

Comparative Analysis of YOLOv8 and YOLOv11 on Tree Detection Using UAV RGB and Laser Scanning Data

Esra Sengun¹, Samet Aksoy², Elif Sertel², Johan E. S. Fransson¹

¹ Linnaeus University, Department of Forestry and Wood Technology, Växjö, Sweden – esra.sengun@lnu.se, johan.fransson@lnu.se

² Istanbul Technical University, Department of Geomatics Engineering, 34469 Maslak Istanbul, Turkey – aksoysa@itu.edu.tr, sertele@itu.edu.tr

Keywords: Deep Learning (DL), Tree Detection, LiDAR, Artificial Intelligence (AI), Precision Forestry, YOLO.

Abstract

To promote sustainable forest management planning including biodiversity monitoring and to enable accurate estimates of stem volume, above-ground biomass, and carbon stocks, tree identification is essential to contemporary forest inventory. Deep learning models are now crucial tools for automating tree recognition over large, forested regions due to the growing availability of high-resolution LiDAR data. In order to identify individual trees using LiDAR-derived RGB raster imagery, this work compares two cutting-edge object identification architectures: YOLOv8 and YOLOv11. A total of 82 annotated images were utilized, rasterized at a resolution of 5 cm, and processed using two input resolutions (640×640 pixel and 960×960 pixel), several model configurations (s, m, l, x), and augmentation settings (rotation and horizontal flip). To provide fair comparison, every model was trained and evaluated using the same methodology. Precision, recall, mAP50, and mAP50-95, standard detection metrics, were used to evaluate performance. The results show that YOLOv8 consistently beat YOLOv11 on all metrics, especially in its large and extra-large forms at high resolution. YOLOv8x with 960 pixel resolution and augmentation was the best-performing setup, with 0.974 precision, 0.837 recall, 0.934 mAP50, and 0.821 mAP50-95. The results demonstrate notable improvements in detection accuracy when compared to previous methods that used YOLOv4 or domain-specific structures like YOLOTree. With the use of rasterized UAV laser scanning data, our results highlight the potential of the YOLO architecture as a robust and scalable tool for automated, high-precision forest inventory.

1. Introduction

Forests are among the most important ecosystems on Earth, offering vital functions such as climate regulation, carbon sequestration, soil preservation, water purification, and biodiversity conservation (Luo et al., 2024; Satama-Bermeo et al., 2025; Straker et al., 2023; Sun et al., 2022; Topgül et al., 2025). As global environmental challenges such as climate change, deforestation, and biodiversity loss worsen, the significance of sustainable forest management and accurate forest inventories grows (Satama-Bermeo et al., 2025; Sun et al., 2022). Forest inventories are the systematic collection of detailed data on forest features such as species composition, tree size, health condition, and spatial distribution (Luo et al., 2024; Straker et al., 2023). Such thorough inventories are essential for making informed decisions about resource allocation, conservation strategies, pest control, selective logging, and climate change mitigation (Luo et al., 2024; Topgül et al., 2025).

Individual tree identification is a critical component of modern forest inventories, allowing for precise estimates of stem volume, above-ground biomass, and carbon stocks, as well as targeted interventions for biodiversity conservation, forest health monitoring, and sustainable management practices (Satama-Bermeo et al., 2025; Straker et al., 2023; Sun et al., 2022). Traditional techniques of tree identification, such as human field surveys and aerial imaging interpretation, while accurate, are labor-intensive, time-consuming, and costly, especially when scaled across large or inaccessible forest regions (Luo et al., 2024; Satama-Bermeo et al., 2025; Topgül et al., 2025).

By making it possible to collect detailed structural and spectral data across large forested landscapes, recent developments in remote sensing technologies, such as satellite imagery, airborne laser scanning (ALS), and high-resolution unmanned aerial vehicle (UAV) imagery, have significantly increased the effectiveness and scope of forest inventories (Luo et al., 2024; Satama-Bermeo et al., 2025; Sun et al., 2022). Numerous techniques, such as marker-controlled algorithms, point cloud clustering, and watershed segmentation, may be used to extract information from these remote sensing datasets (Sun et al., 2022).

