Journal of Forestry
https://doi.org/10.1007/s44392-025-00065-6

ORIGINAL PAPER

Individual Tree Biomass Estimation using Single-Scan Terrestrial Laser Scanner with Efficient Projection-Based Deep Learning

Minyoung Jung[1] · Jaewan Choi[2] · Joshua Carpenter[3] · Songlin Fei[4] · Jinha Jung[1]

Received: 4 May 2025 / Accepted: 27 October 2025

Abstract

Accurate estimation of individual tree aboveground biomass (AGB) is critical for improving our understanding of forest structure and function, especially in forests with diverse tree species spanning a wide range of sizes. Terrestrial laser scanning (TLS) has proven effective at the tree scale thanks to its ability to provide high-density point cloud data. However, most existing methods rely on high-quality multi-scan TLS data, which requires labor-intensive data collection and complicated preprocessing to achieve high precision, thereby limiting their practical applicability. In contrast, single-scan TLS data offers a simpler and more efficient alternative, though it presents significant challenges due to incomplete coverage, uneven point density, and frequent occlusions. This study presents a novel and effective framework for estimating individual tree AGB from single-scan TLS data by integrating a deep learning network (CoAtNet) with three key innovations: an optimized 3D-to-2D projection strategy, a 3D point cloud-based data augmentation, and an ensemble approach to enhance robustness under data-scarce conditions in local natural forest environments. The framework was evaluated in a mixed hardwood forest and demonstrated strong predictive performance with a coefficient of determination (R2) of 0.73, a median percentage error of 0.99%, and a median absolute percentage error of 25.06%, outperforming two comparison models, Random Forest and Point Transformer. The results indicate the potential of single-scan TLS data to support reliable individual tree biomass when combined with deep learning, even in the presence of data imperfections. The proposed framework provides a practical, scalable approach

Study Implications This study demonstrates that single-scan terrestrial laser scanning data, when paired with an appropriately designed deep learning framework, can enable accurate and scalable estimation of individual tree aboveground biomass, particularly in natural hardwood forests. The proposed method offers a practical and scalable alternative to labor-intensive multi-scan approaches, making it well-suited for operational use in diverse and data-limited forest environments. Given the critical role of forests as major carbon sinks, this study provides a valuable tool for improving forest ecosystem monitoring and management, thereby contributing to climate change mitigation efforts.

Extended author information available on the last page of the article

Published online: 19 November 2025

that contributes to the advancement of remote sensing methodologies for sustainable forest management and facilitates improved monitoring of ecological changes in the context of accelerating climate change.

Keywords Aboveground biomass (AGB) · Single-scan terrestrial laser scanner (TLS) · Deep learning · Natural hardwood forest

Introduction

Forests serve as the primary terrestrial carbon sink and have received substantial attention in recent decades due to the escalating impacts of climate change on global ecosystems. Numerous studies have been dedicated to estimating forest biomass across various spatial scales, ranging from regional to global, given its critical role as an indicator of forest carbon flux (Houghton 2005; Lu et al. 2016). Forest biomass comprises both above- and below-ground biomass; however, due to the inherent challenges of measuring belowground biomass, most forest biomass studies focus on aboveground biomass (AGB) (Lu et al. 2016). While traditional destructive sampling methods remain the most primary means of quantifying tree AGB, their applicability is limited to harvested trees. Alternatively, non-destructive approaches, represented as allometric models using non-destructively measurable tree attributes, such as diameter at breast height (DBH) and height, have been developed and applied to enhance our understanding of AGB in living, conservated forests (Chojnacky et al. 2014; Sileshi 2014).

Recent advancements in remote sensing technologies have further revolutionized non-destructive approaches, enabling spatially comprehensive and highly accurate assessments of forest AGB. Among the various remote sensing platforms, light detection and ranging (lidar) has been widely used in AGB research across multiple modalities, including spaceborne platforms, airborne platforms, and terrestrial laser scanners (TLS), owing to its ability to capture detailed structural information of forest canopies and tree architecture. Numerous studies have demonstrated the utility of spaceborne and airborne lidar data for large-scale AGB estimation (Gleason and Im 2012; Cao et al. 2016; Dorado-Roda et al. 2021; Dubayah et al. 2022; Shendryk 2022). However, large-scale biomass estimates cannot sufficiently provide insights into local ecology (e.g., individual tree variability), especially for areas characterized by diverse tree species with varying sizes. Accordingly, the importance of individual-tree-level analysis has been increasingly recognized, with using unmanned aerial system (UAS) and TLS data, due to their exceptionally high point density, emerging as crucial tools for precise tree delineation and AGB estimation (Kankare et al. 2013; Lu et al. 2020; Brede et al. 2022; Lian et al. 2022; Lin et al. 2022; Zhuo et al. 2022).

TLS is particularly suited for detailed tree structure characterization and larger-scale validation studies (Duncanson et al. 2019). As collected from the ground, TLS can provide detailed structural information of the tree components, such as trunks, that exhibit strong correlations with biomass. Previous studies

have demonstrated that key tree attributes required for allometric AGB estimation can be extracted from TLS data accurately (Kankare et al. 2013; Srinivasan et al. 2014). Furthermore, several studies have advanced TLS-based AGB estimation methodologies by reconstructing detailed three-dimensional (3D) models of entire trees, including not only major structural elements (e.g., trunks) but also finer branches (Kaasalainen et al. 2014; Calders et al. 2015; Fan et al. 2020; Kükenbrink et al. 2021). These reconstructions enable precise tree volume measurements, leading to highly accurate AGB estimates when multiplied by species-specific wood density values. These AGB estimates could be regarded as reliable reference datasets for developing an AGB estimation framework at broader regional scales, such as UAS lidar (Brede et al. 2022). However, such reconstruction methods typically require high-quality point cloud data, such as that with high point density and fewer occlusions, to guarantee precise results, and thus have usually been applied to multi-scan TLS data. The quality of multi-scan TLS data directly affects the analysis results, as Burt et al. (2013) and Demol et al (2022) demonstrated that registration errors result in overestimating tree volume. Therefore, labor-intensive data acquisition and processing, such as multiple scanning from different positions and precisely registering them as guided by Wilkes et al. (2017), are involved, thereby raising practical difficulties with applying the prior reconstruction methods to local-scale forest investigation.

In contrast, single-scan TLS data can be an attractive alternative due to its relatively easy acquisition (Liang et al. 2016). It suffers from data quality issues such as frequent occlusions and irregular point density, which limit its broader applications, including AGB estimation. Nevertheless, this type of data retains valuable structural information, and several studies have demonstrated its potential (Liang et al. 2016; Seidel et al. 2016; Batchelor et al. 2023). In particular, Astrup et al. (2014) demonstrated that individual tree volumes could be estimated using single-scan TLS, mainly for softwood species, such as Scots pine (*Pinus sylvestris L.*) and Norway spruce (*Picea abies L. Karst)*. Liang et al. (2018) noted that single-scan data can provide sufficient information for stem modeling of individual trees that are visible in the scan. These previous studies underscore the potential of single-scan TLS for AGB estimation and suggest the need for innovative approaches to fully unlock its capabilities.

In this context, this study developed a novel deep learning-based approach to estimate individual tree AGB from single-scan TLS data, considering the benefit of deep learning, which has demonstrated considerable success in forest research (Liu et al. 2022; Oehmcke et al. 2024; Schäfer et al. 2024). Specifically, we designed an estimation model based on the CoAtNet architecture, integrated with three key innovations: (1) an optimized 3D-to-2D (two-dimensional) projection method, (2) a 3D point cloud-based data augmentation strategy, and (3) an ensemble approach to enhance robustness and accuracy. The superiority of the proposed method was demonstrated through comparison with two baseline models, Random Forest and Point Transformer, in the *Experimental Results* section. In addition, the effectiveness of the three components was thoroughly examined through a series of controlled experiments presented in the *Discussion* section. Collectively, this study demonstrates the feasibility of single-scan TLS data for reliable individual tree AGB estimation

and provides a practical, scalable alternative to conventional reconstruction-based approaches relying on multi-scan TLS data.

Data Preparation

Study Site and Field Campaign

The site of this study is an approximately 2-square-kilometer northern mixed hardwood forest, Martell Forest, located in Indiana, USA (Fig. 1; 40.441, −87.033). Martell Forest has been managed for research, education, and conservation purposes by Purdue University since 1958 and mainly comprises forests (70% natural forest and 21% plantation forest) with different land covers, such as open water, prairie, and roads. It has three distinct terrain regions: the flat upland plains, the flat lowland flood plain, and the steep slopes with a 4.2-km stream.

