. Unlike unsupervised or weakly supervised pipelines, MST learns shared structural representations of tree components, yielding pronounced gains on the *wood* category. Although the morphology of woody structures is generally influenced by ecological growth principles, it remains highly heterogeneous across different forest stands and tree species. Compared with Sen Net, MST achieves comparable overall performance, which can be attributed to the limited structural complexity of the publicly available Lin3D plots, characterized by single-layer canopies and relatively homogeneous semantic class distributions.

Additionally, the MST framework demonstrates consistent performance for instance segmentation tasks when benchmarked against several state-of-the-art methods, including YOLO_v5-based tree crown segmentation (Straker et al., 2023), ForAINet (Xiang et al., 2024), SegmentAnyTree (Wielgosz et al., 2024), TreeLearn (Henrich et al., 2024), Point2tree (Wielgosz et al., 2023), TLS2trees (Wilkes et al., 2023), and two unsupervised graph-based methods, Treeiso (Xi and Hopkinson, 2022), and SSSC (Wang, 2020). As shown in Table 7, MST consistently achieves superior segmentation accuracy across most forest plots. In scenarios where MST did not achieve the highest accuracy, it consistently attained near-optimal performance. Specifically, within the FOR-Instance dataset, MST delivered highly satisfactory segmentation results, including achieving the second-best performance in the particularly challenging TU Wien plot, characterized by complex, unmanaged, multilayered mixed-deciduous forest structures. This ecological complexity, characterized by overlapping asymmetric crowns and interspersed understory vegetation, challenges conventional segmentation methods that rely on geometric crown priors. Additionally, MST also demonstrated generalizable performance to MLS datasets. MST outperforms Point2Tree and TLS2Trees by 20% in completeness on the NIBIO dataset, reflecting its capacity to generalize across heterogeneous forest architectures. This enhanced performance stems from the capability of MST to capture stable, generalized structural representations of trees during pre-training on synthetic forest point cloud datasets. Consequently, MST effectively handles the inherent heterogeneity across datasets derived from various platforms, underscoring its reliability and suitability for precise ecological assessments and forest management applications. Compared with unsupervised graph-based approaches, MST still exhibits stronger and more consistent performance across all datasets, as shown in Table 7. Treeiso obtains tree instances through graph partitioning (e.g., cut-pursuit and normalized cut (Xi and Hopkinson, 2022; Yao et al., 2012)), whereas SSSC relies on graph pathing (e.g., Dijkstra shortest paths (Tao et al., 2015; Wang, 2020)). Although these methods do not require a learning stage, their rule-driven designs are prone to failure in complex forest stands. This limitation is particularly evident in SSSC on the FOR-Instance dataset, where sparse stem observations in ULS point clouds hinder the detection of tree-root seed nodes, leading to frequent segmentation failures. Moreover, unsupervised graph-based pipelines typically involve extensive parameter tuning, which can hinder practical deployment across different forests and platforms.

Additionally, to further validate the cross-platform generalization of the proposed method, we compare Adaptive Batch Normalization (AdaBN) (Li et al., 2016) and Maximum Mean Discrepancy (MMD) (Ghifary et al., 2014) against the baseline (mix training) and our MST, as summarized in Table 8. These two strategies were selected as representative adaptation baselines because they are comparable to MST in that they can be incorporated into the present framework without redesigning the original segmentation pipeline. By contrast, a rigorous adversarial adaptation baseline would require additional domain discriminators and extra optimization objectives, and its consistent extension to both semantic segmentation and grouping-based instance segmentation would call for a substantially different architectural design. AdaBN is a representative architecture-based adaptation strategy that mitigates domain gaps by recalibrating BN statistics with target-domain data at inference time (Wang et al., 2025). MMD, in contrast, represents a distribution-alignment strategy that explicitly minimizes the discrepancy between source- and target-domain feature distributions at selected layers during training, thereby encouraging more domain-invariant representations (Wang and Deng, 2018). As shown in Table 8, both AdaBN and MMD consistently improve semantic segmentation over baseline, but remain clearly below MST across all platforms, indicating that distribution alignment yields a stronger gain than BN recalibration in this setting. A similar trend is observed for instance segmentation: both AdaBN and MMD substantially improve completeness and F1-score relative to baseline, while MST achieves the best performance across platforms. Notably, MMD provides stronger gains than AdaBN on ALS and ULS data, whereas on MLS data the improvement of MMD is smaller than AdaBN. This suggests that feature-level distribution matching can be susceptible to over-alignment in instance segmentation. By reducing cross-platform discrepancies, MMD may suppress geometry variations that are informative for instance boundaries and local separability, thereby impairing the quality of grouping/clustering on platforms where fine-grained structural cues differ. In contrast, MST outperforms these adaptation strategies as it follows a token-conditioned representation learning paradigm, which learns shared structural priors while retaining platform-aware modulation. Therefore, platform heterogeneity is addressed through controlled conditioning rather than forced global alignment, leading to stable, accurate cross-platform performance in both semantic and instance

Table 8
Cross-platform semantic and instance segmentation results on Boreal3D, comparing baseline with domain-adaptation strategies (AdaBN and MMD) and the proposed MST. The bold numbers indicate the best performance. Unit: %.

Model	Test dataset	Instance segmentation				Semantic segmentation		
		Completeness	Omission	Commission	F1-Score	mIoU	mAcc	oAcc
Mix training	ALS	8.11	91.89	35.81	14.39	50.46	60.69	85.35
	ULS	81.15	18.85	25.72	77.57	71.16	76.66	91.54
	MLS	68.00	32.00	15.77	75.25	74.26	83.58	85.37
AdaBN (test-time BN recalibration)	ALS	28.79	71.21	24.86	41.63	51.97	61.48	85.77
	ULS	84.49	15.51	15.48	84.50	73.24	78.37	92.34
	MLS	80.48	19.52	12.26	83.95	76.22	84.98	87.46
MMD (Train-time feature alignment)	ALS	36.44	63.56	16.45	50.75	53.74	64.21	86.31
	ULS	89.74	10.26	10.33	89.70	76.22	80.12	93.67
	MLS	75.35	24.65	15.04	79.87	80.64	85.89	90.58
MST (ours)	ALS	43.86	56.14	0.96	**60.79**	**60.12**	69.58	87.15
	ULS	98.39	1.61	0.43	**98.98**	**83.82**	86.99	95.70
	MLS	97.88	2.12	9.57	**94.01**	**87.13**	91.20	94.49

segmentation.

5.2. Performance across different proportions of labeled data

To comprehensively validate the effectiveness and efficiency of our MST framework, we further investigated its performance for both semantic and instance segmentation tasks across varying proportions of real-world labeled data. In this experiment, we used the representative FOR-Instance dataset, which contains both semantic and instance annotations. Specifically, experiments were conducted using randomly selected subsets corresponding to 80%, 60%, 40%, 20%, and 10% of the available labeled data for fine-tuning, with performance evaluated on the same testing dataset. In addition, we compared fine-tuning results with different proportions of labeled data to training from scratch on the same subsets. Detailed segmentation results under these scenarios are summarized in Table 9 and Fig. 15.

Our analysis reveals that the MST framework consistently outperforms models trained directly from scratch, exhibiting substantial resilience to reductions in labeled training data. For the instance segmentation, completeness declined by 3% at 20% labeled data, while a pronounced degradation of 7% in completeness and 8% in F1-score occurred at 10% labeled data. These thresholds reflect the minimal supervision required to align synthetic priors with ecological reality. In contrast, training from scratch under these same data-limiting conditions resulted in substantial decreases of roughly 20% in both completeness and F1-score, underscoring the reliability of the MST framework and its strong ability to capture inherent ecological structures from limited real-world data. Similarly, for semantic segmentation, utilizing 80% and 60% of labeled data resulted in negligible performance drops (less than 1% in both mIoU and oAcc). Moreover, even with only 20% of labeled data, performance degradation remained modest at approximately 3%. Nonetheless, a significant 6.6% decline in mIoU was observed when fine-tuning was performed with only 10% labeled data. Conversely, models trained from scratch suffered considerably greater performance deterioration, with mIoU decreasing by over 10% under identical circumstances.

The stability observed at lower labeling proportions can primarily be attributed to the comprehensive, generalized structural features acquired by the MST framework during pre-training. By leveraging multi-platform synthetic data, the framework successfully captures universal structural regularities and distills tree morphology at the tree and component levels. Consequently, the pre-acquired knowledge significantly enhances the model capability to adapt to diverse real-world datasets, substantially mitigating labeling demands. This outcome highlights the practical utility of MST, significantly reducing annotation workload, thereby offering considerable advantages for large-scale ecological research and sustainable forest management.