Convolutional neural networks (CNNs), a type of deep learning (DL) approach, have lately been quite effective at automating tasks involving the recognition of individual trees. When compared to conventional techniques, these models greatly improve the scalability, accuracy, and efficiency of tree identification procedures (Luo et al., 2024; Straker et al., 2023; Topgül et al., 2025). Notably, even in complex environments with overlapping canopies and diverse growth patterns, object detection models such as You Only Look Once (YOLO) architectures (YOLOv5, YOLOv7, YOLOv8, and YOLOv9) have demonstrated remarkable potential in accurately identifying individual tree crowns (Luo et al., 2024; Satama-Bermeo et al., 2025; Sertel and Topgul, 2025; Topgül et al., 2025).

In this paper, deep learning methods for individual tree detection are investigated and evaluated for coniferous forest in northern Sweden, focusing on the importance of advancements towards precision forestry, forest inventory, and sustainable management of global forest resources (Satama-Bermeo et al., 2025; Straker et al., 2023; Sun et al., 2022).

2. Materials and Methods

2.1 Study Area

The study was carried out at the Svartberget Experimental Forest, located northwest of Vindeln in northern Sweden (64.24°N, 19.77°E). The forest is owned by the forest company Sveaskog and is a part of Vindeln Experimental Forests run by the Swedish University of Agricultural Sciences. In the boreal zone, Svartberget is a long-term study site with an emphasis on sustainable forest management, catchment-scale hydrology, and mire ecology. The area is characterized by elevations ranging from 160 m to 320 m above sea level and consists predominantly of mixed coniferous forest with Norway spruce (*Picea abies* (L.) H. Karst.) and Scots pine (*Pinus sylvestris* L.). The underlying bedrock is composed almost entirely of gneiss, while the soils are primarily formed from moraines of varying thickness.

2.2 Dataset

In this study, the dataset consists of 82 raster images created from LiDAR point clouds captured via UAV laser scanning on DJI Matrice 350 RTK drone equipped with DJI Zenmuse L1 (i.e., a Livox LiDAR module and a 20 MP RGB camera). Firstly, raw data are processed through DJI Terra software and then the raw point clouds were processed in Cloud-Compare software, rendering RGB raster images at a 5 cm spatial resolution. This rasterizing method preserves the structural and spatial attributes of individual tree crowns while transforming the 3D point cloud into a 2D form that can be input into CNNs for object detection.

Due to the variability in forest conditions in terms of tree density, crown overlap, and canopy structure, it was essential to ensure the robustness of the model. Annotated bounding boxes were placed around every individual tree in each image for ground-truthing in supervised learning. The dataset was split into training (70%), validation (20%), and testing (10%) subsets, with a representative distribution of forest complexity maintained across all splits. In total, 4600 trees were annotated. Even though raster image count is limited, many trees were populated in these images. Therefore, there are a total of 3103 trees in the training dataset, 974 trees in the validation dataset, and 523 trees in the test dataset.

Data augmentation was applied to the training set as a means of model generalization. Each annotated image was augmented three times with horizontal flipping and random rotations, -15° to +15° on either side. The augmented variability further aided the models in detecting trees under varying orientations and spatial arrangements.

2.3 Methodology

This study compares the two object-detection architectures, YOLOv8 and YOLOv11, introduced by Jocher et al. (2023) for detecting individual trees using the raster images derived via UAV RGB and LiDAR data. The same training and evaluation framework was implemented for both models to ensure a consistent basis for comparison. The reason for selecting real-time object-detection models is their scalability of this study. Since forests covers large areas, the inference time of these models can add up significantly over time (Sertel and Topgul, 2025; Terven et al., 2023).

2.3.1 Model Architecture and Training Configuration:

The model families used in this study, YOLOv8 and YOLOv11 (Jocher et al., 2023), were evaluated under four configurations,

s-small, m-middle, l-large, and x-extra-large. These variants differ in terms of depth and width so the effect of the complexity of each model could be assessed in terms of detection performance (Terven et al., 2023). All models are pretrained with Common Objects in Context dataset before any further training.