The forest inventory campaign has been consistently conducted in the mixed natural forest of the study site. The most recent field campaign was conducted during the leaf-off season, between March and April 2022. Although 112 historic

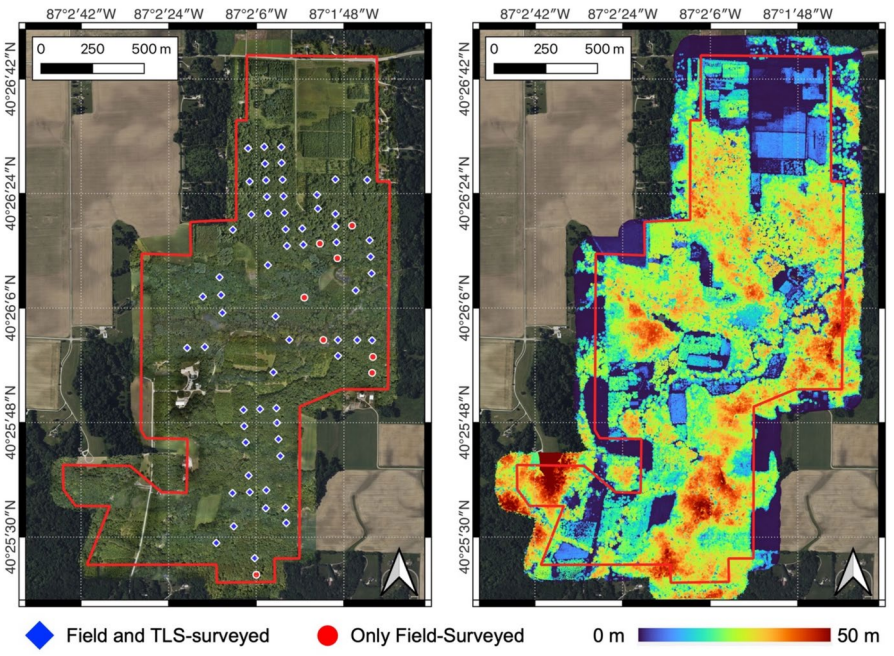

Fig. 1 Study site (Martell Forest, Indiana, USA) and distribution of tree inventory plots. (Left) Plots on an orthomosaic image obtained using an unmanned aerial system (UAS). (Right) Canopy height model (CHM) derived from UAS lidar data. The red line denotes the boundary of Martell Forest. Minor errors are present in the CHM near the boundary, particularly in the southeastern area, due to unavoidable restrictions during UAS data acquisition

inventory plots were initially established within the study area, only 69 circular plots (marked in Fig. 1) were surveyed in 2022 because some could not be located conclusively on the ground, and others had been obliterated by erosion. Each field-surveyed plot center was measured using a navigation-grade GNSS receiver, and all living trees larger than 12.7 cm (5 inches) in diameter within 16 m of the plot center were investigated. The tree inventory from the field campaign includes the individual tree locations, DBHs, and species. Tree locations were measured by compass bearing and tape and presented in the azimuth angles to the North and the distances from each plot center.

Through the field campaign, the study site was identified as a natural forest composed exclusively of hardwood species. A total of 37 hardwood species were recorded, with Sugar Maple (*Acer saccharum*) being the most dominant, followed by White Oak (*Quercus alba*). The species composition is shown in Fig. 2(a), where all species other than Sugar Maple and White Oak account for less than 10% of the total number of trees. Considerable variation in tree size was also observed, with DBH ranging from 12.7 cm to 94.5 cm, as illustrated in Fig. 2(b). The basal area and tree density were estimated to be 29.83 m^2/ha and 286.88 trees/ha, respectively. Additionally, canopy height variability across the site was confirmed using a canopy height model (CHM) derived from UAS-lidar data (Fig. 1).

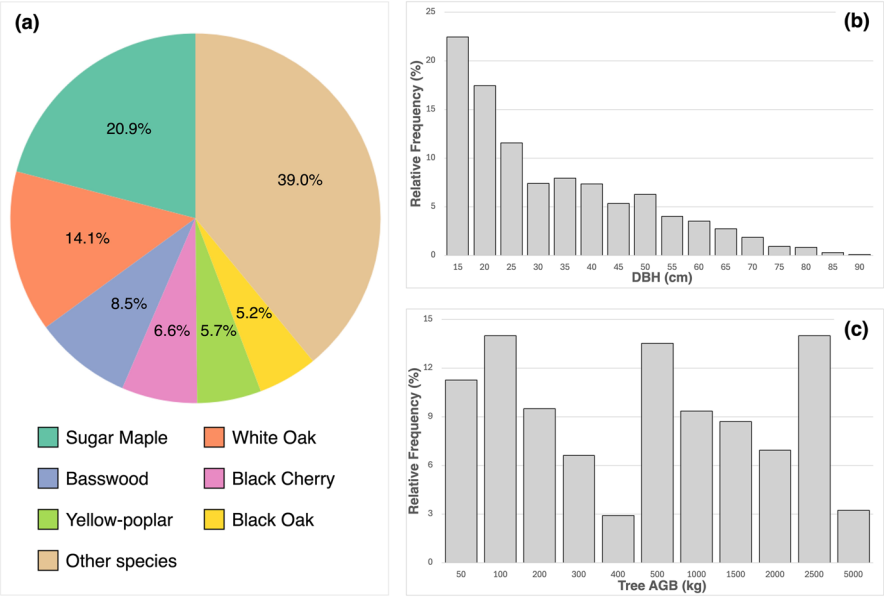

Fig. 2 Distribution of (**a**) tree species, (**b**) diameter at breast height (DBH, cm), and (**c**) aboveground biomass (AGB, kg) of individual trees surveyed during the 2022 field campaign. Tree species representing less than 5% of the total number of surveyed trees (31 species in total) are grouped into a single *Other species* category. The entire species are listed in Table 1. The histograms represent relative frequency (%). Note that the bin sizes in (c) are unequal due to the high variability in tree AGB within the study site

Reference AGB Derived from Field Campaign

AGB estimation was conducted for 37 tree species identified within the tree inventory plots. Due to the absence of locally adapted AGB models that account for the taxonomic and structural diversity of the study site, the generalized AGB models for North American species presented by Chojnacky et al. (2014) were adopted to calculate individual tree AGB based on the DBH measurements from the field campaign. Although previous studies have reported that applying national-scale allometric equations at local scales can introduce considerable estimation errors (Stovall et al. 2017), this limitation was not expected to significantly impact the primary objective of this study, which was to demonstrate the potential of single-scan TLS for individual tree AGB estimation by proposing a novel deep learning approach. Since deep learning relies on supervised learning, any systematic bias in the reference AGB values can be addressed in future applications by training a deep learning model with more accurate, locally adapted equations when available.

The generalized models are expressed as shown in Eq. (1), with species-specific coefficients summarized in Table 1. While the generalized models cover most North American species, certain species, such as Boxelder and Hackberry, were not explicitly modeled. For these species, appropriate generalized models were selected based on basic specific gravity values reported by Miles and Smith (2009) and the recommendations by Hoover and Smith (2016). The AGB values, calculated from DBH using Eq. (1) and the corresponding coefficients (Table 1), were regarded as the dependent variables for the AGB estimation models described in the *AGB Estimation Models* section. Hereafter, these values are referred to as *the reference AGB*.

$$\ln(\text{biomass, kg}) = b_0 + b_1 \ln(\text{diameter, cm}) \tag{1}$$

Individual Tree Point Cloud from Single-scan TLS

Single-scan TLS data were collected in 2022, during the same period as the field campaign, using a FARO Focus3D X330 laser scanner (wavelength: 1,550 nm; beam divergence: 0.011°; beam diameter: 2.25 mm). The instrument provides a full scanning range of 0.6–330 m and a ranging accuracy of ± 2 mm and includes a multi-sensor package (dual-axis compensator, barometric height sensor, electronic compass, and GPS), enabling precise orientation and georeferencing. For each plot, the scanner was positioned directly above each plot center using a plum bob. Point clouds for each plot were generated with the scanner's origin aligned at the plot center and a rough orientation toward the north by the scanner's internal compass. Each scanned point cloud was then cropped to the plot dimensions in post-processing. Due to safety concerns and unstable terrain conditions at some sites, eight plots were excluded from TLS data acquisition. Consequently, TLS data were successfully collected for 61 plots, indicated by blue marks in Fig. 1.

The high point density of the single-scan TLS data enabled effective segmentation of individual trees within the 61 inventory plots. Segmentation was performed

Table 1 Coefficients of the generalized AGB estimation models of Eq. (1) used in this paper

Model based on taxa grouping	Coefficients		Species
	b_0	b_1	
Aceraceae < 0.50 spg	−2.0470	2.3852	Boxelder*
Aceraceae ≥ 0.50 spg	−1.8011	2.3852	Maple (black*, sugar)
Cornaceae/Ericaceae/Lauraceae/Platanaceae/Rosaceae/Ulmaceae	−2.2118	2.4133	American sycamore, Black cherry, Dogwood, Elm (American, red), Hackberry*, Hawthorn*, Red mulberry*, Sassafras
Fabaceae/Juglandaceae, Carya	−2.5095	2.6175	Hickory (bitternut, mockernut, pignut*, shagbark)
Fabaceae/Juglandaceae, other	−2.5095	2.5437	Black walnut*, Eastern redbud*, Ironwood*, Kentucky coffee tree*, Locust (black, honey*)
Fagaceae, deciduous	−2.0705	2.4410	American beech, Oak (black, bur, chinkapin*, red, shingle*, Shumard*, white)
Hippocastanaceae/Tiliaceae	−2.4108	2.4177	Basswood, Ohio buckeye*
Magnoliaceae	−2.5497	2.5011	Yellow-poplar
Oleaceae < 0.55 spg	−2.0314	2.3524	Green ash
Oleaceae ≥ 0.55 spg	−1.8384	2.3524	White ash
Salicaceae ≥ 0.35 spg	−2.4441	2.4561	Bigtooth aspen*, Cottonwood

The coefficients are identical to those in Chojnacky et al. (2014). Species denoted with asterisk marks (*) did not have a generalized estimation model and were assigned to one of Chojnacky et al. (2014) models based on their specific gravities. spg indicates specific gravity

using the unsupervised canopy-to-root pathing (UCRP) algorithm, a graph-based method that applies least-cost routing from canopy points down to the ground to delineate individual trees simultaneously (Carpenter et al. 2022). Unlike many earlier methods, UCRP does not require an explicit trunk detection step. Given its previously demonstrated effectiveness in Martell Forest, the study site, it was expected to perform well on the single-scan TLS data collected from the same region. The user-defined UCRP parameters were initially set to the defaults reported in Carpenter et al. (2022) to ensure consistency with the original implementation. However, since certain plots had heavy ground vegetation, the parameters were adjusted to more conservative values to reduce the amount of ground vegetation incorporated into segmented trees. The voxel-size parameter (*s*) was maintained at the default of 0.3 m, the ground threshold (*g_max*) was reduced from 1.2 m to 1.0 m, and the merge distance (*d*) was reduced from 0.9 m to 0.7 m.