5.3. Ablation study and comparative experiments

To further validate the effectiveness of the MST framework, we conducted ablation and comparative experiments on the FOR-Instance dataset. We examined the role of different contextual integration strategies, initialization schemes, CPAT dimensionalities, and performed baseline comparisons. As illustrated in Appendix S4 Table S2, removing the contextual integration strategy and directly injecting CPAT embeddings into the backbone network caused performance drops of 5.1% in semantic segmentation mIoU and 4.7% in instance segmentation F1-score. This highlights the critical role of the contextual integration strategy in effectively aligning CPATs with the backbone features. When replacing CIM with a cross-attention mechanism that computes weighted correlations between CPAT embeddings and feature maps, segmentation performance also declined by 2.36% in F1-Score and 1.89% in mIoU. This suggests that cross-attention lacks the adaptive normalization property of CIM, which better balances platform-specific adjustments and shared structural learning. Furthermore, as evidenced by the loss curves of Appendix S4 Fig. A10, different contextual integration settings resulted in unstable optimization characterized by pronounced oscillations and poor convergence, highlighting the stability advantage of CIM.

Furthermore, when the zero-initialization strategy for CPATs was removed, performance decreased by 1% in mIoU and F1-score for semantic and instance segmentation, respectively (Appendix S4 Table S2). Although the reduction in accuracy is modest, the practical importance of zero initialization lies primarily in improving optimization stability rather than directly increasing segmentation performance. This confirms that zero initialization stabilizes the early optimization process by allowing the backbone to first focus on learning shared structural representations before CPATs gradually inject platform-specific adjustments. In contrast, random initialization introduced noisy signals during early training, which leads to marginally reduced accuracy and training instability. This is evidenced by the oscillation loss curve and slower convergence observed in Appendix S4 Fig. A10, where zero initialization yields smoother convergence with less oscillation, while random initialization exhibits stronger fluctuations during training. These results collectively validate that zero initialization provides a more stable and reliable foundation for the training process.

We also evaluated the effect of CPAT dimensionality on segmentation accuracy and computational complexity. As summarized in Appendix S4 Table S3 and illustrated in Appendix S3 Fig. A10, the best performance was achieved with 256-dimensional CPATs. Using 128 or 512 dimensions resulted in approximately 1% decreases in both semantic and instance segmentation performance. However, increasing the dimensionality to 1024 results in a more pronounced 2% decline in the instance segmentation F1-score. This reduction is likely due to

Table 9
Instance and semantic segmentation quantitative results of the MST framework under different proportions of real-world annotations. Unit: %. The underlined numbers denote training from scratch.

Proportion	Instance segmentation				Semantic segmentation		
	Completeness	Omission	Commission	F1-Score	mIoU	mAcc	oAcc
100%	88.29	11.71	10.91	88.69	69.42	74.78	94.65
	90.24	9.76	6.67	91.76	72.54	78.11	94.80
80%	85.59	14.41	12.84	86.36	68.38	73.59	94.04
	89.19	10.81	7.48	90.83	72.26	77.23	94.73
60%	81.98	18.02	15.74	83.11	66.93	72.21	93.82
	88.29	11.71	6.67	90.74	71.74	76.74	94.52
40%	77.48	22.52	19.63	78.90	64.08	70.64	92.75
	87.39	12.61	7.62	89.81	71.09	76.02	94.15
20%	72.07	27.93	24.53	73.73	62.62	67.24	91.51
	86.96	13.04	13.04	86.96	69.71	74.80	93.75
10%	67.57	32.43	29.91	68.81	59.21	63.58	90.04
	83.78	16.22	16.96	83.41	65.90	73.10	92.34

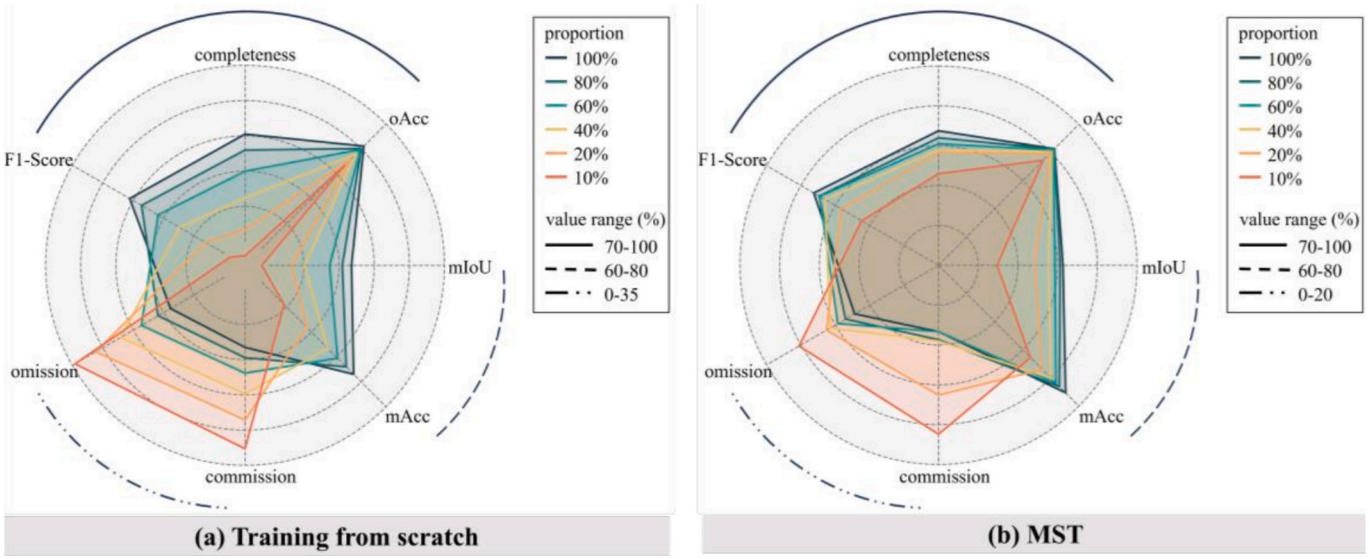

Fig. 15. Results of instance segmentation and semantic segmentation evaluation metrics with different proportions of real-world labeled data. (a) Evaluation results with training from scratch. (b) Evaluation results with MST.