The model was trained using the input resolutions of 640×640 pixel and 960×960 pixel to assess how spatial detail may influence model accuracy. In order to further improve generalization and to provide robustness against overfitting, data augmentation was employed. This involved augmenting each input sample three times, with horizontal flipping and random rotation in the range of -15° to +15°. These augmentation is selected for comparability of previous research by Topgöl et al. (2025).

It was deemed essential to apply all models by supervised learning, with the same optimization setting in terms of learning rate scheduling, batch size, and loss functions. This allows to factor out other variables affecting architecture and input resolution.

2.3.2 Performance Evaluation Metrics: The model has been evaluated using precision, recall, and mean Average Precision (mAP), all of which are widely regarded as standard metrics for evaluating object detection performance. Together with other complements, those metrics can be used to judge detection accuracy, robustness, and the ability of a given model to both localize and classify individual trees.

The metrics are defined as:

- **Precision** evaluates the ratio of correctly predicted trees among all predicted bounding boxes:

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

- **Recall** quantifies the proportion of true trees that were successfully detected:

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

- **Average Precision (AP)** is computed as the area under the precision–recall curve for a given IoU threshold:
- **Mean Average Precision (mAP)** is the average of AP values over different Intersection over Union (IoU) thresholds. In this study, two versions were used:
 - **mAP50**, calculated at a fixed IoU threshold of 0.5.
 - **mAP50–95**, averaged over multiple IoU thresholds ranging from 0.50 to 0.95 in 0.05 increments:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

All metrics were computed using the test set. This evaluation framework enables a comprehensive comparison of the YOLOv8 and YOLOv11 architectures in the context of tree detection from rasterized UAV LiDAR imagery.

3. Results

The performance results of the YOLOv8 models for individual tree detection are presented in Table 1. The table reports results for four commonly used object detection performance metrics: precision, recall, mAP50, and mAP50-95 evaluated across a range of model sizes (s, m, l, x), input resolutions (640×640 pixel and 960×960 pixel), and augmentation settings (with and without augmentation).

3.1 YOLOv8 Performance

YOLOv8 models showed a significant increase in recall and mAP50-95 values as the model size and resolution increase. Among YOLOv8 models, the best mAP50-95 value (0.821) was achieved at 960 pixel resolution without increasing the crown localization generalization ability, which was demonstrated by YOLOv8x. Similar excellent performances were achieved for YOLOv8l and YOLOv8x at 960 pixel resolution with augmentation, yielding mAP50-95 values of 0.819 and 0.814, respectively (Table 1).

In most cases, only slight increases in mAP50-95 values were observed with data augmentation; the largest increase occurred for YOLOv8m (640 pixel), with the value rising from 0.753 to 0.780. On the other hand, for YOLOv8x (960 pixel), the accuracy decreased from 0.974 to 0.959 when augmentation was applied. The benefits of augmentation include improved detection across all robustness levels, but there is also a possibility of geometric distortion, which may cause a slight loss in prediction reliability.

Table 1. Metric results of YOLOv8 experiments.

Dataset	Model	Precision	Recall	mAP 50	mAP 50-95
640	YOLOv8s	0.949	0.813	0.906	0.678
640	YOLOv11s	0.926	0.841	0.919	0.659
640aug	YOLOv8s	0.973	0.816	0.909	0.703
640aug	YOLOv11s	0.936	0.828	0.916	0.702
640	YOLOv8m	0.966	0.845	0.930	0.753
640	YOLOv11m	0.957	0.846	0.925	0.724
640aug	YOLOv8m	0.947	0.861	0.935	0.780
640aug	YOLOv11m	0.939	0.843	0.922	0.728
640	YOLOv8l	0.953	0.824	0.919	0.765
640	YOLOv11l	0.958	0.816	0.915	0.715
640aug	YOLOv8l	0.954	0.838	0.926	0.723
640aug	YOLOv11l	0.930	0.844	0.923	0.744
640	YOLOv8x	0.965	0.835	0.923	0.788
640	YOLOv11x	0.942	0.813	0.902	0.717
640aug	YOLOv8x	0.952	0.827	0.914	0.798
640aug	YOLOv11x	0.958	0.826	0.921	0.744
960	YOLOv8s	0.959	0.851	0.927	0.743
960	YOLOv11s	0.946	0.870	0.925	0.741
960aug	YOLOv8s	0.961	0.849	0.919	0.726
960aug	YOLOv11s	0.952	0.857	0.918	0.729
960	YOLOv8m	0.979	0.870	0.925	0.790
960	YOLOv11m	0.962	0.862	0.929	0.781
960aug	YOLOv8m	0.951	0.853	0.912	0.786
960aug	YOLOv11m	0.952	0.865	0.927	0.789
960	YOLOv8l	0.963	0.860	0.919	0.800
960	YOLOv11l	0.954	0.842	0.921	0.763
960aug	YOLOv8l	0.957	0.849	0.928	0.819
960aug	YOLOv11l	0.959	0.851	0.936	0.796
960	YOLOv8x	0.974	0.837	0.934	0.821
960	YOLOv11x	0.967	0.839	0.927	0.775
960aug	YOLOv8x	0.959	0.852	0.922	0.814
960aug	YOLOv11x	0.930	0.836	0.909	0.742