Figure 3(a) provides examples of individually segmented trees within a 16-m-radius inventory plot. Both large mature trees and smaller trees with DBHs less than 12.7 cm (which were not included in the field inventory) were successfully delineated. The presence of understory shrubs varied slightly across plots; however, as illustrated in this example, shrubs were not present at a density during the leaf-off season that would interfere with segmentation from a single scan for most of the plots. The segmented point clouds, exemplified in Fig. 3(b), served as the input data for the AGB estimation models described in the *AGB Estimation Models* section.

As shown in Fig. 3(b), the segmentation process effectively isolated individual trees, capturing detailed structural information from the ground to the canopy as

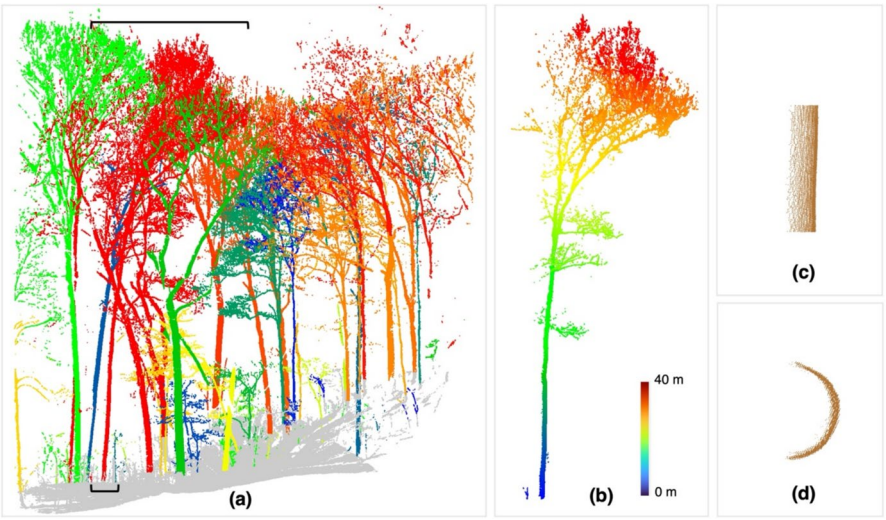

Fig. 3 Examples of segmented trees: **(a)** All segmented trees within a single tree inventory plot, where different colors represent individual segmented trees, and gray represents the ground points; **(b)** A single segmented tree from the plot. Its location is indicated by the black brackets in (a). Point colors represent height, ranging from the ground (blue) to the canopy top (red); **(c)** An enlarged view of the tree trunk from the segmented tree in (b); and **(d)** A cross-sectional view of the tree trunk

collected during the leaf-off season. However, several artifacts inherent to single-scan TLS remain. Point density decreases with increasing distance from the scanner. To address this limitation, an AGB estimation approach that is less sensitive to irregular point density (or a preprocessing step to homogenize point density across varying distances) is required. Furthermore, as illustrated in Figs. 3(c) and (d), the single-scan setup captures incomplete trunk representations, particularly occlusions on the side facing away from the scanner, hindering the application of conventional AGB estimation methods based on geometric reconstruction, such as cylinder fitting (Liu et al. 2018) and quantitative structure models (Kaasalainen et al. 2014; Fan et al. 2020). Accordingly, this study proposes a novel deep learning-based approach designed to overcome these inherent limitations of single-scan TLS data and enable accurate estimation of individual-tree AGB.

Labeling and Sampling

The final step of data preparation involved assigning the reference AGB values to the corresponding segmented tree point clouds, which were individually collected and preprocessed. The linked dataset was then divided into three subsets (training, validation, and testing) for the AGB estimation models introduced in the *AGB Estimation Models* section. Individual tree locations in the tree inventory were not straightforwardly matched to the corresponding segmented tree point clouds due to unavoidable human errors in measuring distances and azimuth angles. To address this challenge, we employed a modified version of a marriage algorithm developed by Gale and Shapely (1962) to match tree point clouds and their reference AGB values based on geographic proximity and similarity in tree size. Given the structural complexity of natural forests, an additional filtering step was applied after matching to ensure accurate labeling. As a result, 616 segmented tree point clouds were retained with assigned reference AGB values.

Because this study focused on individual trees in a natural forest, where tree sizes and species vary widely, the calculated reference AGB values exhibited substantial variability, as illustrated in Fig. 2(c). These values were distributed exponentially and unevenly, with approximately half below 1,000 kg and the remainder ranging from 1,000 to 10,000 kg. This highly variable distribution posed challenges such as potential overfitting within specific AGB ranges when randomly splitting the data into the three subsets. To mitigate this issue, we adopted a stratified sampling strategy based on the skewed distribution of the dependent variable. Specifically, the prepared data were grouped into five bins based on the AGB values, and samples within each bin were randomly assigned to training, validation, and testing subsets in an 8:1:1 ratio. This procedure resulted in 496 samples for training, 60 for validation, and 60 for testing. The training set was used to fit three AGB estimation models introduced in the following section, while the validation set was used to determine the final models. The testing dataset was used for a fair comparison of model performance. By employing stratified sampling, the training and testing datasets better presented the full range of AGB variability within the study site, ensuring more balanced and robust model development.

The exponential increase observed in the reference AGB values could pose an additional challenge, as deep learning models for regression typically assume a linearly distributed dependent variable. To address this issue, a logarithmic transformation was first applied to the AGB values to approximate linearity. The log-transformed values were then normalized using their mean and standard deviation. Furthermore, a small constant (0.005) was added to account for the possibility of AGB values smaller than the minimum observed in the reference data. This study's AGB estimation models were trained to predict normalized AGB values, and the predictions were subsequently converted back to the original scale by reversing the applied transformations.

AGB Estimation Models

Proposed Deep Learning Approach

Deep learning networks specifically designed for point cloud data are expected to fully exploit 3D information and thereby achieve meaningful learning outcomes, as demonstrated by Oehmcke et al. (2024). However, simultaneously, there remains a common consensus that such networks are less mature compared to those developed for raster data, owing to their relatively recent emergence (Guo et al. 2020). Consequently, the effectiveness of point cloud-based deep learning networks may vary depending on the characteristics of the input data. Although this study collected a reasonable number of individual tree point clouds using single-scan TLS data, the overall dataset remains limited compared to the large training datasets typically required for deep learning, particularly given the study's focus on a localized natural forest. In this context, we employed CoAtNet (Dai et al. 2021), a deep learning network originally developed for 2D raster data, which has demonstrated superior performance across various applications, including scenarios with limited training data.

(1) Converting three-dimensional point clouds to two-dimensional projected images

Because CoAtNet is designed for 2D raster data, the segmented tree point clouds needed to be transformed into 2D images (rasterized) to serve as inputs for the model. Various projection strategies have been adopted in previous studies involving point cloud data and deep learning in forestry, depending on the properties of the input data and the research objectives (Seidel et al. 2021; Fan et al. 2023; Schäfer et al. 2024). Similarly, we developed a tailored projection strategy to accurately estimate tree AGB from single-scan TLS data.

Figure 4 illustrates the proposed projection method, wherein a segmented tree point cloud is transformed into a three-band image, with each pixel representing the mean height of the corresponding lidar points. The 2D projection was generated based on vertical cross-sections (parallel to the z-axis) rather than horizontal ones for several reasons. First, as trees are generally characterized by greater height than width, horizontal cross-sections tend to lose more detailed structural information

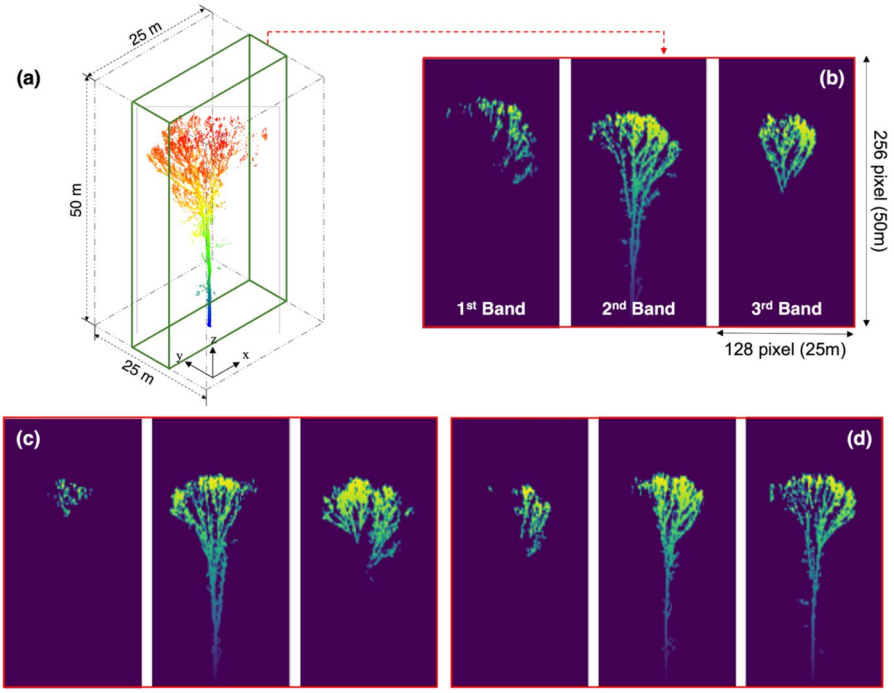

Fig. 4 Projection of a three-dimensional tree point cloud into three-band, two-dimensional images. (**a**) Front view of a tree point cloud, with point colors denoting height from the ground (blue) to the canopy top (red). The green box indicates the cross-sectional region corresponding to the second band of the projected image shown in (b), connected by the red dashed line. (**b**) Three-band projected image, where each pixel represents the mean height of the corresponding points in (a). (**c**) Projected image generated from (a) after a 30° rotation around the z-axis. (**d**) Projected image generated from (a) after a 120° rotation around the z-axis

than vertical cross-sections. Second, because the reference AGB values were derived from species-specific allometric equations (Eq. (1), Table 1), the proposed model needed to infer tree species to achieve more accurate estimates. Previous studies on species classification have demonstrated the effectiveness of vertical cross-sectional images in capturing species-related structural traits (Fan et al. 2023). Accordingly, vertical cross-sectional images were used in this study to enable the model to extract and utilize richer, higher-dimensional information relevant to both tree species and biomass.