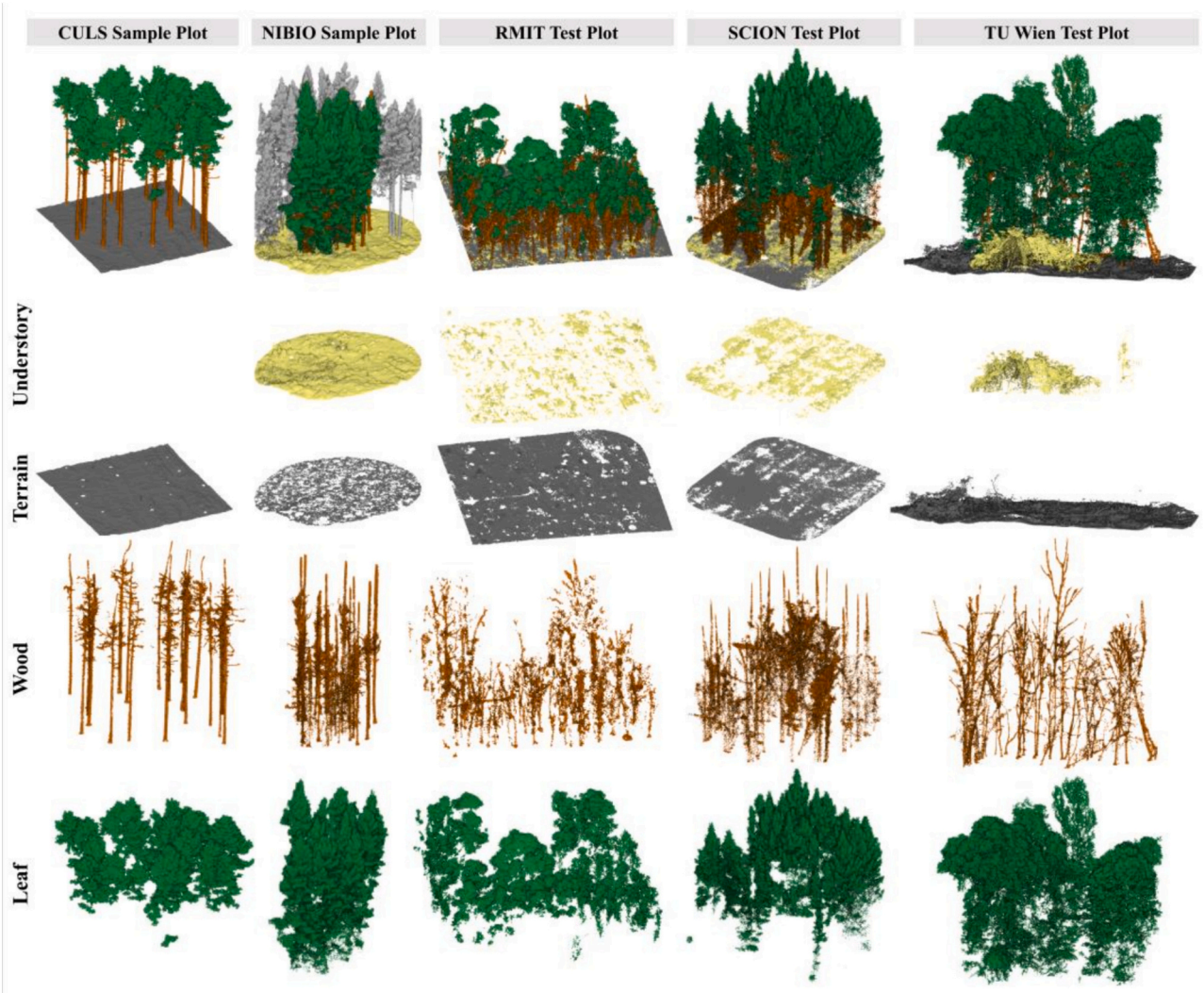

Fig. 16. Visualization of different semantic components in the FOR-Instance dataset.

overfitting, in which CPATs' excessive model capacity captured spurious platform-specific variations rather than meaningful shared structural patterns. Notably, compared with the baseline model, the inclusion of CPATs introduced only about 30 k additional parameters, representing a negligible increase in computational cost relative to the overall network size.

5.4. *Understanding segmentation performance variability across forest plots*

Significant variability in the performance of our segmentation framework across diverse forest plots prompted a detailed analysis of the underlying factors influencing both semantic and instance segmentation outcomes. Specifically, for the semantic segmentation task, the observed mIoU ranged from 71% to 93%, as shown in Table 3. A more detailed analysis revealed consistent and effective segmentation performance for the *leaf* and *wood* categories across all tested forest plots. In contrast, an inverse performance relationship emerged between the *understory* and *terrain* categories, particularly prominent in sample plots such as NIBIO, RMIT, and TU Wien. This phenomenon, on the one hand, can be attributed to the structural complexity of forest stands, which inherently challenges the distinction between understory vegetation and ground surfaces, as both components often intermingle spatially. On the other hand, differences in annotation standards across datasets significantly influenced segmentation outcomes. Fig. 16 visually illustrates these discrepancies using separated semantic classes from the FOR-Instance dataset. In the NIBIO forest plot, a substantial proportion of the point cloud was classified as understory, predominantly due to widespread grass cover, which made terrain points appear sparse and discretely distributed. Conversely, within the RMIT plot, a fraction of points belonging to understory vegetation were categorized as terrain. This misclassification likely stems from the inherent difficulty of manually delineating understory vegetation from terrain in dense, structurally complex forests. In the TU Wien plots, the annotation guidelines explicitly classified shrubs as *understory*, whereas dense grass layers and coarse woody debris were categorized under *terrain*, reflecting dataset-specific labeling conventions for delineating terrain boundaries. Consequently, the observed variability in semantic segmentation performance across different datasets can be partially attributed to inconsistencies in semantic annotation criteria, combined with the inherent ecological complexity of forest components differentiation. This analysis underscores the need to harmonize annotation standards and highlights the importance of ecological understanding in interpreting segmentation results across diverse forest environments.

Furthermore, the virtual synthetic dataset (Boreal3D) was utilized for pretraining in the MST framework. The observed performance

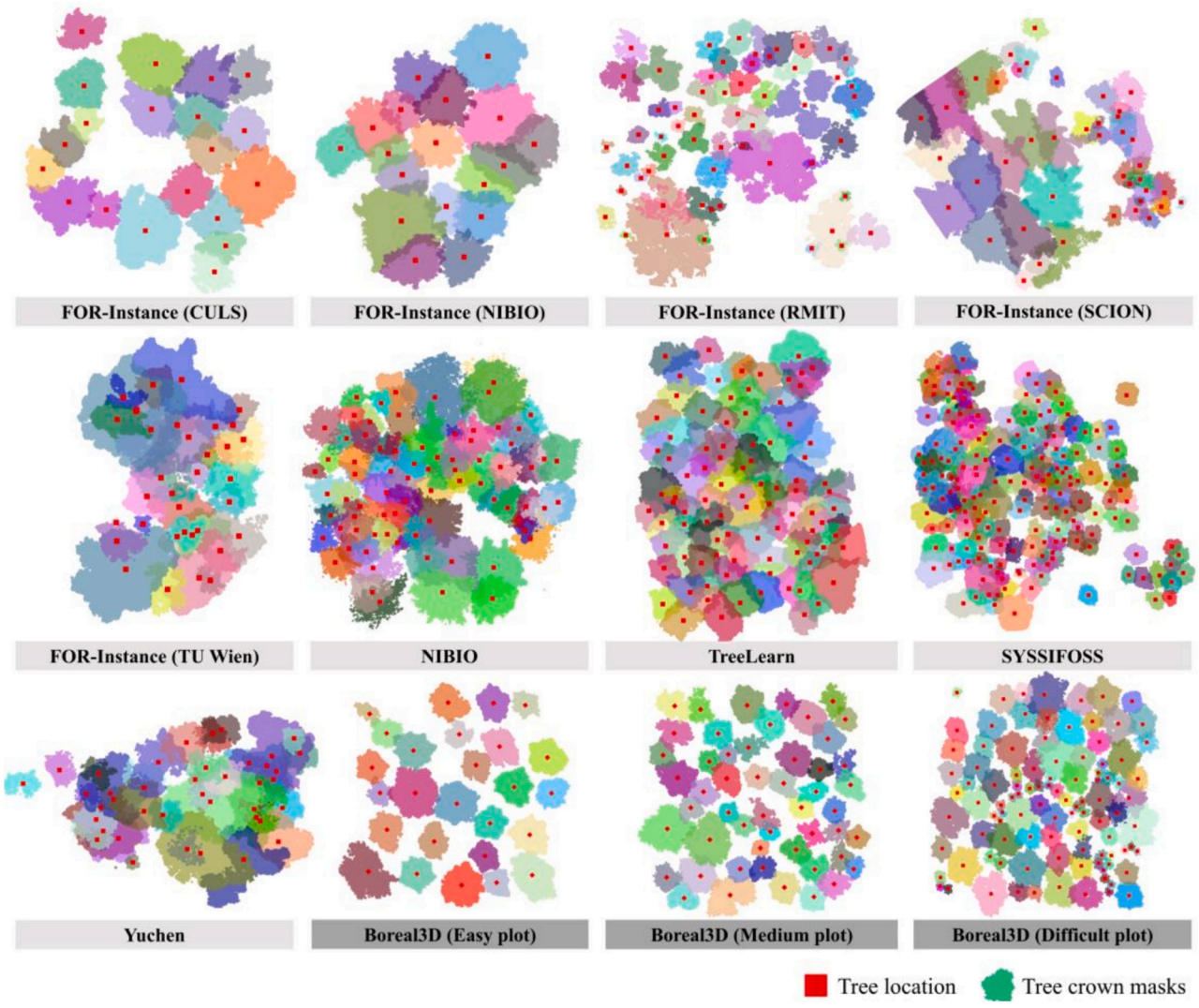

Fig. 17. Visualization of tree crown masks and tree locations in individual tree segmentation datasets. Dark gray boxes represent the virtual synthetic dataset (Boreal3D), whereas the light gray boxes represent real-world datasets.

variability can be interpreted in terms of ecological similarity (forest type and dominant species) and acquisition similarity (platform-specific sampling characteristics and point density). The highest mIoU values are obtained on datasets that exhibit strong consistency with Boreal3D in either species composition or sampling characteristics. In particular, CULS within FOR-Instance (mIoU = 93.38%) and NIBIO (MLS) (mIoU = 90.42%) are conifer-dominated stands with dominant species overlapping Boreal3D's boreal composition (pine/spruce/birch). However, despite similar conifer composition, NIBIO (ULS) in FOR-Instance shows substantially lower performance (mIoU = 72.54%). A plausible explanation is the pronounced mismatch in point density between Boreal3D's ULS data (~900 pts./m^2) and the NIBIO (ULS) data (~9500 pts./m^2). Such a density mismatch likely weakens transfer by altering local neighborhood statistics, geometric continuity, and class boundaries. In addition, strong semantic performance can also emerge when ecological similarity is weaker but platform and density are well aligned: Lin3D v0.2 (ULS) reaches mIoU = 87.38% in a subtropical evergreen broadleaf forest, while its point density (944 pts./m^2) closely matches Boreal3D's ULS simulation (908 pts./m^2), suggesting that pretraining on platform-consistent sampling patterns can yield transferable geometric representations even across different forest types. In contrast, lower semantic performance is generally observed when both ecological conditions and acquisition characteristics deviate from the pretraining distribution, or when key semantic categories are insufficiently observed (e.g., ULS plots with sparse or occluded ground returns), which limits the extent to which pretraining can regularize downstream learning.