3.2 YOLOv11 Performance

In YOLOv11 models, it was observed that increasing model size and resolution does not always result in gains in detection metrics. For example, while moving from small to medium at 640×640 pixel, large and extra-large performed nearly identically.

Both YOLOv11s and YOLOv11m gave competitive numbers at both resolutions, but the additional gain was very clear when going from 640×640 pixel to 960×960 pixel. For recall and mAP50-95, improved performance is very evident at high spatial detail, suggesting a benefit from a higher resolution for moderate capacity models. Generally, augmentation contributed to higher mAP scores; however, as observed with YOLOv8, it occasionally incurs a penalty in precision (Table 1).

3.3 Comparative Insights

The comparative evaluation of YOLOv8 and YOLOv11 across four experimental conditions, 640×640 pixel without augmentation (Figure 1), 640×640 pixel with augmentation (Figure 2), 960×960 pixel without augmentation (Figure 3), and 960×960 pixel with augmentation (Figure 4), reveals several consistent patterns. At the baseline resolution of 640×640 pixel, YOLOv8 uniformly outperforms YOLOv11 for every model scale. Specifically, mAP50-95 scores for the small, medium, large, and extra-large YOLOv8 variants register at 0.678, 0.753, 0.765, and 0.788, respectively, compared to YOLOv11's 0.659, 0.724, 0.715, and 0.717 (Figure 1). This advantage—ranging from +0.02 for the smallest model to +0.07 for the extra-large model—demonstrates the superior capability of YOLOv8's architecture to resolve individual tree crowns even under lower-resolution constraints.

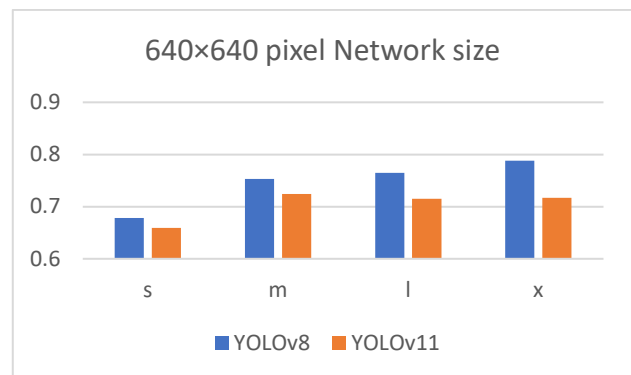


Figure 1. Baseline mAP50-95 comparison YOLOv8 versus YOLOv11 across small (s), medium (m), large (l), and extra-large (x) model variants at 640×640 pixel resolution without data augmentation.

When standard data-augmentation techniques are applied at 640×640 pixel, both architectures experience performance shifts (Figure 2). YOLOv8s and YOLOv8m benefit most (+0.02–0.03 mAP), while YOLOv8l exhibit slight declines (−0.02). In contrast, YOLOv11 displays more uniform gains of up to +0.03 across its model sizes. Despite these divergent responses to augmentation, YOLOv8 retains its lead across all scales.