The input image size for CoAtNet was set to 128 pixels in width and 256 pixels in height, reflecting two key considerations: the typical tree morphology (i.e., greater height than width) and the common deep learning practice of selecting dimensions that are powers of two. Given the strong correlation between tree size and AGB, a fixed scale factor was applied to all tree point clouds, ensuring consistency in tree area representation across images. Since the tallest trees within the study area reach just over 45 m, the 256-pixel image height was designed to represent 50 m in the point cloud space, resulting in an image spatial resolution

of approximately 0.19 m per pixel. To prevent axis-based distortion, the same resolution was maintained along the *x*-axis. Consequently, a fixed space (50 m in height and 25 m in width as well as depth, as tree canopy is likely circular rather than oval) centered on each individual tree point cloud was projected into a three-band, 128-by-256 image. The fixed space was first divided into three equal sections along the *y*-axis, as depicted in Fig. 4(a). For each section, a corresponding band image was created. The 2D image coordinates of each point were calculated by dividing the *x*- and *z*-coordinates by the spatial resolution. Then, the pixel value was set as the mean of the height (*z*-coordinate) of all the points falling within that pixel. Pixels without any corresponding points were assigned a value of zero. Figures 4(b)-(d) exemplify the resulting projected images from a point cloud.

(2) Data augmentation on 3D point cloud

As focusing on detailed, local-scale forests, the amount of training data available in this study is relatively small compared to previous large-scale AGB estimation efforts using deep learning (Oehmcke et al. 2024). To address this data scarcity, we employed data augmentation, a common approach for improving deep learning model robustness under limited training conditions. Rather than applying augmentation directly to the 2D images, which is a typical approach, we increased the number of input projected images by transforming the original point clouds prior to projection. Specifically, the point clouds were rotated around the *z*-axis by different angles, effectively altering their *x*- and *y*-coordinates and generating varied projected images from a single point cloud, as illustrated in Figs. 4(c) and (d). Rotations were applied in 15-degree increments from 0 to 180 degrees, resulting in 12 augmented versions of each point cloud. As a result, the final training and validation datasets consisted of 5,952 and 720 samples, respectively, which were subsequently used for CoAtNet model development in this study.

(3) CoAtNet for AGB estimation

Since the development of AlexNet (Krizhevsky et al. 2017), one of the most fundamental and representative convolutional neural networks (ConvNets), various deep learning architectures have been proposed across numerous fields. More recently, Vision Transformers (ViTs), which utilize self-attention mechanisms, have shown remarkable performance and emerged as competitive alternatives to conventional ConvNets (Bazi et al. 2021; Wang et al. 2022). CoAtNets (Dai et al. 2021) hybrids convolutional kernels (from ConvNets) and attention mechanisms (from ViTs) in a unified framework that balances generalization and model capacity, allowing them to retain strong performance while reducing the need for extensive training data. The CoAtNet architecture combines depthwise convolutional blocks—specifically, mobile inverted bottleneck convolutions (MBConv; Sandler et al. 2018)—with Transformer layers based on self-attention, stacked in a vertical sequence. Various ablation studies and benchmarking experiments have

shown that this hybrid architecture achieves higher efficiency and accuracy than state-of-the-art ConvNet- or Transformer-based models.

In this study, we adopt the CoAtNet architecture to estimate individual tree AGB from single-scan TLS data using a nonlinear regression approach, as AGB is a continuous variable. Among the CoAtNet family, we selected the CoAtNet-1 model to address potential overfitting and regularization challenges. Compared to large-scale open datasets commonly used in computer vision, our dataset is inevitably limited in size, making model training more prone to overfitting. Therefore, the depth and head dimensions of the model were adjusted to enable efficient learning from small datasets. The overall architecture of the CoAtNet used in this study is illustrated in Fig. 5. The Python code for the proposed CoAtNet implementation is available via a GitHub repository (https://github.com/gdslab/agbm_tls_dl).

(4) Ensemble for final tree AGB estimates

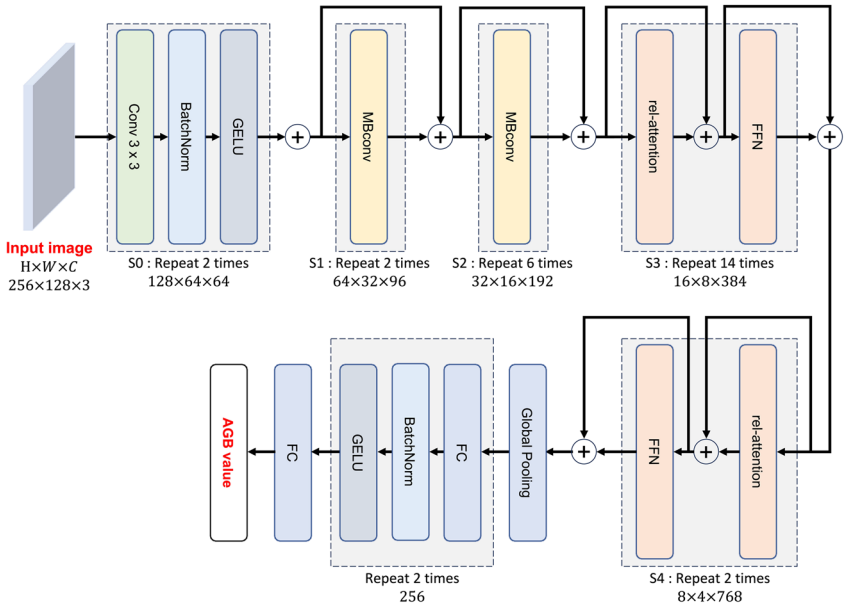

Fig. 5 Architecture of CoAtNet-1 for individual-tree AGB estimation. The input feature map (256 × 128 × 3; height (H) × width (W) × channels (C), as shown in Figs. 4(b)–(d)) passes through the stem stage (S0), two MBConv stages (S1 and S2), and two transformer stages with re-attention and FFN blocks (S3 and S4). The resulting 8 × 4 × 768 feature map undergoes global pooling and the head module (FC layer, BatchNorm, and GELU), followed by a final FC layer to predict AGB. Dropout (0.5) is applied to all convolutional, transformer, and FC layers. Stage repetition counts are indicated below each block. *Conv*, *BatchNorm*, *GELU*, *MBConv*, *re-attention*, *FFN and FC* denote convolution, batch normalization, Gaussian error linear unit, mobile inverted bottleneck convolution, relative attention, feed-forward network, and fully connected, respectively

In contrast to conventional data augmentation methods for 2D images based on geometric transformation (e.g., shifts, rotations, shears), this study applied geometric transformation directly to the original point cloud, generating twelve distinct and realistic projected images per tree, as shown in Figs. 4(b)-(d). While these augmented images served primarily to enhance model robustness during training, they also provided multiple observations for each tree during testing or practical application. For each test point cloud, twelve projected images were created through rotational augmentation along the *z*-axis, consistent with the data augmentation of the training datasets. The trained CoAtNet model produced an individual AGB estimate for each projected image, resulting in twelve predictions per tree. Recognizing the limitations of single-scan TLS data, particularly its incomplete coverage, an ensemble strategy was implemented to improve estimation accuracy. The final AGB value for each tree was calculated by averaging the twelve predictions after excluding outliers (i.e., any values that deviated by more than three standard deviations from the mean). This ensemble approach is expected to enhance overall robustness and effectively leverage the variability introduced by geometric augmentation to mitigate the uncertainties inherent in single-scan TLS data.

Other Comparison Models

(1) Random forest

Random forest has been widely adopted in recent AGB estimation studies due to its ease of implementation and demonstrated effectiveness. Previous studies developing deep learning-based AGB estimation models have often employed Random Forest as a baseline model. In a similar manner, this study established a Random Forest model using lidar features extracted from segmented point clouds in the training dataset, implemented through Scikit-Learn. Most lidar features commonly utilized in previous AGB estimation studies (Oehmcke et al. 2024; Schäfer et al. 2024) were calculated, including *the mean, the median, and the maximum of heights, the standard deviation, the entropy, the kurtosis, and the skewness of the height distribution, the percentage of returns with a height higher than canopy base height, the percentage of returns above the mean height, the 5th to 95th height percentiles in increments of 5.* Due to the absence of return number information in the single-scan TLS data used in this study, features based on return number were excluded. Additionally, features directly linked to point density, such as the total number of points, were not considered, given the inherent variability in point density relative to the distance from the scanner. Hyperparameters for the Random Forest model were optimized based on the best-performing configuration on the validation dataset.

(2) Point transformer

Although this study sought to preserve 3D information of point clouds by converting them into multi-band 2D images, the proposed method may not fully exploit

the valuable structural details compared to the deep learning networks explicitly designed for 3D point cloud processing (Wang et al. 2019; Zhao et al. 2021). Consequently, the performance of the proposed 2D projection-based approach may not be comparable to that of the 3D deep learning models. To provide a quantitative comparison, this study evaluated the proposed method alongside a representative 3D deep learning model, the Point Transformer (Zhao et al. 2021). Point Transformer is one of the representative self-attention networks, like CoAtNet employed in the proposed model. For this comparison, the individual point clouds were subsampled to 2,048 points, following common practices in prior Point Transformer implementations. Data augmentation techniques, including random rotations, point shuffling, and random point dropout, were applied to enhance the model's robustness, consistent with previous studies employing Point Transformer.