In the instance segmentation task, considerable differences in segmentation performance were similarly observed across the forest plots, as highlighted in Table 4. Specifically, the CULS plot achieved 100% completeness, whereas the TU Wien plot showed a notably lower 66%. These significant discrepancies prompted further qualitative investigations focusing on forest stand complexity, tree density, and canopy overlap, as illustrated in Fig. 17 and Fig. 18. Additionally, we examine whether the observed performance variability is related to the similarity in stand complexity between Boreal3D and the real-world datasets. The forest structure in CULS and NIBIO plots of the FOR-Instance dataset exhibited low complexity, predominantly comprising dominant trees with few suppressed individuals. Additionally, tree distribution in these plots was relatively uniform, resulting in minimal canopy overlap, which is conducive to accurate individual tree segmentation. This structural profile is most consistent with the easy plots in the Boreal3D dataset, where tree crowns are well separated and suppressed layers are limited, suggesting that pretraining on Boreal3D provides instance-level geometric priors that transfer effectively to low-overlap, low-stratification stands. Consequently, our framework achieved 100% and 90% detection rates for the CULS and NIBIO plots, respectively.

Conversely, the RMIT and TU Wien plots in FOR-Instance, SYSSIFOSS, as well as the Yuchen dataset, exhibit markedly higher stand complexity, characterized by multi-layered canopies, a larger fraction of suppressed trees, and substantially stronger crown interlacing. These structural characteristics increase occlusion and ambiguity in tree boundaries, thereby increasing both omission and commission errors. Consistent with this observation, instance segmentation performance is notably lower on these plots, with F1-scores of 70.31% (RMIT), 73.02% (TU Wien), 81.58% (SYSSIFOSS), and 82.50% (Yuchen). Moreover, although Boreal3D provides three stand-complexity levels (easy/medium/difficult) spanning increasing tree density, crown overlap, and vertical stratification (dark gray in Figs. 17–18), even the difficult Boreal3D plots exhibit less severe stand complexity than the most challenging real-world case. This discrepancy indicates that a complexity gap remains between the synthetic pretraining distribution and certain real-world forests, implying that pretraining alone cannot fully resolve the extreme boundary ambiguity introduced by heavy occlusion and strong inter-crown interpenetration. These results indicate that forest structural complexity and vertical stratification are critical determinants

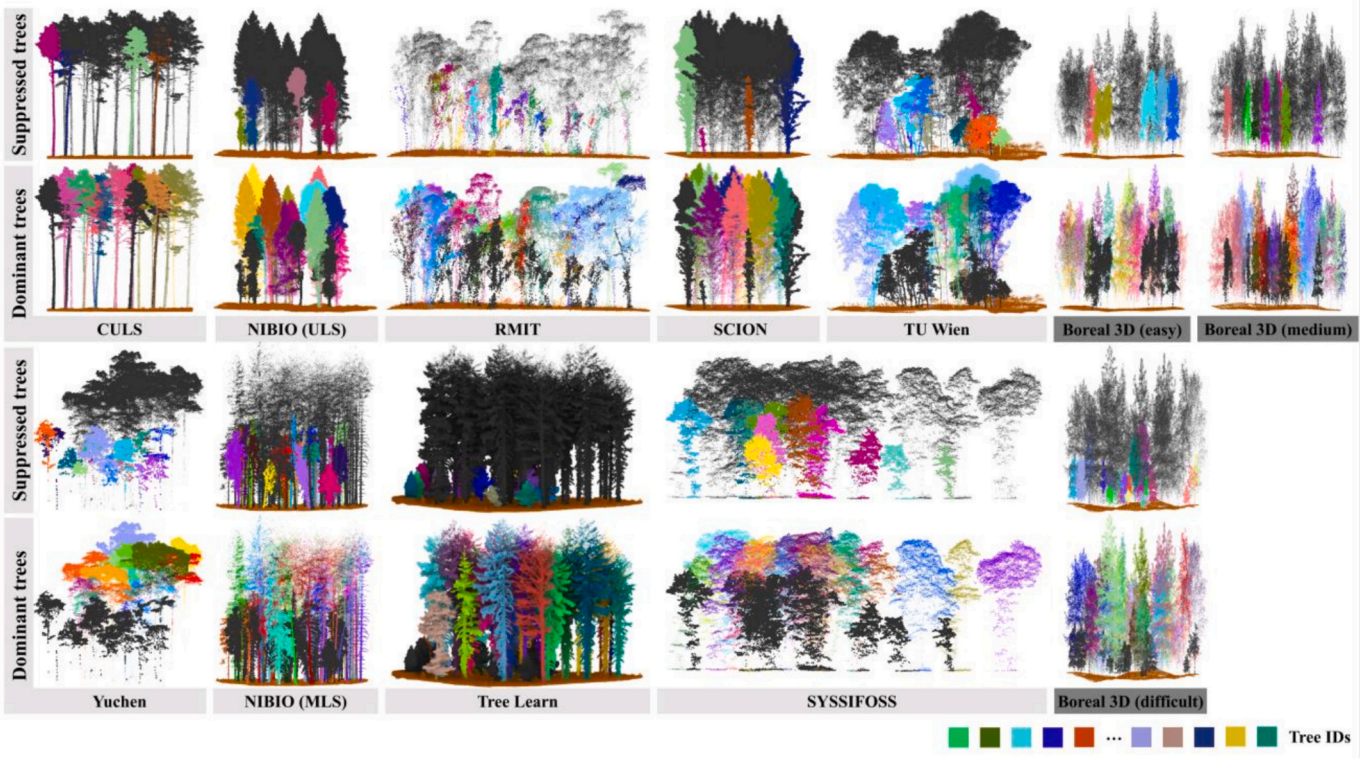

Fig. 18. Visualization of the forest stand complexity in individual tree segmentation datasets. In each sample plot, the upper subplot represents suppressed trees and the bottom subplot represents dominant trees. The dark gray box represents the virtual synthetic dataset (Boreal3D), whereas light gray boxes represent real-world datasets.

of instance segmentation accuracy, and that pretraining alone cannot fully compensate for severe synthetic-to-real complexity gaps.

5.5. Limitations and future work

The MST framework presented in this study effectively addresses the fundamental challenges posed by heterogeneity across multi-platform forest point cloud data. It mitigates negative transfer issues typically encountered in the traditional mixed-dataset training method. The demonstrated capability of MST in both semantic and instance segmentation tasks across multiple data acquisition platforms further validates its strong generalization potential. Moreover, our framework achieves near fully supervised performance with 20% real-world labeled data, underscoring its practical effectiveness and suitability for applications that require resource-efficient annotation strategies.

Despite these advancements, several opportunities remain for further enhancement. Overall, while our novel training framework effectively captures generalized forest structure representations from multi-platform datasets, the underlying segmentation architectures were not specifically designed for the complexities inherent to forest ecosystems. Indeed, several mature point cloud-based deep learning networks have demonstrated strong adaptability in downstream tasks. However, the diverse topological characteristics, species heterogeneity, intricate canopy structures, and multi-layered vegetation dynamics typical of forest environments impose stringent demands on network architectures explicitly tailored to forest environments. Therefore, future work should focus on designing specialized segmentation networks grounded in ecological and dendrological principles to better accommodate the structural nuances of forests. Additionally, although our training framework independently proved effective in semantic and instance segmentation tasks, these two tasks inherently complement each other. Developing a unified network capable of concurrently addressing both semantic and instance segmentation tasks (e.g., panoptic segmentation), such as ForAINet, could substantially enhance segmentation performance and efficiency in forest environments.