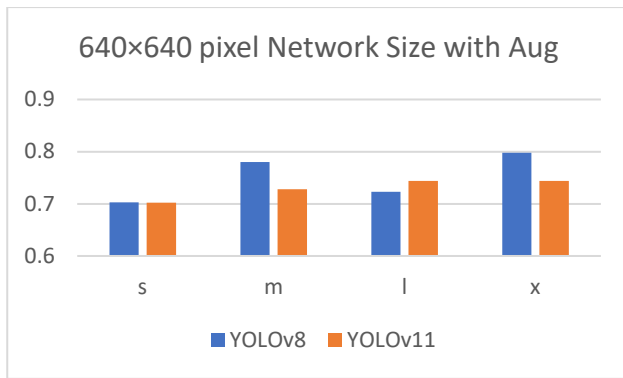


Figure 2. Impact of standard data-augmentation on mAP50-95 for YOLOv8 and YOLOv11 at 640×640 pixel resolution.

Escalating the input resolution to 960×960 pixel significantly elevates detection accuracy for both networks (Figure 3). YOLOv8's scores rise to 0.743, 0.790, 0.800, and 0.821, whereas YOLOv11 attains 0.741, 0.781, 0.763, and 0.775 for the corresponding scales. The extra-large YOLOv8 model, thus, enjoys a +0.04 mAP margin over YOLOv11x, underscoring its enhanced capacity to exploit finer spatial detail.

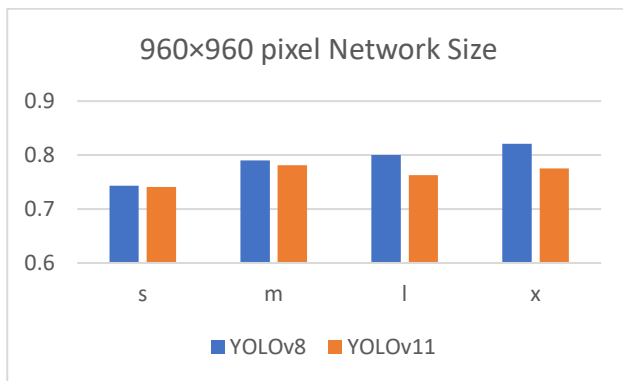


Figure 3. Baseline mAP50-95 comparison of YOLOv8 and YOLOv11 across all scales at 960×960 pixel input resolution without augmentation.

Finally, when high resolution is combined with augmentation (Figure 4), further gains are observed: YOLOv8l and YOLOv8x

achieve approximately +0.02 mAP improvements, while YOLOv11m and YOLOv11l record up to +0.03. Nevertheless, YOLOv8 continues to secure the highest absolute mAP50-95 values. Collectively, these findings confirm that YOLOv8 not only consistently surpasses YOLOv11 across diverse configurations, but also exhibits greater robustness to input resolution and augmentation strategies in complex forest-scene crown detection.

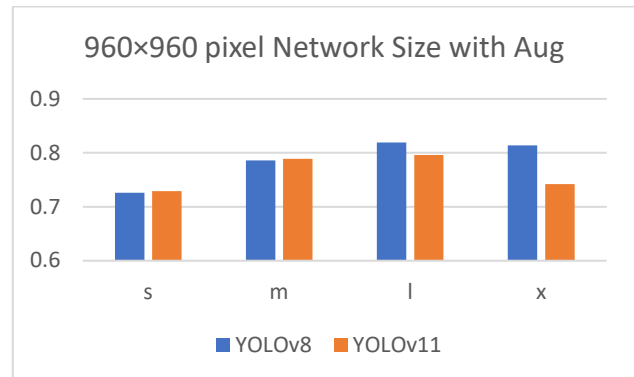


Figure 4. Combined effect of 960×960 pixel resolution and data augmentation on mAP50-95 for YOLOv8 and YOLOv11 across model sizes.