Experimental Results

Training Deep Learning Models

The CoAtNet-1 model was implemented with hardware (e.g., NVIDIA Quadro RTX A5000) and hyperparameters described in Table 2. For training the model, the AdamW optimizer, which is known for its excellent generalization performance, was applied (Loshchilov and Hutter 2017). To further improve the model's performance and stability, the learning rate was managed using the CyclicLR method (Smith 2017), employing a specific function where the learning rate cyclically increases and decreases, aiding in efficiently finding the optimal solution during training. The root mean squared error (RMSE) was used as a loss function. The model was trained for

Table 2 Hardware and hyperparameters for implementing CoAtNet and Point Transformer. The learning rate (*lr*) controls the magnitude of weight updates during training, and *base_lr* and *max_lr* define the lower and upper bounds of the learning rate used during training. The learning rate follows an exponential increase–decrease cycle during training (*exp_range*)

Hardware	CPU	Intel Xeon W-2235(3.8 GHz)	
	GPU	NVIDIA Quadro RTX A5000	
	RAM	128 GB	
	OS	Linux	
	Framework	Pytorch	
Hyperparameter	Batch size	12	
	Optimizer	AdamW	
	Learning rate (lr)	base lr	0.0001
		max lr	0.005
		step size up	10
		mode	exp_range
		gamma	0.995
	Epoch	200	
	Loss function	Root mean squared error (RMSE)	

200 epochs, with each epoch taking approximately 50 s. To prevent overfitting, the deep learning model used for AGB estimation was selected based on the epoch with the lowest loss function value on the validation data. The Point Transformer model was also trained using the same hardware and hyperparameters as the CoAtNet.

Evaluation Metrics

The proposed models and the other two comparison models were evaluated using the test dataset, which was not involved in model development. The test dataset's reference tree AGB values range approximately from 60 kg to 6,000 kg, with a mean of 1,305 kg, a median of 700 kg, and a standard deviation of 1,423 kg. The substantial difference between the mean and the median, along with the standard deviation being higher than the mean, complicates the evaluation of the estimation results. Therefore, the performance of the three models was assessed through a comprehensive analysis of the following multiple evaluation metrics commonly employed in AGB estimation: coefficient of determination (R^2),

$$R^2 = 1 - \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \tag{2}$$

mean bias (MB),

$$\mathrm{MB} = \frac{1}{N} \sum_i \left(y_i - \hat{y}_i\right) \tag{3}$$

RMSE,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_i \left(y_i - \hat{y}_i\right)^2} \tag{4}$$

mean percentage error (MPE),

$$\mathrm{MPE} = \frac{100}{N} \sum_i \frac{y_i - \hat{y}_i}{y_i} \tag{5}$$

and mean absolute percentage error (MAPE),

$$\mathrm{MAPE} = \frac{100}{N} \sum_i \frac{\left|y_i - \hat{y}_i\right|}{y_i} \tag{6}$$

where y_i and $\hat{y}_i$ indicate the reference and predicted AGB of the ith tree, respectively. $\bar{y}$ means the mean value of the reference AGBs. N denotes the total number of trees (i.e., 60) used for the evaluation in this study. In addition to the commonly used metrics described above, we computed median values of biases (MedB), percentage errors (MedPE), and absolute percentage errors (MedAPE), to account for the skewed distribution of AGB values and the potential presence of outliers resulting from the inherent limitations of single-scan TLS data. While a higher R^2 value

reflects better estimation performance, higher values of the error-related metrics, such as RMSE, MAPE, and MedAPE, indicate poorer performance. MB, MedB, MPE, and MedPE were further used to evaluate systematic tendencies in the predictions, with negative values indicating overestimation and positive values indicating underestimation.

Tree AGB Estimation Results

Table 3 summarizes each model's performance based on the evaluation metrics. Overall, all three models demonstrated reliable predictive capabilities, with R^2 values exceeding 0.6, indicating the potential of single-scan TLS data to enable meaningful estimation of tree AGB despite limitations in data quality. Predictive performances are further illustrated in Fig. 6, which presents scatter plots comparing the reference and predicted AGB values for each model. The plots show close alignment between predictions and observations, with no conspicuous outliers, supporting the effectiveness of each modeling approach.

The Random Forest model presented the poorest performance, with a lower R^2 (0.61) and higher RMSE, MAPE, and MedAPE (884.43 kg, 41.88%, and 32.04%, respectively) compared to the other two deep learning models (Table 3). This result expresses the inherent limitations of Random Forest in estimating tree AGB based on single-scan TLS data for the study site. While deep learning models can capture high-dimensional information beyond what conventional methods capture, Random Forest relies solely on manually extracted lidar features, limiting its ability to fully exploit the rich information embedded in dense point clouds. In addition, the absence of data augmentation during Random Forest training constrained the model's robustness. Despite these limitations, Fig. 6 shows that Random Forest

Table 3 Comparison of methods with the evaluation metrics on the test dataset. R^2 and RMSE denote a coefficient of determination and a root mean squared error, respectively

Metric		Proposed	Random Forest	Point Transformer
R^2		**0.73**	0.61	0.68
RMSE (kg)		**731.02**	884.43	797.96
MB (kg)		277.55	211.85	**141.53**
MedB (kg)		3.55	**1.57**	−6.35
MPE (%)		**−1.01**	−11.60	−1.62
MedPE (%)		0.99	1.05	**−0.77**
MAPE (%)	Overall	**29.40**	41.88	32.44
	Group ≤ 1000 kg	30.54	45.06	**29.58**
	Group > 1000 kg	**28.00**	37.98	35.93
MedAPE (%)	Overall	**25.06**	32.04	27.05
	Group ≤ 1000 kg	**19.98**	24.75	25.01
	Group > 1000 kg	**29.68**	32.49	39.32

MB, MedB, MPE, and MedPE stand for mean and median bias and mean and median percentage error. MAPE and MedAPE are mean and median absolute percentage errors. The best results are highlighted in bold font

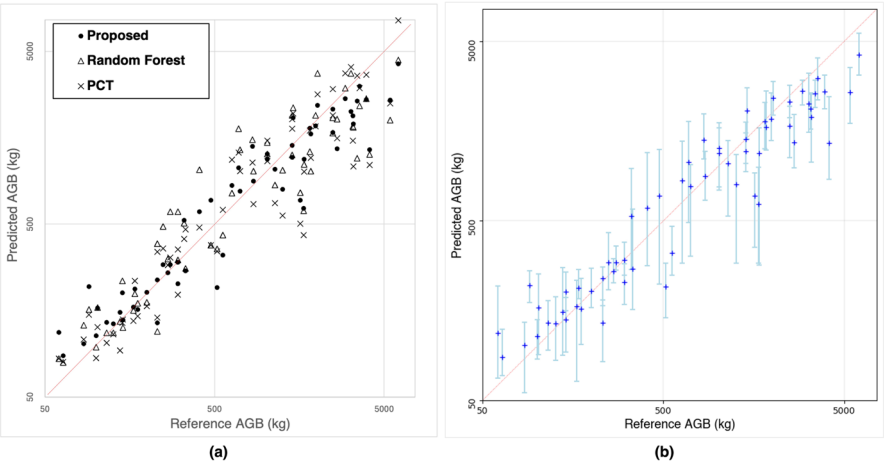

Fig. 6 Scatter plot comparing reference and predicted AGB values. (**a**) All three models, and (**b**) only the proposed model. The red dashed line represents the 1:1 line, indicating perfect agreement between the two values. The right blue vertical lines indicate the range of twelve predictions for each tree, and the crosses indicate the final predicted AGB values (i.e., the mean values). Both axes are on a logarithmic scale to account for the wide range of AGB values

predictions maintained a reasonable agreement with reference AGB values, indicating some potential for Random Forest. Random Forest could become a viable option for tree AGB estimation with larger datasets and the new advanced lidar features that better represent the deliquescent structure of hardwood trees. However, given that hardwood trees exhibit considerable variation across sites and species (Shao et al. 2018), achieving such improvements would require substantial efforts.

In contrast, the proposed CoAtNet-based model exhibited the best predictive performance, achieving an R^2 of 0.73 and the lowest RMSE of 731.02 kg. Although the RMSE may appear relatively high (approximately half the mean reference tree AGB of 1,305 kg), this result reflects the broad and sparse distribution of the dependent variables rather than poor model performance. Furthermore, the model's MAPE of 29.40% and MedAPE of 25.06% confirmed its strong capacity to accurately estimate individual tree AGB using single-scan TLS data. To examine any potential bias in estimation, the histogram of percentage errors for the test data is shown in Fig. 7(a). As exhibited by its distribution centered around 0%, the proposed model did not show a clear bias. This is also evidenced by the values close to 0 (an MPE of −1.01% and a MedPE of 0.99%) summarized in Table 3.