Moreover, the virtual synthetic dataset (Boreal3D) was utilized as our pre-training dataset. However, the current study does not provide a dedicated sensitivity analysis to identify which specific properties of the synthetic data are influential for MST, such as species composition and stand density, the strength and form of simulated sensor noise, point density induced by acquisition settings, and occlusion patterns arising from canopy size and tree spatial arrangement. As a result, the relative contribution of these data attributes to cross-platform transfer and real-world generalization remains insufficiently characterized, and clarifying these dependencies constitutes an important direction for future work. The tree-species composition of the Boreal3D pre-training dataset is highly consistent with that of the NIBIO dataset. Together with MST's ability to learn shared structural representations, this ecological similarity likely contributes to the strong semantic segmentation performance observed on NIBIO, where the mIoU exceeds 90%. This suggests that species composition in the pre-training data can affect downstream performance. However, the Boreal3D dataset primarily focuses on boreal forest ecosystems, particularly boreal conifers such as *Pinus* and *Picea*. This limited representation of tree architectures constrains broader generalization across forest structures. Future enhancements could involve extending Boreal3D to incorporate a broader range of tree species and structural forms, thereby improving the model's generalization and applicability across diverse forest environments. More broadly, future progress in cross-platform forest point cloud segmentation is likely to depend more on expanding the diversity of real-world data than on simply increasing the volume of synthetic data. In particular, incorporating more heterogeneous real-world datasets, potentially through weakly supervised, semi-supervised, or partially annotated settings, may provide a more effective path toward improving representation learning and cross-platform generalization.

More specifically, an interesting observation regarding semantic segmentation involves suboptimal performance for the *wood* category, particularly the stem class. Intuitively, linear and distinctive woody structures are assumed to be easier to segment. However, segmentation performance for this class remains unexpectedly low across various methods, including SenNet (43% IoU) and ForAINet (55.6% IoU). Our framework, although improved, still achieved only moderate accuracy (66.5% IoU). This suggests a persistent limitation in representing linear structural features, which should be addressed in future work through modules tailored to woody vegetation. Regarding instance segmentation, while MST exhibited stable cross-platform generalization, performance notably declined in complex, mixed-species, multi-layered forest stands, such as the TU Wien plot in the FOR-Instance dataset. The substantial structural heterogeneity among individual trees, along with the stratified vegetation structure, likely contributes to segmentation difficulties. This highlights the need for more advanced architectures and morphological priors for complex forest stands. Furthermore, the extent to which differing scanning configurations induce data heterogeneity and negative transfer in forest point clouds warrants further investigation. Differences in scanning parameter settings, for example, scanning mechanism (e.g., rotating mirror, fiber array, oscillating mirror, or conic mirror), scan angle (vertical or oblique), and scan altitude (low or high platforms), can markedly affect spatial resolution, point density distributions, and occlusion patterns. These factors influence the visibility of structural elements and alter feature distributions, thereby affecting model generalization. Beyond platform- and sensor-related factors, heterogeneity across forest types (e.g., broadleaf, coniferous, or mixed forests) may also contribute to negative transfer. Future work should explore conditioning models on scanning settings and perform fine-grained feature alignment, as these are promising avenues for enhancing model transferability.

Lastly, exploring the practical implications of the MST framework within forestry remains essential. Pre-trained on synthetic multi-platform forests, the MST learns cross-platform shared structural representations and achieves competitive accuracy even with limited labeled data. This capability allows MST to be readily deployed for forest inventories across plot, regional, and national scales. Through semantic and instance segmentation, MST provides key biophysical attributes, which feed allometric models for biomass and carbon stock estimation.

6. Conclusion

Deep learning has revolutionized forestry applications, largely facilitated by the increasing availability of well-annotated forest point cloud datasets collected from diverse LiDAR platforms. However, notable inter-platform heterogeneity poses considerable challenges, as models trained on platform-specific datasets typically exhibit limited generalization when applied to other platforms. Traditional methodologies that integrate multi-platform data into mixed training suffer from negative transfer, ultimately degrading segmentation performance. To address this critical limitation, this study introduced a novel, model- and data-driven, multi-platform synergistic training framework for generalized forest point cloud segmentation across datasets from different LiDAR platforms. Extensive experimental evaluations conducted across multi-platform forest point cloud benchmarks have demonstrated that MST consistently achieves exceptional performance in both semantic segmentation and instance segmentation tasks. Such consistent results across diverse benchmarks indicate the effectiveness of MST in capturing generalized, ecologically coherent structural representations at both the tree and component levels. Moreover, MST shows strong transferability to TLS data despite the complete absence of TLS-specific training samples during pre-training. Notably, MST can achieve segmentation accuracy comparable to that of fully supervised models while using only 20% of the annotated real-world data, significantly reducing annotation burden. Overall, the proposed MST framework represents a meaningful advancement in forest scene segmentation. It offers a scalable and environmentally sound solution to support efficient cross-

platform LiDAR-based forest inventories, thereby promoting sustainable forest management practices.

CRediT authorship contribution statement

Jundi Jiang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Yueqian Shen:** Writing – review & editing, Validation, Supervision, Software, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Jinhu Wang:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Funding acquisition, Formal analysis, Data curation. **W. Daniel Kissling:** Writing – review & editing, Supervision. **Markus Hollaus:** Writing – review & editing, Visualization, Validation. **Hongjun Su:** Supervision, Resources, Project administration. **Jinguo Wang:** Supervision, Project administration. **Vagner Ferreira:** Supervision, Project administration, Funding acquisition. **Norbert Pfeifer:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis.

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.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (Grant No. 42571513, 42201487, W2432026), and the Fundamental Research Funds for the Central Universities (Grant No. B250205035). Jinhu Wang and W. Daniel Kissling acknowledge funding from the European Commission for the MAMBO project (grant agreement number 101060639). Norbert Pfeifer acknowledges funding from the I-DEAL project supported by the European Union's Horizon Europe research and innovation funding programme under the Marie Skłodowska-Curie GA no. 101236355. The authors are grateful to the High-Performance Computing platform at Hohai University for processing point cloud data in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2026.115467.

Data availability

The data used in this study is publicly available at the following links: Boreal3D dataset: https://boreal3d.github.io/; FOR-Instance dataset: https://zenodo.org/records/8287792; Lin3D dataset: https://github.com/bjfu-lidar/Lin3D-Large-scale-forest-scene-INterpretation-3D-point-cloud-dataset; TreeLearn dataset: https://data.goettingen-research-online.de/dataset.xhtml?persistentId=doi:10.25625/VPMPID; NIBIO_MLS dataset: is available at: https://zenodo.org/records/12754726; SYSSIFOSS dataset: https://doi.pangaea.de/10.1594/PANGAEA.942856?format=html#download; Wytham Woods dataset: https://zenodo.org/records/7307956; Yuchen dataset: https://zenodo.org/records/8398853. The code for the Multi-platform Synergistic Training (MST) framework is available at: https://github.com/jdjiang312/MST.

References

Ali, M., Lohani, B., Hollaus, M., Pfeifer, N., 2025. A hybrid approach for enhanced tree volume estimation of complex trees using terrestrial LiDAR. GISci. Remote Sens. 62, 2474836.

Amiri, N., Polewski, P., Yao, W., Krzystek, P., Skidmore, A., 2017. Detection of single tree stems in forested areas from high density ALS point clouds using 3D shape descriptors. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. 4, 35–42.

Bai, Y., Durand, J.-B., Vincent, G., Forbes, F., 2023. Semantic segmentation of sparse irregular point clouds for leaf/wood discrimination. Adv. Neural Inf. Proces. Syst. 36, 48293–48313.

Balestra, M., Marselis, S., Sankey, T.T., Cabo, C., Liang, X., Mokroš, M., Peng, X., Singh, A., Stereńczak, K., Vega, C., Markus, H., 2024. LiDAR data fusion to improve forest attribute estimates: a review. Curr. For. Rep. 10, 281–297.

Berman, M., Triki, A.R., Blaschko, M.B., 2018. The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4413–4421.

Beyene, S.M., Hussin, Y.A., Kloosterman, H.E., Ismail, M.H., 2020. Forest inventory and aboveground biomass estimation with terrestrial LiDAR in the tropical forest of Malaysia. Can. J. Remote. Sens. 46, 130–145.