4. Discussion

The study reveals encouraging results with mAP50 scores up to 0.936 and precision levels above 0.95 in the best configurations, evaluating YOLOv8 and YOLOv11 models on UAV RGB and LiDAR-derived raster images for individual tree detection. Compared to Sun et al. (2022), who used an improved YOLOv4 model on tree height maps derived directly from airborne laser scanning data and reported an overall accuracy of 81.4% with scores of recall and precision at 83.6%, our models exhibited superior precision and mAP performance, especially at high resolution (960×960 pixel). Whereas Sun et al. (2022) have introduced advanced augmentation using generative adversarial networks (GANs) and fitted overlapping tree segmentation more precisely by paraboloid fitting, our pipeline was simplistic with augmentation like flipping and rotation, but nevertheless produced competitive or even better detection accuracy. This indicates considerable improvements in detection quality

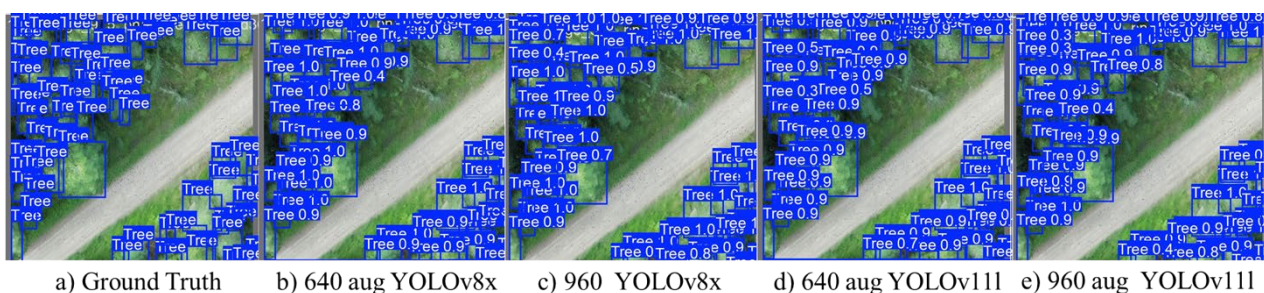


Figure 5. Best results for the YOLOv8 and the YOLOv11 models.

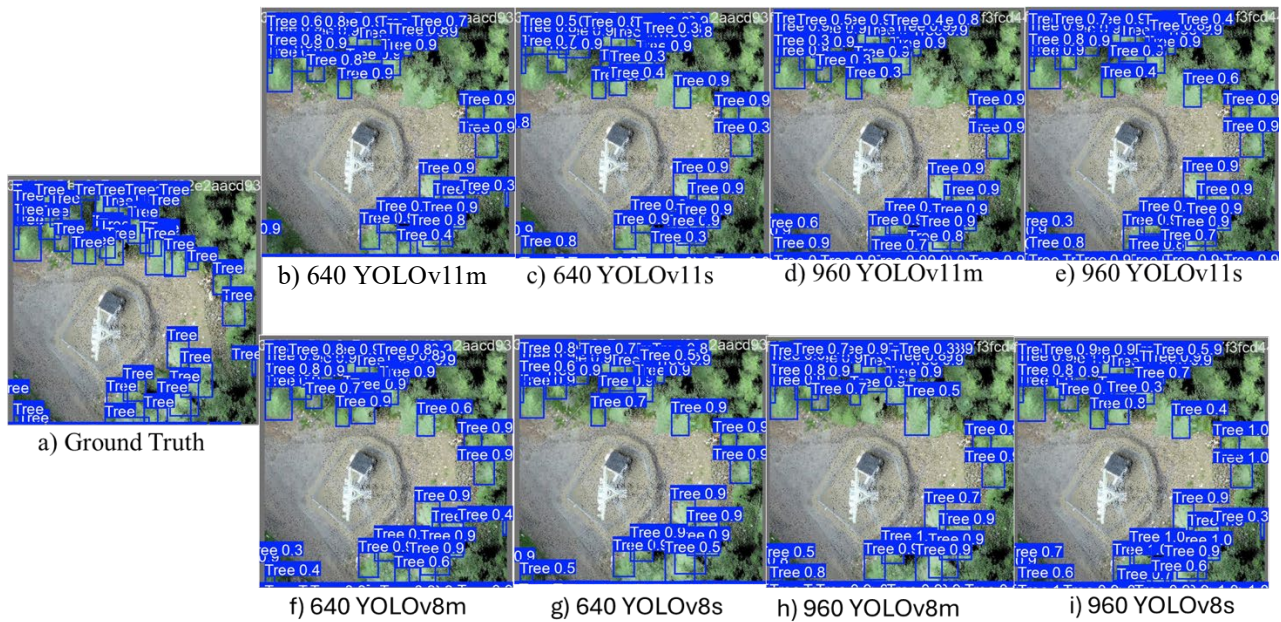


Figure 6. Effects of resolution on tree detection performance.