Similar to the proposed model, the Point Transformer model did not exhibit a clear bias, but showed a wider spread of percentage errors (Fig. 7(b)), suggesting relatively lower overall accuracy. This is consistent with its lower R^2 (0.68) and higher RMSE (797.96 kg) compared to the proposed model. A closer examination of the MAPE values by tree size showed that the Point Transformer performed slightly better for small trees (less than or equal to 1,000 kg) than the proposed CoAtNet model, with MAPE values of 29.58% and 30.54%, respectively. This was primarily due to the proposed model's exceptionally large errors for small trees, as indicated

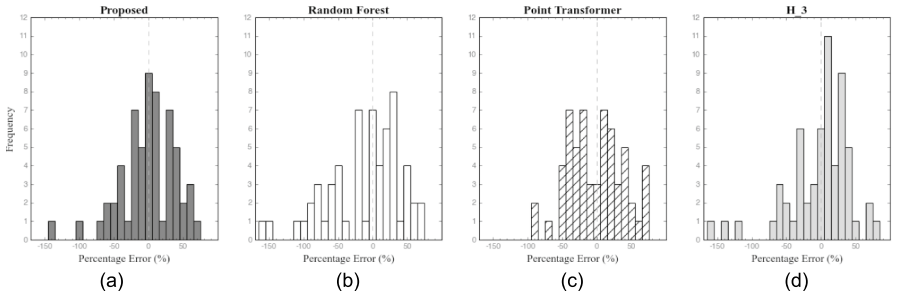

Fig. 7 Histograms of percentage errors (%) for three AGB estimation models: **(a)** Proposed CoAtNet-based model; **(b)** Random Forest model; **(c)** Point Transformer model; **(d)** Proposed model without the ensemble step. The H_3 model is explained in the Discussion. The dashed line indicates a percentage error of 0%

by percentage errors around and below −100% (Fig. 7(a)). However, in other metrics, including MAPE for large trees with AGB greater than 1,000 kg, the proposed model consistently exhibited superior robustness and overall predictive accuracy compared to the Point Transformer model.

From a computational cost aspect, the proposed method was also more efficient than the Point Transformer model. Computational costs were assessed using model parameters and FLOPs (floating point operations per second), representing the number of operations required for a single inference. The proposed CoAtNet model required 4.484 Giga-FLOPs with 32.8 million parameters, whereas the Point Transformer model required 9.418 Giga-FLOPs with 3.166 million parameters. Despite having more parameters, the proposed model demanded less computational time for both training and inference, owing to its lower FLOPs compared to the Point Transformer model.

Meanwhile, it is notable that the proposed model failed to accurately estimate AGBs for certain trees, as shown by two percentage errors around and below −100% (Fig. 7(a)). The two large errors caused the noticeable difference of more than 10% between the MAPE and the MedAPE values (30.54% and 19.98%, respectively) for the smaller trees (AGB ≤ 1000 kg), as well as the differences between the MAPE and the MedAPE values (29.40% and 25.06%, respectively) for all test trees. The exceptionally large errors were caused by segmentation errors, particularly for small trees. The high rate of occlusion in single-scan data occasionally caused the UCRP algorithm (Carpenter et al. 2022), adopted in this study, to missegment individual trees. Portions of trees, detached by occlusion from their main structure, might be mistakenly included in the segmentation results of neighboring trees. For smaller trees, such missegmentation could introduce anomalously high pixel values in the projected mean height images, thereby leading to substantial AGB overestimation. Although emerging deep learning-based tree segmentation methods (Henrich et al. 2024; Wielgosz et al. 2024) could potentially mitigate this issue, such approaches would need to be specifically adapted to address the unique characteristics of single-scan TLS point clouds collected in natural hardwood forests. Developing such adapted approaches remains an important future study. Nevertheless, considering

that such noticeable issues occurred in only two out of sixty test samples (and specifically in small trees that contribute relatively little to total forest AGB), the experimental results demonstrate both the effectiveness and reliable predictive performance of the proposed CoAtNet-based model, as well as the potential of single-scan TLS data for individual tree AGB estimation.

Discussion

This study proposed an effective framework for estimating individual tree AGB from single-scan TLS data by introducing (1) a suitable 3D-to-2D projection strategy, (2) a novel data augmentation method, and (3) an ensemble approach associated with the CoAtNet architecture. The effectiveness of these three components was evaluated through a series of comparative experiments. When projecting a 3D point cloud into a 2D image, various lidar features (e.g., 3D coordinates, intensity, return numbers) can be encoded as pixel values, and diverse projected encoding based on these features with different numbers of bands can be generated, resulting in different learning performances. Therefore, this study first compared the performances of the different input image encodings to determine the most effective one for the proposed CoAtNet. We specifically examined the effects in two aspects: the number of bands and the lidar attributes (features) of projected images. Table 4 compares the CoAtNet models with various projected images based on the evaluation metrics of the test dataset.

Table 4 Comparison of CoAtNet models (with different projection strategies, with typical augmentation, without ensemble) on the test dataset

Model	R^2	RMSE (kg)	MAPE (%)	MedAPE (%)
H_1	0.67	806.84	39.60	33.24
H_3	**0.74**	**722.38**	33.62	26.59
H_5	0.57	920.24	38.27	32.26
H_7	0.56	937.56	81.55	37.00
B_1	0.59	902.58	<u>32.11</u>	31.69
B_3	0.65	829.26	33.31	<u>26.27</u>
B_5	−1.23	2,105.19	38.94	35.21
B_7	0.49	1,004.85	82.97	36.99
D_1	0.58	917.48	35.92	34.97
H_3 + D_1	0.67	806.41	38.14	28.13
Proposed	<u>0.73</u>	<u>731.02</u>	**29.40**	**25.06**
Conventional	−2.60	2,678.29	41.64	33.19

The best results are in bold font, and the second-best results are underlined. 'H_n', 'B_n', and 'D_n' indicate the n-band projected images, whose pixel values were decided based on mean height, point existence (binary), and depth range, respectively. 'Conventional' indicates the CoAtNet model based on the three-band mean height images used for the proposed model, but with a traditional, typical augmentation strategy on images, not point clouds

Increasing the number of bands in the projected images is expected to better preserve 3D structural information by more finely capturing variations of the middle sections of trees. Consequently, a greater number of bands may improve the AGB estimation model's performance. To examine this effect, projected images were generated with one, three, five, and seven bands, based on the mean height. In Table 4, 'H_*n*' refers to CoAtNet models trained using *n*-band mean height images. The three-band model, H_3, achieved the best performance, with the highest R^2 value (0.74) and lowest RMSE, MAPE, and MedAPE (722.38 kg, 33.62%, and 26.59%, respectively). The model's performance reaches its best at three bands and decreases thereafter. A similar trend was observed for the models based on binary images, indicating point existence ('B_*n*' in Table 4). The model with three-band binary images performed the best, exhibiting the highest R^2 of 0.65 and the lowest RMSE of 829.26 kg, and performance metrics worsened with additional bands. These results imply that three-band images are an optimal trade-off between information retention and model complexity for AGB estimation in this study area. However, because the optimal number of vertical divisions may vary with tree size distribution and the amount of data available for training, future applications should adjust the number of bands to reflect local forest conditions.

The lidar attributes (or features) also affect the performance of deep learning, as Fan et al. (2023) demonstrated that an additional lidar feature, depth, could improve the accuracy of tree species classification. To evaluate the impact of feature selection on AGB estimation, this study compared two alternative lidar features: *point existence* and *depth range*, with the mean height used in the proposed model. The projected image inherently includes height information through image coordinates, where higher pixel rows correspond to greater heights in the 3D space. This suggests the use of simpler binary images based on point existence because the use of mean heights may be redundant. To generate a binary image, each pixel was assigned a value of one if it contained any lidar points, and zero otherwise. The depth range, another tested feature (denoted 'D_1' in Table 4), was defined as the difference between the maximum and minimum y-coordinates of the points within each pixel. This feature captured the thickness of tree structures, potentially allowing the model to infer tree volume more effectively and thereby improve AGB estimates. Intensity and return number were excluded due to their limitations in single-scan TLS: intensity is highly sensitive to factors such as incidence angle and range, and return number can be misleading due to occlusions from nearby trees within the same scan.

Contrary to expectations, neither the simple binary images nor the depth range images improve the model's performance, as B_3 showed the most promising results across all models with different projected images. It is notable that additional depth information did not improve but rather worsened the model performance, as the B_3 + D_1 model (using a three-band height image along with a one-band depth range image) exhibits a lower R^2 (0.67) and the higher RMSE (806.4 1 kg), MAPE (38.14%), and MedAPE values (28.13%) than those of the B_3 model. This may be attributed to structural occlusion inherent in single-scan TLS data, which often results in incomplete tree reconstructions on the side opposite the scanner (as illustrated in Figs. 3(c) and (d)). Consequently, the depth range feature may not reliably reflect true crown depth or volume in such conditions. Accordingly, this study

converted a point cloud into a three-band mean height image based on vertical cross-sections, exhibiting the most accurate and robust overall capability to estimate tree AGB among the projection approaches.

In addition to optimizing the 3D-to-2D projection strategy, this study introduced a novel data augmentation method that transformed the original 3D point cloud rather than the 2D projected images. As illustrated in Figs. 4(b)-(d), this approach successfully expanded the training dataset. The effectiveness of this proposed augmentation was evident when comparing the performance of the proposed model against a conventional model that used standard 2D image augmentations (Table 4). Both models employed the CoAtNet architecture (Fig. 5) and were trained with projected three-band mean height images, but used different data augmentation. The conventional model applied random shifts (±5 pixels), rotations (±5 degrees), and shearing (0–0.2) directly to the 2D images.

The conventional augmentation approach yielded the poorest performance among all models evaluated, with a negative R^2 value (−2.60) and an exceptionally high RMSE (2,678.29 kg), indicating model collapse. The failure of conventional augmentation implies that geometric distortions introduced at the image level can disrupt the spatial relationships and structural features essential for accurate AGB estimation. In contrast, the proposed augmentation method, which operates in 3D space prior to projection, preserved the structural integrity of tree forms and substantially improved model generalizability and robustness, as shown by better metric values. Consequently, from a practical perspective, we suggest that augmentation strategies in forestry applications—especially those involving single-scan TLS data—should operate at the point cloud level to maximize the utility of available training data, where acquiring large, annotated datasets is logistically challenging.