Bruggisser, M., Wang, Z., Ginzler, C., Webster, C., Waser, L.T., 2024. Characterization of forest edge structure from airborne laser scanning data. Ecol. Indic. 159, 111624.

Calders, K., Verbeeck, H., Burt, A., Origo, N., Nightingale, J., Malhi, Y., Wilkes, P., Raumonen, P., Bunce, R.G., Disney, M., 2022. Laser scanning reveals potential underestimation of biomass carbon in temperate forest. Ecol. Solut. Evid. 3, e12197.

Caruana, R., 1997. Multitask learning. Mach. Learn. 28, 41–75.

Chang, W.-G., You, T., Seo, S., Kwak, S., Han, B., 2019. Domain-specific batch normalization for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7354–7362.

Chang, L., Fan, H., Zhu, N., Dong, Z., 2022a. A two-stage approach for individual tree segmentation from TLS point clouds. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 8682–8693.

Chang, R., Wang, Y.-X., Ertekin, E., 2022b. Towards overcoming data scarcity in materials science: unifying models and datasets with a mixture of experts framework. npj Comput. Mater. 8, 242.

Chen, X., Jiang, K., Zhu, Y., Wang, X., Yun, T., 2021. Individual tree crown segmentation directly from UAV-borne LiDAR data using the PointNet of deep learning. Forests 12, 131.

Chen, Q., Zhang, Z., Chen, S., Wen, S., Ma, H., Xu, Z., 2022. A self-attention based global feature enhancing network for semantic segmentation of large-scale urban street-level point clouds. Int. J. Appl. Earth Obs. Geoinf. 113, 102974.

Chen, S., Hou, W., Khan, S., Khan, F.S., 2024a. Progressive semantic-guided vision transformer for zero-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 23964–23974.

Chen, X., Yin, H., Chen, Q., Chen, L., Shen, C., 2024b. Multi-source subdomain negative transfer suppression and multiple pseudo-labels guidance alignment: a method for fault diagnosis under cross-working conditions. ISA Trans. 154, 389–406.

Coops, N.C., Tompalski, P., Goodbody, T.R., Queinnec, M., Luther, J.E., Bolton, D.K., White, J.C., Wulder, M.A., van Lier, O.R., Hermosilla, T., 2021. Modelling Lidar-derived estimates of forest attributes over space and time: a review of approaches and future trends. Remote Sens. Environ. 260, 112477.

de Paula Pires, R., Olofsson, K., Persson, H.J., Lindberg, E., Holmgren, J., 2022. Individual tree detection and estimation of stem attributes with mobile laser scanning along boreal forest roads. ISPRS J. Photogramm. Remote Sens. 187, 211–224.

Ding, W., Huang, R., Yao, W., Zhang, W., Heurich, M., Tong, X., 2025. A simple oriented search and clustering method for extracting individual forest trees from ALS point clouds. Ecol. Inform. 86, 102978.

Dobbs, H., Batchelor, O., Green, R., Atlas, J., 2023. Smart-tree: Neural medial axis approximation of point clouds for 3D tree skeletonization. In: Iberian Conference on Pattern Recognition and Image Analysis. Springer, pp. 351–362.

Dong, Y., Ma, Z., Xu, F., Chen, F., 2022. Unsupervised semantic segmenting TLS data of individual tree based on smoothness constraint using open-source datasets. IEEE Trans. Geosci. Remote Sens. 60, 1–15.

dos Santos, J.M.F., Vega, P.J.S., Mota, G.L.A., da Costa, G.A.O.P., 2025. Adversarial domain adaptation for deforestation detection in remote sensing imagery. Ecol. Inform. 89, 103124.

El Mendili, L., Daniel, S., Badard, T., 2025. Context-aware feature adaptation for mitigating negative transfer in 3D LiDAR semantic segmentation. Remote Sens. 17, 2825.

Eldar, Y., Lindenbaum, M., Porat, M., Zeevi, Y.Y., 1997. The farthest point strategy for progressive image sampling. IEEE Trans. Image Process. 6, 1305–1315.

Fekry, R., Yao, W., Cao, L., Shen, X., 2022. Ground-based/UAV-LiDAR data fusion for quantitative structure modeling and tree parameter retrieval in subtropical planted forest. For. Ecosyst. 9, 100065.

Gaikadi, S., Selvaraj, V.K., 2024. Allometric model based estimation of biomass and carbon stock for individual and overlapping trees using terrestrial LiDAR. Model. Earth Syst. Environ. 10, 1771–1782.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. J. Mach. Learn. Res. 17, 1–35.

Gavilán-Acuna, G., Coops, N.C., Tompalski, P., Mena-Quijada, P., Varhola, A., Roeser, D., Olmedo, G.F., 2024. Characterizing annual leaf area index changes and volume growth using ALS and satellite data in forest plantations. Sci. Remote Sens. 10, 100159.

Ghifary, M., Kleijn, W.B., Zhang, M., 2014. Domain adaptive neural networks for object recognition. In: Pacific Rim International Conference on Artificial Intelligence. Springer, pp. 898–904.

Ghorbani, F., Chen, Y.-C., Hollaus, M., Pfeifer, N., 2024. A robust and automatic algorithm for TLS–ALS point cloud registration in forest environments based on tree locations. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 4015–4035.
Gong, M., Zhang, K., Liu, T., Tao, D., Glymour, C., Schölkopf, B., 2016. Domain adaptation with conditional transferable components. In: International Conference on Machine Learning. PMLR, pp. 2839–2848.
González-Quiñones, J.J., Polidori, L., Ariza-López, F.J., Ureña-Cámara, M.A., Reinoso-Gordo, J.F., 2024. Influence of tree density and terrain slope on ground point density in LiDAR point clouds: a simulation-based study with Helios++. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 48, 197–202.
Han, X., Liu, C., Zhou, Y., Tan, K., Dong, Z., Yang, B., 2024. WHU-Urban3D: An urban scene LiDAR point cloud dataset for semantic instance segmentation. ISPRS J. Photogramm. Remote Sens. 209, 500–513.
Henrich, J., van Delden, J., Seidel, D., Kneib, T., Ecker, A.S., 2024. TreeLearn: a deep learning method for segmenting individual trees from ground-based LiDAR forest point clouds. Eco. Inform. 84, 102888.
Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S., 2019. Parameter-efficient transfer learning for NLP. In: International Conference on Machine Learning. PMLR, pp. 2790–2799.
Huang, W., Chen, C., Li, Y., Li, J., Li, C., Song, F., Yan, Y., Xiong, Z., 2023. Style projected clustering for domain generalized semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3061–3071.
Hui, Z., Jin, S., Xia, Y., Wang, L., Ziggah, Y.Y., Cheng, P., 2021. Wood and leaf separation from terrestrial LiDAR point clouds based on mode points evolution. ISPRS J. Photogramm. Remote Sens. 178, 219–239.
Jarahizadeh, S., Salehi, B., 2025. Advancing tree detection in forest environments: A deep learning object detector approach with UAV LiDAR data. Urban For. Urban Green. 105, 128695.
Jiang, L., Zhao, H., Shi, S., Liu, S., Fu, C.-W., Jia, J., 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.
Jiang, T., Zhang, Q., Liu, S., Liang, C., Dai, L., Zhang, Z., Sun, J., Wang, Y., 2023. LWSNet: a point-based segmentation network for leaf-wood separation of individual trees. Forests 14, 1303.
Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., Stuetzle, W., 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. Remote Sens. Environ. 113, 1148–1162.
Komárek, J., Lagner, O., Klouček, T., 2024. UAV leaf-on, leaf-off and ALS-aided tree height: a case study on the trees in the vicinity of roads. Urban For. Urban Green. 93, 128229.
Kükenbrink, D., Marty, M., Rehush, N., Abegg, M., Ginzler, C., 2025. Evaluating the potential of handheld mobile laser scanning for an operational inclusion in a national forest inventory–a Swiss case study. Remote Sens. Environ. 321, 114685.
Lahoud, J., Ghanem, B., Pollefeys, M., Oswald, M.R., 2019. 3d instance segmentation via multi-task metric learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9256–9266.
Lei, L., Chai, G., Yao, Z., Li, Y., Jia, X., Zhang, X., 2025. A novel self-similarity cluster grouping approach for individual tree crown segmentation using multi-features from UAV-based LiDAR and multi-angle photogrammetry data. Remote Sens. Environ. 318, 114588.
Li, S., Fang, H., 2025. Mapping global leaf inclination angle (LIA) based on field measurement data. Earth Syst. Sci. Data 17, 1347–1366.
Li, W., Guo, Q., Jakubowski, M.K., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud. Photogramm. Eng. Rem. Sensing 78, 75–84.
Li, Y., Wang, N., Shi, J., Liu, J., Hou, X., 2016. Revisiting batch normalization for practical domain adaptation. arXiv preprint. *arXiv:1603.04779*.
Li, Y., Wang, N., Shi, J., Hou, X., Liu, J., 2018. Adaptive batch normalization for practical domain adaptation. Pattern Recogn. 80, 109–117.
Li, G., Kang, G., Wang, X., Wei, Y., Yang, Y., 2023a. Adversarially masking synthetic to mimic real: Adaptive noise injection for point cloud segmentation adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20464–20474.
Li, Y., Xie, D., Wang, Y., Jin, S., Zhou, K., Zhang, Z., Li, W., Zhang, W., Mu, X., Yan, G., 2023b. Individual tree segmentation of airborne and UAV LiDAR point clouds based on the watershed and optimized connection center evolution clustering. Ecol. Evol. 13, e10297.
Liang, X., Qi, H., Deng, X., Chen, J., Cai, S., Zhang, Q., Wang, Y., Kukko, A., Hyyppä, J., 2025. ForestSemantic: a dataset for semantic learning of forest from close-range sensing. Geo-spat. Inf. Sci. 28, 185–211.
Lin, Y., Liu, J., Zhou, J., 2020. A novel tree-structured point cloud dataset for skeletonization algorithm evaluation. arXiv preprint arXiv:2001.02823.
Liu, J., Wang, D., Gong, H., Wang, C., Zhu, J., Wang, D., 2026. A synthetic data generation framework for deep learning-based LiDAR forest structure analysis. Remote. Sens. Environ. 341, 115436.
Liu, Y., Chen, D., Na, J., Peethambaran, J., Pfeifer, N., Zhang, L., 2025. Segmentation of individual trees in TLS point clouds via graph optimization. IEEE Trans. Geosci. Remote Sens. 63, 1–21.
Lu, J., Deng, J., 2025. Relation3D: enhancing relation modeling for point cloud instance segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 8889–8899.
Lu, D., Jiang, X., 2024. A brief overview and perspective of using airborne Lidar data for forest biomass estimation. Int. J. Image Data Fusion 15, 1–24.
Lu, X., Guo, Q., Li, W., Flanagan, J., 2014. A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data. ISPRS J. Photogramm. Remote Sens. 94, 1–12.
Lu, J., Deng, J., Wang, C., He, J., Zhang, T., 2023. Query refinement transformer for 3d instance segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 18516–18526.
Lu, Y., Sun, Z., Shao, J., Guo, Q., Huang, Y., Fei, S., Chen, V., 2024. Lidar-forest dataset: Lidar point cloud simulation dataset for forestry application. In: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, pp. 112–116.
Lu, H., Li, B., Yang, G., Fan, G., Wang, H., Pang, Y., Wang, Z., Lian, Y., Xu, H., Huang, H., 2025. Towards a point cloud understanding framework for forest scene semantic segmentation across forest types and sensor platforms. Remote Sens. Environ. 318, 114591.
Ma, Y., Zhao, Y., Im, J., Zhao, Y., Zhen, Z., 2024. A deep-learning-based tree species classification for natural secondary forests using unmanned aerial vehicle hyperspectral images and LiDAR. Ecol. Indic. 159, 111608.
Maes, J., Bruzón, A.G., Barredo, J.I., Vallecillo, S., Vogt, P., Rivero, I.M., Santos-Martín, F., 2023. Accounting for forest condition in Europe based on an international statistical standard. Nat. Commun. 14, 3723.
Meng, X., Zeng, J., Yang, Y., Zhao, W., Ma, H., Letu, H., Zhu, Q., Liu, Y., Wang, P., Peng, J., 2024. High-resolution soil moisture mapping through passive microwave remote sensing downscaling. Innov. Geosci. 2, 100105.
Mo, L., Zohner, C.M., Reich, P.B., Liang, J., De Miguel, S., Nabuurs, G.-J., Renner, S.S., Van Den Hoogen, J., Araza, A., Herold, M., 2023. Integrated global assessment of the natural forest carbon potential. Nature 624, 92–101.
Ngo, T.D., Hua, B.-S., Nguyen, K., 2023. Isbnet: a 3d point cloud instance segmentation network with instance-aware sampling and box-aware dynamic convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13550–13559.
Oehmcke, S., Li, L., Trepekli, K., Revenga, J.C., Nord-Larsen, T., Gieseke, F., Igel, C., 2024. Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR. Remote Sens. Environ. 302, 113968.
Panagiotidis, D., Abdollahnejad, A., Slavík, M., 2022. 3D point cloud fusion from UAV and TLS to assess temperate managed forest structures. Int. J. Appl. Earth Obs. Geoinf. 112, 102917.
Paris, C., Valduga, D., Bruzzone, L., 2016. A hierarchical approach to three-dimensional segmentation of LiDAR data at single-tree level in a multilayered forest. IEEE Trans. Geosci. Remote Sens. 54, 4190–4203.
Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A., 2018. Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence.
Polewski, P., Yao, W., Cao, L., Gao, S., 2019. Marker-free coregistration of UAV and backpack LiDAR point clouds in forested areas. ISPRS J. Photogramm. Remote Sens. 147, 307–318.
Puliti, S., Pearse, G., Surový, P., Wallace, L., Hollaus, M., Wielgosz, M., Astrup, R., 2023. For-instance: a UAV laser scanning benchmark dataset for semantic and instance segmentation of individual trees. arXiv preprint arXiv:2309.01279.
Qin, C., You, H., Wang, L., Kuo, C.-C.J., Fu, Y., 2019. Pointdan: a multi-scale 3D domain adaption network for point cloud representation. Adv. Neural Inf. Proces. Syst. 32.
Rauch, L., Braml, T., 2025. Multi-dataset synergistic in supervised learning to pre-label structural components in point clouds from shell construction scenes. arXiv preprint. *arXiv:2502.14721*.
Rodda, S.R., Nidamanuri, R.R., Mayamanikandan, T., Rajashekar, G., Jha, C.S., Dadhwal, V.K., 2024. Non-destructive allometric modeling for tree volume estimation in tropical dry deciduous forests of India using terrestrial laser scanner. J. Indian Soc. Remote Sens. 52, 825–839.
Ruoppa, L., Oinonen, O., Taher, J., Lehtomäki, M., Takhtkeshha, N., Kukko, A., Kaartinen, H., Hyyppä, J., 2025. Unsupervised deep learning for semantic segmentation of multispectral LiDAR forest point clouds. arXiv preprint. *arXiv:2502.06227*.
Shao, J., Yao, W., Wan, P., Luo, L., Wang, P., Yang, L., Lyu, J., Zhang, W., 2022. Efficient co-registration of UAV and ground LiDAR forest point clouds based on canopy shapes. Int. J. Appl. Earth Obs. Geoinf. 114, 103067.
Shao, J., Habib, A., Fei, S., 2023. Semantic segmentation of UAV Lidar data for tree plantations. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 48, 1901–1906.
Shen, Y., Ji, S., Wang, J., Liu, W., Wang, J., Chen, Y., Deng, Z., Fu, S., Chen, D., 2024. Coarse-to-fine separation of wood and leaf from MLS street tree point clouds using branch tilt prior and enhanced shortest path ttacing. IEEE Trans. Geosci. Remote Sens. 62, 1–18.
Shendryk, I., Broich, M., Tulbure, M.G., Alexandrov, S.V., 2016. Bottom-up delineation of individual trees from full-waveform airborne laser scans in a structurally complex eucalypt forest. Remote Sens. Environ. 173, 69–83.
Straker, A., Puliti, S., Breidenbach, J., Kleinn, C., Pearse, G., Astrup, R., Magdon, P., 2023. Instance segmentation of individual tree crowns with YOLOv5: a comparison of approaches using the ForInstance benchmark LiDAR dataset. ISPRS Open J. Photogramm. Remote Sens. 9, 100045.
Sun, A., Su, R., Ma, J., Lin, J., 2025. Individual trunk segmentation and diameter at breast height estimation using mobile LiDAR scanning. Forests 16, 582.
Tang, S., Ao, Z., Li, Y., Huang, H., Xie, L., Wang, R., Wang, W., Guo, R., 2024. TreeNet3D: a large scale tree benchmark for 3D tree modeling, carbon storage estimation and tree segmentation. Int. J. Appl. Earth Obs. Geoinf. 130, 103903.
Tao, S., Wu, F., Guo, Q., Wang, Y., Li, W., Xue, B., Hu, X., Li, P., Tian, D., Li, C., 2015. Segmenting tree crowns from terrestrial and mobile LiDAR data by exploring ecological theories. ISPRS J. Photogramm. Remote Sens. 110, 66–76.
Tong, F., Zhang, Y., 2025. Individual tree crown delineation in high resolution aerial RGB imagery using StarDist-based model. Remote Sens. Environ. 319, 114618.
Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9.