Figure 7. Effects of augmentation on tree detection performance.

when transitioning from YOLOv4 to newer versions such as YOLOv8, although trained on RGB rasterized LiDAR data instead of tree height maps, primarily focusing on optimized model size and resolution.

Unlike the YOLOTree model proposed by Luo et al. (2024), which combines UAV RGB imagery and LiDAR point clouds for individual tree detection and crown volume modeling, our study focuses solely on individual tree detection performance of YOLOv8 and YOLOv11 architectures applied to rasterized RGB images derived from LiDAR data. While Luo et al. (2024) obtained a mAP50-95 of 52.8% on the TreeLD dataset with their lightweight YOLOTree model, our investigations revealed that YOLOv11l (960×960 pixel with augmentation) reached up to 0.936 mAP50 and 0.796 mAP50-95, indicating a significant increase in accuracy.

Furthermore, our models' accuracy regularly approached about 0.95, exceeding YOLOTree's 90.5% under comparable lightweight restrictions.

While Satama-Bermeo et al. (2025) employed UAVs and AI, specifically YOLOv4 and Faster R-CNN, to automate road signage inventory, our work investigates the recognition of individual trees from UAV RGB and LiDAR-derived raster images using more contemporary models, YOLOv8 and YOLOv11. Both investigations use aerial images to detect objects; however, the target domains, road signs and natural forest elements, present distinct problems in terms of item scale, occlusion, and unpredictability. Unlike signage, which has uniform forms and characteristics, trees have natural inconsistencies, making detection more difficult.

Satama-Bermeo et al. (2025) found mAP scores ranging from 74% to 95% in an urban context using YOLOv4-based models, which were frequently used in RGB or multisensory configurations. Our top-performing setup, when applying YOLOv8x with 960×960 pixel resolution and augmentation, had a mAP50 of 0.934 and a mAP50-95 of 0.821, showing equivalent or even greater performance in the setting of vegetation structure, where trees frequently overlap and vary in size and density.

Furthermore, although the studied signage research focuses on real-time performance and model deployment issues in complicated urban scenarios, our findings indicate that high detection accuracy can be achieved in forested environments using rasterized 2D RGB from LiDAR point cloud data. These findings indicate that updated YOLO designs provide generalizable performance and can serve as strong backbones for environmental inventory tasks like tree recognition.

Figures 5–7 support the quantitative findings. Figure 5 shows the best results for YOLO8 and YOLO11 models. Figure 6 shows the effect of resolution on tree detection performance. In Figure 7, models trained with flip and rotation show improved crown identification under varying orientations and overlapping canopies, especially for smaller YOLO variants. However, slight over-detection in some cases suggests that geometric transformations can introduce small localization noise, which is consistent with the observed decreases in precision. On the other hand, Figure 6 clearly shows the benefit of using higher resolution inputs (960×960 pixel), where finer crown details and tree boundaries are more accurately detected, especially in dense forest patches. Increased spatial accuracy at 960×960 pixel resolution improves the model's ability to discriminate closely located crowns, which is reflected in higher recall and mAP50-95 scores. These qualitative observations confirm that both magnification and resolution play complementary roles in improving model robustness and detection accuracy in complex forest scenes.

5. Conclusions

This study shows how well the YOLOv8 and YOLOv11 object identification architectures can distinguish individual trees from RGB raster images taken from UAV LiDAR point cloud data in order to improve forest management (Fransson et al., 2023). We found that YOLOv8 consistently outperforms YOLOv11 in terms of accuracy, recall, and both mAP50 and mAP50-95 scores after conducting a comprehensive examination of model sizes (s, m, l, x), input resolutions (640×640 pixel and 960×960 pixel), and data augmentation procedures. With an accuracy of 0.959, recall of 0.851, mAP50 of 0.936, and mAP50-95 of 0.796, the greatest detection performance was specifically attained by the YOLOv8x model, which was trained at 960×960 pixel resolution without augmentation. These findings demonstrate the model's strong generalization capabilities in complex forest environment, characterized by overlapping crowns and diverse canopy structures.

However, there are limitations. Although using rasterized RGB images from the LiDAR point cloud data makes CNN input simpler, it does not take into account the point clouds' full 3D spatial richness. Furthermore, the findings applicability to other forest types or biomes is restricted by the very limited sample size (82 images). By combining spectral and structural inputs, merging multi-temporal data, or broadening the dataset over many ecological zones, future research might overcome these constraints. Moreover, adding species categorization, tree health evaluation, and 3D crown reconstruction might improve the

usefulness of the models in ecological monitoring and operational forestry. This paper offers a quick, precise, and scalable method for detecting individual trees using data from UAV RGB and laser scanning data, laying a solid basis for the deployment of sophisticated YOLO designs in forest applications.

This study represents one of our initial implementations, and due to time and venue constraints, extensive comparative experiments are beyond the current scope. We plan to conduct these comparisons in future studies by testing our methods on benchmark datasets or reimplementing prior methods on our data for fair evaluation.

Acknowledgements

This work was part of the ForestMap project which is supported under the umbrella of ERA-NET Cofund ForestValue by the Swedish Governmental Agency for Innovation Systems, the Swedish Energy Agency, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning, Academy of Finland, and the Scientific and Technological Research Council of Turkey. ForestValue has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 773324 and from TUBITAK Project No. 221N393.

References

- Fransson, J.E.S., Sertel, E., Ünsalan, C., Salo, J., Holmström, A., Wallerman, J., Nilsson, M., 2023. ForestMap: Mapping Forest Attributes Across the Globe - First Case Study. *2023 IEEE Int. Geoscience & Remote Sensing Symp.*, 3395–3397. doi.org/10.1109/IGARSS52108.2023.10283354
- Jocher, G., Qiu, J., Chaurasia, A., 2023. Ultralytics YOLO, Version 8.0.0. ultralytics.com (20 July 2025).
- Luo, T., Rao, S., Ma, W., Song, Q., Cao, Z., Zhang, H., Xie, J., Wen, X., Gao, W., Chen, Q., Yun, J., Wu, D., 2024. YOLOTree-Individual Tree Spatial Positioning and Crown Volume Calculation Using UAV-RGB Imagery and LiDAR Data. *Forests* 15, 1375. doi.org/10.3390/f15081375
- Satama-Bermeo, G., Lopez-Guede, J.M., Rahebi, J., Teso-Fz-Betño, D., Boyano, A., Akizu-Gardoki, O., 2025. PRISMA Review: Drones and AI in Inventory Creation of Signage. *Drones* 9, 221. doi.org/10.3390/drones9030221
- Sertel, E., Topgul, S.N., 2025. Comparative Analysis of Deep Learning Approaches for Forest Stand Type Classification: Insights from the New VHRTreeSpecies Benchmark Dataset. *International Journal of Digital Earth* 18, 2522394. doi.org/10.1080/17538947.2025.2522394
- 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 Journal of Photogrammetry and Remote Sensing* 9, 100045. doi.org/10.1016/j.ophoto.2023.100045
- Sun, C., Huang, C., Zhang, H., Chen, B., An, F., Wang, L., Yun, T., 2022. Individual Tree Crown Segmentation and Crown Width Extraction from a Heightmap Derived from Aerial Laser Scanning Data Using a Deep Learning Framework. *Front. Plant Sci.* 13, 914974. doi.org/10.3389/fpls.2022.914974

Terven, J., Córdova-Esparza, D.-M., Romero-González, J.-A., 2023. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning & Knowledge Extraction* 5, 1680–1716. doi.org/10.3390/make5040083

Topgül, Ş.N., Sertel, E., Aksoy, S., Unsalan, C., Fransson, J.E.S., 2025. VHRTrees: A New Benchmark Dataset for Tree Detection in Satellite Imagery and Performance Evaluation with YOLO-Based Models. *Front. For. Glob. Change*, 7:1495544. doi.org/10.3389/ffgc.2024.1495544