Lastly, the benefits of the proposed ensemble strategy were also clearly demonstrated. In the H_3 model, the AGB was estimated based solely on a single projected image from the original point cloud (without rotation along the z-axis). In the proposed approach, however, AGB estimates were obtained from multiple augmented images and then aggregated. Figure 6(b) shows the range of twelve predictions for each test tree, illustrating how AGB predictions vary depending on the projection angle. The distribution of these predictions is generally centered around the reference AGB, suggesting that the model captures the underlying pattern despite input variation. This uncertainty can be mitigated through aggregation, allowing for more stable and reliable AGB predictions. Although the proposed model has a slightly lower R^2 and a higher RMSE than the H_3 model (by 0.01 and 8.64 kg, respectively), its lower MAPE and MedAPE values (by 4.22% and 1.53%, respectively) highlight the effectiveness of the ensemble strategy (Table 4). Moreover, the H_3 model shows a tendency toward underestimation, as indicated by its percentage error distribution being centered around 10% (Fig. 7(d)). In contrast, the proposed model produces unbiased AGB predictions. The H_3 model also generates more frequent and notably larger percentage errors compared to the proposed model, confirming that integrating multiple perspectives through data augmentation enhances robustness and reduces uncertainty. In summary, the combination of (1) an optimized three-band projection strategy, (2) 3D-based data augmentation, and (3) ensemble learning substantially enhanced

the CoAtNet model's ability to accurately estimate individual tree AGB from single-scan TLS data. These findings demonstrate that carefully tailored pre-processing and learning strategies can maximize the predictive potential of deep learning models, even under the inherent limitations of single-scan data acquisition.

Conclusion

This study developed and evaluated a novel and effective framework for estimating individual tree AGB from single-scan TLS data by integrating a deep learning network (CoAtNet) with three key innovations: (1) an optimized 3D-to-2D projection strategy that converts 3D point clouds into three-band vertical cross-sectional images based on mean height; (2) a 3D point cloud–based data augmentation technique that applies geometric transformations directly to point clouds rather than to the projected images; and (3) an ensemble approach that aggregates predictions from multiple augmented inputs to enhance robustness. By combining these three components, the proposed model achieved strong predictive performance ($R^2 = 0.73$, MPE = 0.99%, and MAPE = 25.06%), surpassing both Random Forest and Point Transformer baselines. Despite the inherent limitations of single-scan TLS, such as occlusions and uneven point density, the experimental results in a hardwood forest demonstrate that carefully designed preprocessing and learning strategies can effectively harness the structural information embedded in such data. While the framework performed well overall, its accuracy in estimating small trees was limited by segmentation errors. These challenges may be addressed in future work by modifying and incorporating emerging tree segmentation methods given the characteristics of single-scan TLS and hardwood forests. The developed method is expected to enable reliable and scalable AGB estimation at the individual-tree level without the need for labor-intensive multi-scan TLS. This framework provides a practical solution for forest researchers and practitioners seeking to monitor forest structure and biomass in diverse, data-scarce environments. It contributes to advancing remote sensing methodologies in support of sustainable forest management and carbon accounting, particularly in the context of climate-driven ecological change.

Acknowledgements This work is based upon efforts supported by the PERSEUS grant, #2023-68012-38992 under USDA NIFA. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NIFA or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein

Authors Contributions M. Jung – Conceptualization, Investigation, Data curation, Methodology, Software, Validation, Visualization, Formal analysis, Writing (original draft, review & editing); J. Choi – Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing (original draft, review & editing); J. Carpenter – Data curation, Software, Visualization, Writing (review & editing); S. Fei – Data curation, Writing (review & editing), Resources, Funding acquisition; J. Jung – Conceptualization, Methodology, Writing (review & editing), Resources, Funding acquisition, Supervision. All authors read and approved the final manuscript.

Data Availability The single-scan TLS data and the corresponding tree inventory data are publicly available through the Data-to-Science STAC (SpatioTemporal Asset Catalogs) server: https://stac.d2s.org/collections/martell-forest-benchmark-data

Declarations

Competing Interests None of the authors are aware of financial or non-financial competing interests.

References

Astrup, R., M. J. Ducey, A. Granhus, T. Ritter, and N. von Lüpke. 2014. Approaches for estimating stand-level volume using terrestrial laser scanning in a single-scan mode. *Canadian Journal of Forest Research* 44 (6): 666–676. https://doi.org/10.1139/cjfr-2013-0535.

Batchelor, J. L., T. M. Wilson, M. J. Olsen, and W. J. Ripple. 2023. New structural complexity metrics for forests from single terrestrial LiDAR scans. *Remote Sensing* 15:145. https://doi.org/10.3390/rs15010145.

Bazi, Y., L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan. 2021. Vision transformers for remote sensing image classification. *Remote Sensing* 13:516. https://doi.org/10.3390/rs13030516.

Brede, B., L. Terryn, N. Barbier, H. M. Bartholomeus, R. Bartolo, K. Calders, G. Derroire, S. M. K. Moorthy, A. Lau, S. R. Levick, P. Raumonen, H. Verbeeck, D. Wang, T. Whiteside, J. Van Der Zee, and M. Herold. 2022. Non-destructive estimation of individual tree biomass: Allometric models, terrestrial and UAV laser scanning. *Remote Sensing of Environment* 280 : 113180. https://doi.org/10.1016/j.rse.2022.113180.

Burt, A, M. I. Disney., P. Raumonen., J. Armston., K. Calders, and P. Lewis. 2013. Rapid characterisation of forest structure from TLS and 3D modelling. In 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS 3387–3390. https://doi.org/10.1109/IGARSS.2013.6723555

Calders, K., G. Newnham, A. Burt, S. Murphy, P. Raumonen, M. Herold, D. Culvenor, V. Avitabile, M. Disney, J. Armstone, and M. Kaasalainen. 2015. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution* 6:98–208. https://doi.org/10.1111/2041-210X.12301.

Cao, L., N. C. Coops, J. L. Innes, S. R. Sheppard, L. Fu, H. Ruan, and G. She. 2016. Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data. *Remote Sensing of Environment* 178:158–171. https://doi.org/10.1016/j.rse.2016.03.012.

Carpenter, J., J. Jung, S. Oh, B. Hardiman, and S. Fei. 2022. An unsupervised canopy-to-root pathing (UCRP) tree segmentation algorithm for automatic forest mapping. *Remote Sensing* 14:4274. https://doi.org/10.3390/rs14174274.

Chojnacky, D. C., L. S. Heath, and J. C. Jenkins. 2014. Updated generalized biomass equations for North American tree species. *Forestry* 87:129–151. https://doi.org/10.1093/forestry/cpt053.

Dai, Z., H. Liu., Q. Le, and M. Tan. 2021. CoAtNet: Marrying convolution and attention for all data sizes. In Proceedings of the 35th Conference on Neural Information Processing Systems.

Demol, M., P. Wilkes, P. Raumonen, S. M. Krishna Moorthy, K. Calders, B. Gielen, and H. Verbeeck. 2022. Volumetric overestimation of small branches in 3D reconstructions of *Fraxinus excelsior*. *Silva Fennica* 56 : 10550. https://doi.org/10.14214/sf.10550.

Dorado-Roda, I., A. Pascual, S. Godinho, C. A. Silva, B. Botequim, P. Rodríguez-Gonzálvez, E. González-Ferreoiro, and J. Guerra-Hernández. 2021. Assessing the accuracy of GEDI data for canopy height and aboveground biomass estimates in Mediterranean forests. *Remote Sensing* 13:2279. https://doi.org/10.3390/rs13122279.

Dubayah, R., J. Armston, S. P. Healey, J. M. Bruening, P. L. Patterson, J. R. Kellner, L. Duncanson, S. Saarela, G. Ståhl, Z. Yang, H. Tang, J. B. Blair, L. Fatoyinbo, S. Goetz, S. Hancock, M. Hansen, M. Hofton, G. Hurtt, and S. Luthcke. 2022. Gedi launches a new era of biomass inference from space. *Environmental Research Letters* 17 : 095001. https://doi.org/10.1088/1748-9326/ac8694.

Duncanson, L., J. Armston, M. Disney, V. Avitabile, N. Barbier, K. Calders, S. Carter, J. Chave, M. Herold, T. W. Crowther, M. Falkowski, J. R. Kellner, N. Labrière, R. Lucas, N. MacBean, R. E. Mc Roberts, V. Meyer, E. Næsset, J. E. Nickeson, K. I. Paul, O. L. Phillips, M. Réjou-Méchain, M. Román, S. Roxburgh, S. Saatchi, D. Schepaschenko, K. Scipal, P. R. Siqueira, A. Whitehurst, and M. Williams. 2019. The importance of consistent global forest aboveground biomass product validation. *Surveys in Geophysics* 40:979–999. https://doi.org/10.1007/s10712-019-09538-8.

Fan, G., L. Nan, Y. Dong, X. Su, and F. Chen. 2020. AdQSM: A new method for estimating aboveground biomass from TLS point clouds. *Remote Sensing* 12:3089. https://doi.org/10.3390/rs12183089.

Fan, Z., W. Zhang, R. Zhang, J. Wei, Z. Wang, and Y. Ruan. 2023. Classification of tree species based on point cloud projection images with depth information. *Forests* 14:2014. https://doi.org/10.3390/f14102014.

Gale, D., and L. S. Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69:9–15. https://doi.org/10.1080/00029890.1962.11989827.

Gleason, C. J., and J. Im. 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment* 125:80–91. https://doi.org/10.1016/j.rse.2012.07.006.

Guo, Y., H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun. 2020. Deep learning for 3D point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43:4338–4364. https://doi.org/10.1109/TPAMI.2020.3005434.

Henrich, J., J. van Delden, D. Seidel, T. Kneib, and A. S. Ecker. 2024. Treelearn: A deep learning method for segmenting individual trees from ground-based LiDAR forest point clouds. *Ecological Informatics* 84 : 102888. https://doi.org/10.1016/j.ecoinf.2024.102888.