Vicari, M.B., Disney, M., Wilkes, P., Burt, A., Calders, K., Woodgate, W., 2019. Leaf and wood classification framework for terrestrial LiDAR point clouds. Methods Ecol. Evol. 10, 680–694.
Vu, T., Kim, K., Luu, T.M., Nguyen, T., Yoo, C.D., 2022. Softgroup for 3d instance segmentation on point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2708–2717.
Wan, P., Shao, J., Jin, S., Wang, T., Yang, S., Yan, G., Zhang, W., 2021. A novel and efficient method for wood–leaf separation from terrestrial laser scanning point clouds at the forest plot level. Methods Ecol. Evol. 12, 2473–2486.
Wang, D., 2020. Unsupervised semantic and instance segmentation of forest point clouds. ISPRS J. Photogramm. Remote Sens. 165, 86–97.
Wang, M., Deng, W., 2018. Deep visual domain adaptation: a survey. Neurocomputing 312, 135–153.
Wang, P., Yao, W., 2022. A new weakly supervised approach for ALS point cloud semantic segmentation. ISPRS J. Photogramm. Remote Sens. 188, 237–254.
Wang, Z., Dai, Z., Póczos, B., Carbonell, J., 2019. Characterizing and avoiding negative transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11293–11302.
Wang, D., Momo Takoudjou, S., Casella, E., 2020. LeWoS: a universal leaf-wood classification method to facilitate the 3D modelling of large tropical trees using terrestrial LiDAR. Methods Ecol. Evol. 11, 376–389.
Wang, D., Puttonen, E., Casella, E., 2022a. PlantMove: a tool for quantifying motion fields of plant movements from point cloud time series. Int. J. Appl. Earth Obs. Geoinf. 110, 102781.
Wang, H., Tao, C., Qi, J., Xiao, R., Li, H., 2022b. Avoiding negative transfer for semantic segmentation of remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–15.
Wang, L., Li, D., Liu, H., Peng, J., Tian, L., Shan, Y., 2022c. Cross-dataset collaborative learning for semantic segmentation in autonomous driving. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 2487–2494.
Wang, P., Yao, W., Shao, J., 2023. One class one click: quasi scene-level weakly supervised point cloud semantic segmentation with active learning. ISPRS J. Photogramm. Remote Sens. 204, 89–104.
Wang, L., Lu, D., Xu, L., Robinson, D.T., Tan, W., Xie, Q., Guan, H., Chapman, M.A., Li, J., 2024. Individual tree species classification using low-density airborne multispectral LiDAR data via attribute-aware cross-branch transformer. Remote Sens. Environ. 315, 114456.
Wang, P., Yao, W., Shao, J., He, Z., 2025. Test-time adaptation for geospatial point cloud semantic segmentation with distinct domain shifts. ISPRS J. Photogramm. Remote Sens. 229, 422–435.
Weiser, H., Schäfer, J., Winiwarter, L., Krašovec, N., Fassnacht, F.E., Höfle, B., 2022. Individual tree point clouds and tree measurements from multi-platform laser scanning in German forests. Earth Syst. Sci. Data 14, 2989–3012.
Wielgosz, M., Puliti, S., Wilkes, P., Astrup, R., 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.
Wielgosz, M., Puliti, S., Xiang, B., Schindler, K., Astrup, R., 2024. SegmentAnyTree: a sensor and platform agnostic deep learning model for tree segmentation using laser scanning data. Remote Sens. Environ. 313, 114367.
Wild, B., Özkan, T., Ali, M., Pöppl, F., Milenković, M., Hofhansl, F., Pfeifer, N., Lau, A., Hollaus, M., 2026. Evaluating RayCloudTools to estimate single-tree volume. Forestry 99, cpaf087.
Wilkes, P., Disney, M., Armston, J., Bartholomeus, H., Bentley, L., Brede, B., Burt, A., Calders, K., Chavana-Bryant, C., Clewley, D., 2023. TLS2trees: a scalable tree segmentation pipeline for TLS data. Methods Ecol. Evol. 14, 3083–3099.
Winiwarter, L., Pena, A.M.E., Weiser, H., Anders, K., Sánchez, J.M., Searle, M., Höfle, B., 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.
Wu, X., Jiang, L., Wang, P.-S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H., 2024a. Point transformer v3: simpler faster stronger. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4840–4851.
Wu, X., Tian, Z., Wen, X., Peng, B., Liu, X., Yu, K., Zhao, H., 2024b. Towards large-scale 3d representation learning with multi-dataset point prompt training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 19551–19562.
Xi, Z., Hopkinson, C., 2022. 3D graph-based individual-tree isolation (Treeiso) from terrestrial laser scanning point clouds. Remote Sens. 14, 6116.
Xiang, B., Wielgosz, M., Kontogianni, T., Peters, T., Puliti, S., Astrup, R., Schindler, K., 2024. Automated forest inventory: analysis of high-density airborne LiDAR point clouds with 3D deep learning. Remote Sens. Environ. 305, 114078.
Xiang, B., Wielgosz, M., Puliti, S., Král, K., Krůček, M., Missarov, A., Astrup, R., 2025a. ForestFormer3D: a unified framework for end-to-end segmentation of forest LiDAR 3D point clouds. arXiv preprint arXiv:2506.16991.
Xiang, B., Wielgosz, M., Puliti, S., Král, K., Krůček, M., Missarov, A., Astrup, R., 2025b. Forestformer3d: a unified framework for end-to-end segmentation of forest lidar 3d point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 24717–24727.
Xiao, A., Huang, J., Guan, D., Zhan, F., Lu, S., 2022. Transfer learning from synthetic to real lidar point cloud for semantic segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 2795–2803.
Xu, S., Zhou, K., Sun, Y., Yun, T., 2021. Separation of wood and foliage for trees from ground point clouds using a novel least-cost path model. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 6414–6425.
Yang, J., Kang, Z., Cheng, S., Yang, Z., Akwensi, P.H., 2020. An individual tree segmentation method based on watershed algorithm and three-dimensional spatial distribution analysis from airborne LiDAR point clouds. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 1055–1067.
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. Remote Sens. Environ. 123, 368–380.
Yun, T., Jiang, K., Li, G., Eichhorn, M.P., Fan, J., Liu, F., Chen, B., An, F., Cao, L., 2021. Individual tree crown segmentation from airborne LiDAR data using a novel Gaussian filter and energy function minimization-based approach. Remote Sens. Environ. 256, 112307.
Zeng, J., Shen, X., Zhou, K., Cao, L., 2025. FO-net: an advanced deep learning network for individual tree identification using UAV high-resolution images. ISPRS J. Photogramm. Remote Sens. 220, 323–338.
Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., Yan, G., 2016. An easy-to-use airborne LiDAR data filtering method based on cloth simulation. Remote Sens. 8, 501.
Zhang, Z., Alwen, A., Lyu, H., Liu, X., Li, Z., Xie, Z., Xie, Y., Guan, F., Babakhani, A., Pei, Q., 2019. Stretchable transparent wireless charging coil fabricated by negative transfer printing. ACS Appl. Mater. Interfaces 11, 40677–40684.
Zhang, K., Ye, L., Xiao, W., Sheng, Y., Zhang, S., Tao, X., Zhou, Y., 2022. A dual attention neural network for airborne LiDAR point cloud semantic segmentation. IEEE Trans. Geosci. Remote Sens. 60, 1–17.
Zhong, M., Chen, X., Chen, X., Zeng, G., Wang, Y., 2022. Maskgroup: Hierarchical point grouping and masking for 3d instance segmentation. In: 2022 IEEE International Conference on Multimedia and Expo (ICME). IEEE, pp. 1–6.
Zhong, Y., Qin, J., Liu, S., Ma, Z., Liu, E., Fan, H., 2025. An unsupervised semantic segmentation network for wood-leaf separation from 3D point clouds. Plant Phenom. 7 (2), 100064.