Hoover, C. M., and J. E. Smith. 2016. Evaluating revised biomass equations: Are some forest types more equivalent than others? *Carbon Balance and Management* 11:1–20. https://doi.org/10.1186/s13021-015-0042-5.

Houghton, R. A. 2005. Aboveground forest biomass and the global carbon balance. *Global Change Biology* 11:945–958. https://doi.org/10.1111/j.1365-2486.2005.00955.x.

Kaasalainen, S., A. Krooks, J. Liski, P. Raumonen, H. Kaartinen, M. Kaasalainen, E. Puttonen, K. Anttila, and R. Mäkipää. 2014. Change detection of tree biomass with terrestrial laser scanning and quantitative structure modelling. *Remote Sensing* 6:3906–3922. https://doi.org/10.3390/rs6053906.

Kankare, V., M. Holopainen, M. Vastaranta, E. Puttonen, X. Yu, J. Hyyppä, M. Vaaja, H. Hyyppä, and P. Alho. 2013. Individual tree biomass estimation using terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing* 75:64–75. https://doi.org/10.1016/j.isprsjprs.2012.10.003.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60:84–90. https://doi.org/10.1145/3065386.

Kükenbrink, D., O. Gardi, F. Morsdorf, E. Thürig, A. Schellenberger, and L. Mathys. 2021. Aboveground biomass references for urban trees from terrestrial laser scanning data. *Annals of Botany* 128:709–724. https://doi.org/10.1093/aob/mcab002.

Lian, X., H. Zhang, W. Xiao, Y. Lei, L. Ge, K. Qin, Y. He, Q. Dong, L. Li, Y. Han, H. Fan, Y. Li, L. Shi, and J. Chang. 2022. Biomass calculations of individual trees based on unmanned aerial vehicle multispectral imagery and laser scanning combined with terrestrial laser scanning in complex stands. *Remote Sensing* 14:4715. https://doi.org/10.3390/rs14194715.

Liang, X., V. Kankare, J. Hyyppä, Y. Wang, A. Kukko, H. Haggrén, X. Yu, H. Kaartinen, A. Jaakkola, F. Guan, M. Holopainen, and M. Vastaranta. 2016. Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing* 115:63–77. https://doi.org/10.1016/j.isprsjprs.2016.01.006.

Liang, X., J. Hyyppä, H. Kaartinen, M. Lehtomäki, J. Pyörälä, N. Pfeifer, M. Holopainen, G. Brolly, F. Pirotti, J. Hackenberg, H. Huang, H.-W. Jo, M. Katoh, L. Liu, M. Mokroš, J. Morel, K. Olofsson, J. Poveda-Lopez, J. Trochta, D. Wang, J. Wang, Z. Xi, B. Yang, G. Zheng, V. Kankare, V. Luoma, X. Yu, L. Chen, M. Vastaranta, N. Saarinen, and Y. Wang. 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing* 144:137–179. https://doi.org/10.1016/j.isprsjprs.2018.06.021.

Lin, J., D. Chen, W. Wu, and X. Liao. 2022. Estimating aboveground biomass of urban forest trees with dual-source UAV acquired point clouds. *Urban Forestry & Urban Greening* 69 : 127521. https://doi.org/10.1016/j.ufug.2022.127521.

Liu, G., J. Wang, P. Dong, Y. Chen, and Z. Liu. 2018. Estimating individual tree height and diameter at breast height (DBH) from terrestrial laser scanning (TLS) data at plot level. *Forests* 9:398. https://doi.org/10.3390/f9070398.

Liu, B., S. Chen, H. Huang, and X. Tian. 2022. Tree species classification of backpack laser scanning data using the PointNet++ point cloud deep learning method. *Remote Sensing* 14:3809. https://doi.org/10.3390/rs14153809.

Loshchilov, I., and F. Hutter. 2017. Decoupled weight decay regularization, arXiv preprint. arXiv:1711.05101. https://doi.org/10.48550/arXiv.1711.05101

Lu, D., Q. Chen, G. Wang, L. Liu, G. Li, and E. Moran. 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth* 9:63–105. https://doi.org/10.1080/17538947.2014.990526.

Lu, J., H. Wang, S. Qin, L. Cao, R. Pu, G. Li, and J. Sun. 2020. Estimation of aboveground biomass of *Robinia pseudoacacia* forest in the Yellow River Delta based on UAV and Backpack LiDAR point clouds. *International Journal of Applied Earth Observation and Geoinformation* 86 : 102014. https://doi.org/10.1016/j.jag.2019.102014.

Miles, P. D., and W.B. Smith. 2009. Specific gravity and other properties of wood and bark for 156 tree species found in North America. Research Note NRS-38. Newtown Square, PA: USDA Forest Service, Northern Research Station. https://doi.org/10.2737/NRS-RN-38

Oehmcke, S., L. Li, K. Trepekli, J. C. Revenga, T. Nord-Larsen, F. Gieseke, and C. Igel. 2024. Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR. *Remote Sensing of Environment* 302 : 113968. https://doi.org/10.1016/j.rse.2023.113968.

Sandler, M., A. Howard., M. Zhu., A. Zhmoginov, and L. C. Chen. 2018 Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 4510–4520

Schäfer, J., L. Winiwarter, H. Weiser, B. Höfle, S. Schmidtlein, J. Jan Novotný, G. Krok, K. Stereńczak, M. Hollaus, and F. E. Fassnach. 2024. CNN-based transfer learning for forest aboveground biomass prediction from ALS point cloud tomography. *European Journal of Remote Sensing* 57:2396932. https://doi.org/10.1080/22797254.2024.2396932.

Seidel, D., M. Ehbrecht, and K. Puettmann. 2016. Assessing different components of three-dimensional forest structure with single-scan terrestrial laser scanning: A case study. *Forest Ecology and Management* 381:196–208. https://doi.org/10.1016/j.foreco.2016.09.036.

Seidel, D., P. Annighöfer, A. Thielman, Q. E. Seifert, J. H. Thauer, J. Glatthorn, M. Ehbrecht, T. Kneib, and C. Ammer. 2021. Predicting tree species from 3d laser scanning point clouds using deep learning. *Frontiers in Plant Science* 12 : 635440. https://doi.org/10.3389/fpls.2021.635440.

Shao, G., G. Shao, J. Gallion, M. R. Saunders, J. R. Frankenberger, and S. Fei. 2018. Improving lidar-based aboveground biomass estimation of temperate hardwood forests with varying site productivity. *Remote Sensing of Environment* 204:872–882. https://doi.org/10.1016/j.rse.2017.09.011.

Shendryk, Y. 2022. Fusing GEDI with earth observation data for large area aboveground biomass mapping. *International Journal of Applied Earth Observation and Geoinformation* 115 : 103108. https://doi.org/10.1016/j.jag.2022.103108.

Sileshi, G. W. 2014. A critical review of forest biomass estimation models, common mistakes and corrective measures. *Forest Ecology and Management* 329:237–254. https://doi.org/10.1016/j.foreco.2014.06.026.

Smith, L. N. 2017. Cyclical learning rates for training neural networks. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 464–472. https://doi.org/10.1109/WACV.2017.58

Srinivasan, S., S. C. Popescu, M. Eriksson, R. D. Sheridan, and N. W. Ku. 2014. Multi-temporal terrestrial laser scanning for modeling tree biomass change. *Forest Ecology and Management* 318:304–317. https://doi.org/10.1016/j.foreco.2014.01.038.

Stovall, A. E., A. G. Vorster, R. S. Anderson, P. H. Evangelista, and H. H. Shugar. 2017. Non-destructive aboveground biomass estimation of coniferous trees using terrestrial LiDAR. *Remote Sensing of Environment* 200:31–42. https://doi.org/10.1016/j.rse.2017.08.013.

Wang, Y., Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. 2019. Dynamic graph CNN for learning on point clouds. *Acm Transactions on Graphics* 38:1–12. https://doi.org/10.1145/3326362.

Wang, D., Q. Zhang, Y. Xu, J. Zhang, B. Du, D. Tao, and L. Zhang. 2022. Advancing plain vision transformer toward remote sensing foundation model. *IEEE Transactions on Geoscience and Remote Sensing* 61:1–15. https://doi.org/10.1109/TGRS.2022.3222818.

Wielgosz, M., S. Puliti, B. Xiang, K. Schindler, and R. Astrup. 2024. Segmentanytree: A sensor and platform agnostic deep learning model for tree segmentation using laser scanning data. *Remote Sensing of Environment* 313 : 114367. https://doi.org/10.1016/j.rse.2024.114367.

Wilkes, P., A. Lau, M. Disney, K. Calders, A. Burt, J. G. de Tanago, H. Bartholomeus, B. Brede, and M. Herold. 2017. Data acquisition considerations for terrestrial laser scanning of forest plots. *Remote Sensing of Environment* 196:140–153. https://doi.org/10.1016/j.rse.2017.04.030.

Zhao, H., L. Jiang., J. Jia., P. H. Torr., and V. Koltun. 2021 Point transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR) 16259–16268.

Zhou, L., X. Li, B. Zhang, J. Xuan, Y. Gong, C. Tan, H. Huang, and H. Du. 2022. Estimating 3d green volume and aboveground biomass of urban forest trees by UAV-lidar. *Remote Sensing* 14:5211. https://doi.org/10.3390/rs14205211.

Authors and Affiliations

Minyoung Jung[1] · Jaewan Choi[2] · Joshua Carpenter[3] · Songlin Fei[4] · Jinha Jung[1]

✉ Jinha Jung
jinha@purdue.edu

[1] Lyles School of Civil and Construction Engineering, Purdue University, West Lafayette, IN 47907, USA

[2] Department of Civil Engineering, Chungbuk National University, Cheongju, Chungbuk 28644, South Korea

[3] Civil Engineering, The University of Akron, Akron, OH 44325, USA

[4] Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA