*Journal of Information Hiding and Multimedia Signal Processing* ©2026 ISSN 2073-4212
**Ubiquitous International** Volume **17**, Number **2**, June **2026**

# Automated Tree Counting and Geo-Referencing Using YOLO-Based Deep Learning on UAV Imagery

Jignesh Bhavsar[1,*], Amrutbhai Patel[2]

[1]Assistant Professor, Electronics and Communication Department,
Government Engineering College,
Sector 28, Gandhinagar, Gujarat, India
[2]Associate Professor, Electronics and Communication Department,
Ganpat University – U. V. Patel College of Engineering,
Mehsana, Gujarat, India

*Corresponding author: Jignesh Bhavsar (jignesh@gecg28.ac.in)

Received January 26, 2026, revised April 4, 2026, accepted April 16, 2026.

Abstract. Accurate tree detection and counting using aerial imagery is essential for forest inventory, urban planning, plantation monitoring, and ecological assessment. Conventional manual tree enumeration is time-consuming and error-prone, especially over large areas. This paper presents an automated framework for tree detection and counting using a YOLO-based deep learning model applied to geo-referenced UAV imagery. The proposed approach detects individual trees from UAV-captured images and estimates their geo-locations by associating detected bounding boxes with geographic coordinates derived from UAV metadata and pixel-to-ground scale mapping. The method is evaluated against ground truth tree counts and geo-referenced coordinates generated using ArcGIS. Experimental results demonstrate reliable detection and consistent geo-referencing accuracy for the surveyed region. The proposed method estimated the tree count with 92% accuracy with a mean geo-referencing error of $1.25 \pm 0.35$m showing the potential in significantly reducing the manual work required for tree inventory.
**Keywords:** UAV imagery, Tree detection, Tree counting, YOLO, Deep learning, Geo-referencing, ArcGIS, Remote sensing.

1. **Introduction.** The necessity to oversee, control, and preserve the vegetation of our planet has never been more urgent than it is now all over the world. Being one of the basic elements of terrestrial and urban ecosystems, trees are the important indicators of ecological health and stability. They are major drivers of carbon consumption and they capture a large part of the anthropogenic carbon dioxide emissions hence a key factor in mitigating the effects of climate change [1]. Global land-use change and its environmental consequences have also been extensively studied [2]. They anchor biodiversity, which offers the necessary habitats to many species, which is vital in ensuring that hydrological cycles take place, soil erosion, and local and regional climatic balances are maintained[3]. The urban forest, also known as the canopy of trees, offers an essential service to the urban landscape, such as alleviation of the urban heat island effect, elevation of air and water quality, fewer noise sources, and improvement of health and well-being of the population[4]. Urban forest ecosystems also provide significant socio-economic and environmental benefits [5]. The necessity to quantify, efficiently and repeatedly the number, distribution, and location of trees accurately, efficiently and repeatedly is therefore

not just an academic activity; but a basic resource to effective environmental stewardship, sustainable resource management and evidence-based policy-making in the 21st century.

The need to have accurate vegetation information is immense at the national level with forest inventory to inform carbon credit markets and global climate treaties, at the municipal level with green space plans to provide equal access to nature. This information supports a vast number of applications of critical nature. It is used by ecologists to monitor the rates of deforestation and afforestation, ecosystem fragmentation and the distribution of species[6]. Forest managers apply it in estimation of timber volumes, outbreak of pests and diseases, and in planning of the sustainable harvesting activities[7]. It is relied on by urban planners to fulfill the green cover mandates, infrastructure conflict management, and provision of healthier and resilient cities [8]. In the absence of credible and current tree inventories, it is through this that any action that seeks to combate these complicated issues will always be reactive, under-informed and small scale in its effects. The main problem, however, is not to understand the significance of counting trees, but to come up with the methodologies capable of providing this crucial information in the needed scale, speed, and accuracy demanded by the practice of environmental science and management today.

The concept of the tree inventory has changed considerably throughout decades due to the improvement of technologies and the growing demand of massive data. Comprehensive methodologies integrating sampling, GIS, and remote sensing for forest inventory have been discussed in earlier studies [9]. The manual field survey has traditionally been the gold standard of tree counting. In this technique, a group of surveyors go to the location and use equipment such as measuring tapes, dendometers, and GPS equipment to note the location, species, size, and health of individual trees at its exact location. The main benefits of ground survey are that it can be high-accuracy and abundant attributes information about each tree can be collected [10]. On small, clearly defined plots such an approach is unsurpassable in terms of accuracy. But the shortcomings of the manual field survey become clearly visible when used on bigger or more inaccessible areas. Such avenues are amazingly time-consuming, labor-intensive, and, therefore, extremely costly [11]. A detailed survey of a significant forest or the whole city can be achieved only after months or even years, and by that time the information may be obsolete because of the natural development, death or developing action of people. In addition, ground surveys can be inconvenient or hazardous in isolated or rough or even in areas with political instability. They also tend to have inconsistencies due to sampling bias- where the sampled plots might not represent the whole area- and inter-observer variability whereby the various survey parties might gather data with slight but important difference [12]. These limits make conventional methods economically impractical and logistically unfeasible in terms of the constant and extensive coverage of an area that is currently being demanded.

Having identified these drawbacks, scientists and practitioners have resorted to remote sensing technologies. Having identified these drawbacks, scientists and practitioners have resorted to remote sensing technologies. Remote sensing has long supported large-scale forest inventory programs [13]. Initial uses were done in the satellite images that provided an unprecedented synoptic image of the Earth surface. Although satellite data can still be invaluable in mapping the scale of forest cover and land-use change [14], the spatial resolution has historically not been suitable to detect and count Individual Tree Crown (ITC) in a heterogeneous forest or in an urban setting. This has been enhanced more recently by the advent of Very High Resolution (VHR) satellite imagery which can provide this

capability [15] although at a high cost and restricted by cloud cover, making it less useful on-demand. The next more crucial technological advancement was the implementation of Light Detection and Ranging (LiDAR). LiDAR systems by measuring these returns and emitting laser pulses can create a very fine-level three-dimensional point cloud of the landscape and thus accurately measure tree biomass, height, and crown structure. Airborne laser scanning techniques have been widely used for extracting detailed forest inventory data [16]. LiDAR is capable of creating very precise numbers of trees and structural data when it is deployed out of an airplane [17]. Nevertheless, similarly to its predecessors, such technology has serious disadvantages. LiDAR sensors and aerial surveys are quite expensive, and not every research group and agency can afford to purchase and use them. Complex 3D point cloud data also demand specialized software and high amount of computational expertise which makes entry highly barrier to entry [18].

Biomass, height, and crown structure. LiDAR can produce highly accurate quantities of tree counts and structural data in applications as an aerial tool[17]. However, just like its counterparts, this technology has grave downsides. LiDAR sensors and aerial surveys are quite costly and cannot be afforded by the majority of research entities and agencies. Specific software and a high level of computer expertise which is also a high barrier to entry are also required to work with complex 3D point cloud data [18].

The meeting point of two parallel technological revolutions, in Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence (AI) has recently opened a ground-breaking new horizon to environmental monitoring. Drones, also referred to as UAVs, have made aerial data collection democratic by providing a potent trade-off between detailed information on the ground and extensive coverage [19]. However, challenges in UAV data acquisition in complex forest environments have also been reported [20]. They may be launched in minutes, on short notice, and at a fraction of the price of manned planes. They can fly at low altitudes, which allows capturing images with ultra-high spatial resolutions, and this detail level allows distinguishing the crowns of separate trees [21]. Recent advancements in image super-resolution using generative adversarial networks further enhance UAV image quality [22], enabling improved detection performance [6]. Yet, their capability of producing large quantities of high-resolution image data is also the real strength of UAVs, which is a severe analytical bottleneck. It is at this point that the second revolution, AI, is needed. The progress in computer vision, especially in the field of deep learning and Convolutional Neural Networks (CNNs), has offered the momentum to perform such an analysis automatically [23]. Systems Object detection models (including the You Only Look Once (YOLO) family) have shown a high level of success in object identification in real time, and their use has quickly been extended to ecological monitoring [24].

Although there is a rapid development in computer vision , modern algorithms tend to ignore the process of converting image-space detections into real-world metric coordinates. More recent state-of-the-art research, including [25, 26], focuses on mean Average Precision (mAP) but considers geolocational validation to be a visual or secondary task. In particular, the literature does not demonstrate a single, quantitatively consistent framework which is able to: (1) process raw UAV images without the need for ortho-mosaic transformation which is computationally burdensome and noisy, (2) quantify measurements of geo-referencing errors (such as RMSE), (3) do so in a manner which systematically removes duplicates. This research addresses this gap through a two-step pipeline, which utilize YOLOv11 for tree detection and assign geo-referencing to each detected tree, (4)

verify the proposed method using real data set collected from Chandrala site of Gandhinagar District in Gujarat. The proposed solution is more scalable and cost-effective alternative to LiDAR-based workflows in recent research [25].
Unlike conventional approaches that rely on generating computationally intensive orthomosaics, the proposed framework directly estimates geographic coordinates from raw UAV image metadata combined with pixel-to-ground scale transformation, thereby significantly reducing processing time and computational overhead while maintaining geo-referencing accuracy.

The study contribution is as follows:

- To use a state-of-the-art object detector algorithm for identification of individual trees in UAV imagery.
- To propose geo-referencing process for accurate mapping of geographic coordinates of each detected tree.
- To formulate a mechanism for filtering the duplicate detections of same tree in multiple subsequent images from UAV. It assist in precise counting of trees in a surveyed area.
- The results are presented to verify the proposed method. It shows the effectiveness of the proposed method as evidenced from the analysis of real data captured from a field site. The methodology is verified using the geo-references of trees collected from ArcGIS software, a standard followed in the domain.

The rest of this paper is structured in the following way. The relevant studies in the field are summarized in the Literature Review section. The detailed methodology is presented in Methodology section. The implications of these findings are discussed in the Result and Discussion section.

2. **Literature Review.** A shift in the methods of vegetation monitoring towards the automated process is an important paradigm shift in the sphere of environmental science and forestry. Traditional ground-based inventories have been known to be labor-intensive, prohibitively expensive, and unrealistic in terms of large-scale or repeated measures and are commonly perceived to generate outdated or spatially biased data [11]. Earlier image classification approaches relied on traditional machine learning techniques such as decision trees combined with feature extraction methods like Hough transform and genetic algorithms [27]. Aerial imaging from UAV has paved the way to Individual Tree Crown (ITC) detection, which was quite difficult using coarser remote sensing data.

Recent studies have demonstrated the detection of trees from UAV images, however, a significant disjuncture is present in simultaneous detection and mapping them into high-fidelity geo-referencing. As an example, [25] have conducted a comparative study between YOLOv8 and YOLOv11 architectures in identifying trees, but their experiment was based on pre-rasterized images obtained through high-precision LiDAR solutions. Although it is useful in terms of detection benchmarks, it does not focus on the georeferencing issues of raw RGB imagery based on the typical consumer-grade UAVs. On the same note, [26] were able to use YOLOv8 to count Camellia oleifera trees successfully, and their research was based on Orthomosaic (DOM) datasets. Although the advantage of Orthomosaics is that it makes managing long canopy edges geometrically distorted and looks aesthetically better, Orthomosaics is computationally expensive to stitch and needs lots of pre-processing time, which can be avoided by our framework since it processes the raw images through a new coordinate transformation pipeline.

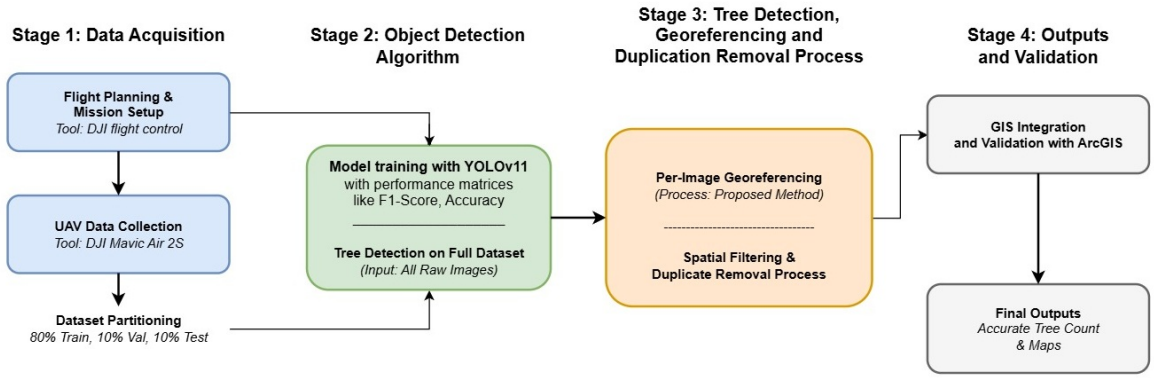


FIGURE 1. Workflow of proposal

Studies have explored the analysis of tree detection which lacks the robust evaluation of detection model due to limited data availability. [28] filled this gap on the unlabeled agricultural data by creating synthetic training environments of counting palms using YOLOv7. Although their use of synthetic data enhances the robustness of the model, the study does not include a mechanism to deal with the redundancy of space in the high-overlap ($\leq$70%) flight paths which may cause overestimation of wide-area surveys. Additionally, [29] investigated the application of Swin Transformers and YOLO in detecting trees, and they were used on Very High Resolution (VHR) satellite images, however, with limited accuracy in geo-referencing of detected tree.

In spite of these achievements in object detection, a critical review of the literature shows that there is a great gap in the creation of an entirely integrated, end-to-end inventory workflow. Most of the current literature focuses mainly on the metrics of detection accuracy (e.g. precision, recall, F1-score) and the next step of georeferencing is seen as a secondary or unaddressed issue. It is not a trivial process of providing a set of geographic coordinates to each tree that has been detected. Although techniques that employ orthomosaics are available, these technologies may cause geometric errors and coarse out fine-scale details of the canopy and may undermine the locational accuracy. More to the point, an even more common problem which has often been addressed improperly is that of duplicate detection. The protocols of Standard UAV data acquisition include a large image overlap (generally larger than 70 %) in order to permit full coverage and be able to reconstruct 3D. One of the side effects is that one tree can be captured and detected in several overlapping images. The end result is that these redundant detections may be hugely overestimated by a weak spatial filtering mechanism, and inventory is therefore no longer useful in practical contexts [30]. As a result, an actual research gap remains regarding a holistic framework that goes beyond its detection. There is a pressing need to develop methodology that do not only detect trees but also combine a geo-referencing process. The primary purpose of this study is to fill this gap by integrating both processes to ensure ready availability of tree inventory .

3. **Methodology.** This paper hypothesizes and confirms a two-phase framework of automated tree inventory of aerial images. The workflow will utilize a state-of-the-art deep learning model to initially identify individual tree crowns (ITCs) and then to obtain precise geographic coordinates of the latter with the help of a novel geo-referencing algorithm, which will incorporate spatial filtering. The general methodology is represented in Figure 1 and is described in the subsections below.

3.1. **Study Area and Data Acquisition.** The framework was built with and tested on a high resolution aerial dataset that was obtained in one of the study locations of Chandrala in the district of Gandhinagar in the state of Gujarat in India. This location has a mixed rural and semi-urban terrain that is heterogeneous.

The data was collected with a DJI Mavic Air 2S Unmanned Aerial Vehicle (UAV) which has a 1-inch CMOS sensor with the capability to record 20-megapixel still images and 5.4K video. Flights of the UAV were done at 90 meters in order to have adequate detail in the detection of ITC. The flight plan was set in such a manner that it would have a 70% forward and 60% side overlapping of the sequential images, which is essential in reducing data lapses and assisting the process of georeferencing. The stored images are of 5472 $\times$ 3648 pixels spatial resolution and 72 DPI horizontal and vertical resolutions. The metadata, such as the latitude and longitude of the UAV and its altitude at the time of the image capture is embedded in the EXIF metadata of each raw image that is about 12 MB in size and in the JPG format. Hence, the experiment was conducted using a dataset collected from Chandrala, Gujarat, with imagery acquired using a DJI Mavic Air 2S sensor, comprising a total of 478 images and 849 labeled tree instances.

3.2. **Model training.**

3.2.1. *Data Pre-processing and Annotation.* To train the supervised object detection model, the raw UAV images were labeled by mannual process to produce the ground-truth labels. LabelImg, an open source program, was applied to manually label the location of every crown of the trees visible with a bounding box. This conversion created annotations in the standard YOLOv11 format, consisting of a class label (tree) and normalized box coordinates (x center, y center, width, height) of the bounding box. It is based on this annotated dataset in order to train a model to recognize and locate trees and detect them accurately.

3.2.2. *Hyper parameters training strategy.* Various hyperparameter optimization techniques such as random search, manual tuning, evolutionary algorithms (used in YOLO), Bayesian optimization, and Hyperband are available. Nevertheless, grid search was adopted in this work due to its systematic and exhaustive evaluation of parameter combinations, enabling effective optimization of the F1-score and reduction of validation loss.

Our proposed framework consists of two primary stages: (1) ITC detection via YOLOv11 model and (2) a precise georeferencing.

3.2.3. *ITC detection via YOLOv11 model.* In the case of ITC detection, this paper utilized the YOLOv11, which is the state-of-the-art one-stage detection model with a high level of accuracy and efficiency in computing. Its design that combines a hybrid Transformer-CNN back-bone and an anchor-free detection head is especially effective when it comes to detecting objects of different sizes and in the presence of tricky backgrounds.

The architecture comprises:

- CNN-Transformer based feature extractor hierarchy.
- feature pyramid and path aggregation neck to aggregate the features across multi-scales; and
- a decoupled, anchor-free classification and regression head for single class 'Tree'.

The network was also instantiated with weights that had been trained with the MS COCO data.

Random shuffling of the annotated dataset was done to avoid bias by partitioning the data into training (80%), validation (10%), and testing (10%) subsets respectively. The

FIGURE 2. Image for Annotation

FIGURE 3. Annotated tree

batch size and initial learning rate used to train the model are 16 and 0.001 respectively. The training was performed by adopting the ADAM optimizer on a machine with an NVIDIA Tesla T4 graphical processing unit (GPU) and 64 GB RAM with the help of the graphics processor, which helped to complete the process within the framework of about one hour.

The training of model was done in the following settings:

- Initial learning rate: 0.001
- Momentum: 0.937
- Weight decay: $5 \times 10^{-4}$
- Batch size: 16
- Number of epochs: 50

The model is evaluated using standard performance metrics, including F1-score (derived from Precision and Recall) and Accuracy. Figure 4 shows confusion matrix evaluates the

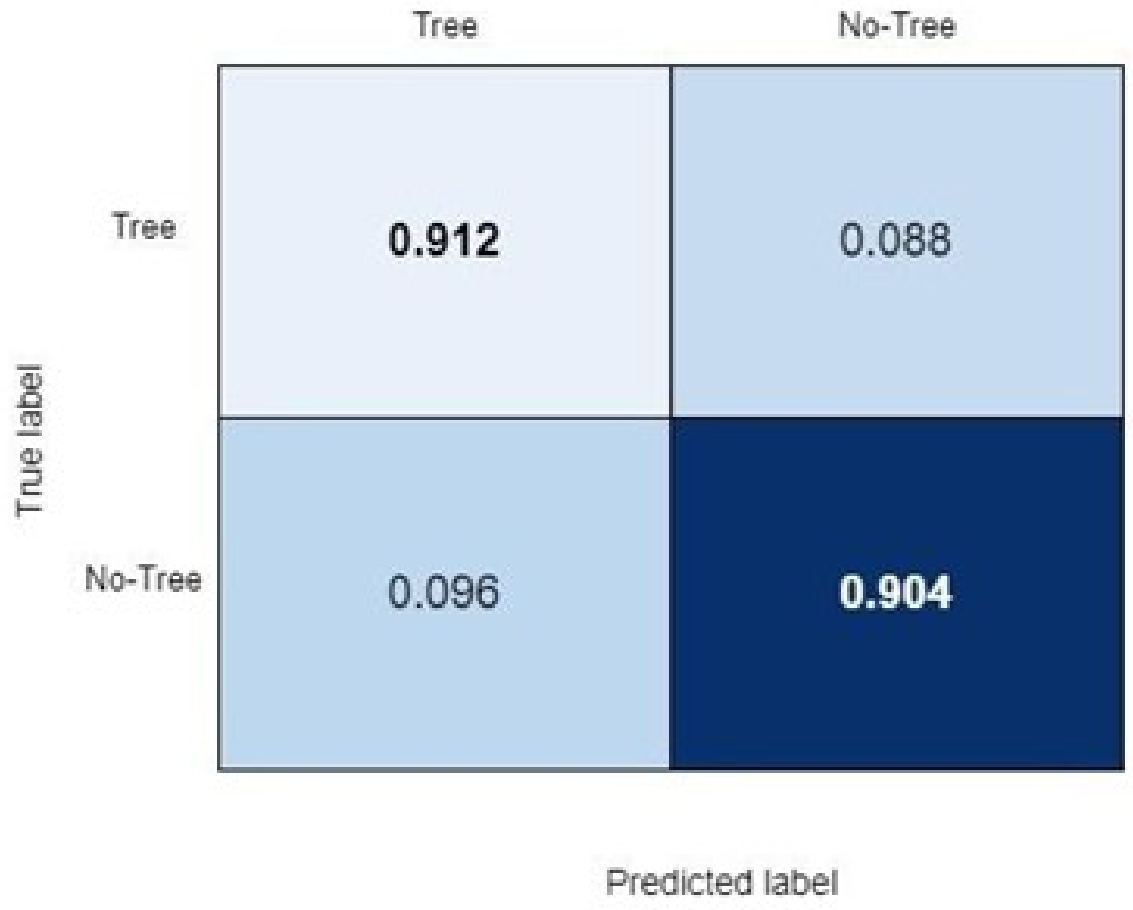


FIGURE 4. Confusion matrix of model

model's ability to distinduish the 'Tree' class from the background (No-Tree). To further assess detection performance, the Precision-Recall (PR) curve is presented in Figure 5. The model achieves a mean Average Precision (mAP) of 90.8%, indicating high detection accuracy. The PR curve demonstrates that precision remains above 0.90 across most of the recall range, highlighting the model's ability to minimize false positives while effectively detecting individual tree crowns.

As per confusion matrix,

- Precision $\approx$ 91%
- Recall/Sensitivity $\approx$91%
- F1 Score $\approx$ 92%
- Accuracy $\approx$ 92%
- mAP $\approx$ 90.8%

Novelty of proposed method is unlike conventional approaches that rely on generating computationally intensive orthomosaics, the proposed framework directly estimates geographic coordinates from raw UAV image metadata combined with pixel-to-ground scale transformation, thereby significantly reducing processing time and computational overhead while maintaining geo-referencing accuracy

3.3. **Georeferencing of Trees.** This section presents the mathematical formulation for transforming pixel centroid positions into the WGS84 coordinate system, facilitating geo-referencing of detected objects from image space to real-world geographic coordinates.

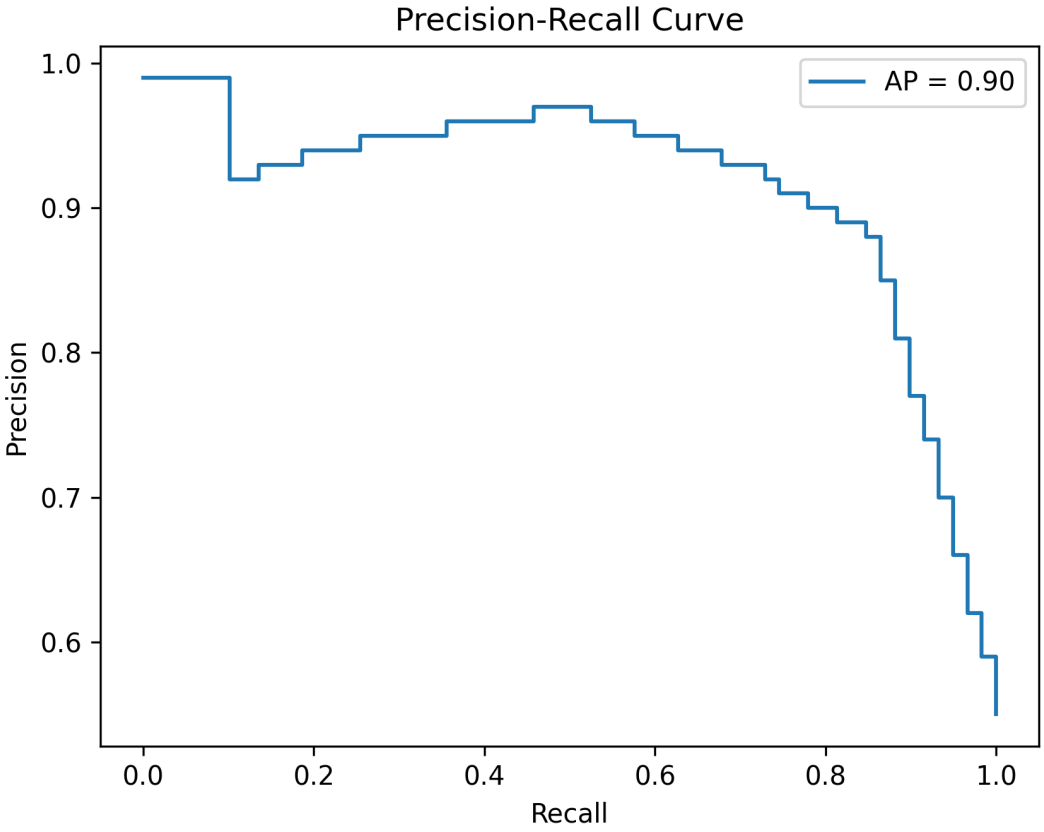


FIGURE 5. AP curve of model

3.3.1. *Calculation of Derived Parameters to Measure Ground Distance.* Drone images are captured in pixel coordinates, whereas geolocation and physical measurements are required in meters. To bridge this gap, intermediate parameters such as effective pixel size, real-world displacement, and yaw-corrected ground distance are derived. These parameters account for camera geometry, UAV flight altitude, and drone orientation, enabling pixel-level detections to be translated into accurate ground measurements.

3.3.2. *Effective Pixel Size.* The effective pixel size ($P$) defines the real-world distance represented by a single pixel. It is computed by effective altitude $H$, and the camera sensor width, image width, and focal length:

$$P = \frac{S_w}{I_w} \times \frac{H}{F} \tag{1}$$

where:

- $H$ = effective altitude (m)
- $S_w$ = physical width of the camera sensor (m)
- $I_w$ = image width (pixels)
- $F$ = focal length of the camera lens (m)
- $P$ = real-world distance represented by one pixel (m)

3.3.3. *Conversion of Normalized Coordinates to Pixel Coordinates.* The normalized bounding box coordinates obtained from YOLO are converted to pixel coordinates as follows:


FIGURE 6. Depiction of Yaw angle

$$x_p = x_n \times I_w \tag{2}$$

$$y_p = y_n \times I_h \tag{3}$$

where:

- $(x_p, y_p)$ = pixel coordinates of the detected tree
- $(x_n, y_n)$ = normalized coordinates of the detected tree
- $I_w$, $I_h$ = image width and height in pixels

3.3.4. *Distance from Image Center to Detected Tree.* The Euclidean distance from the image center to the detected tree in pixel space is computed as:

$$k = \sqrt{(x_p - x_c)^2 + (y_p - y_c)^2} \tag{4}$$

The real-world distance ($\alpha$) is then obtained by scaling with the effective pixel size:

$$d = k \times P \tag{5}$$

where:

- $k$ = distance in pixels
- $d$ = real-world distance from the drone position to the detected tree (m)
- $(x_c, y_c)$ = pixel coordinates of the image center

3.3.5. *Yaw Angle Correction.* The yaw angle ($\theta$) represents the rotational orientation of the drone relative to true north. To align the displacement vectors with geographic directions, a rotation is applied:

$$\alpha_x = d_x \cos(\theta) - d_y \sin(\theta) \tag{6}$$

$$\alpha_y = d_x \sin(\theta) + d_y \cos(\theta) \tag{7}$$

where:

- $(\alpha_x, \alpha_y)$ = yaw-corrected displacement components
- $(d_x, d_y)$ = displacement vector components in the image plane
- $\theta$ = yaw angle

3.3.6. *Tree Coordinate Correction Based on Height.* To account for the vertical discrepancy between the drone altitude and the tree height, a correction factor is applied:

$$\alpha_c = \frac{h}{H} \times d \tag{8}$$

where:

- $\alpha_c$ = corrected horizontal displacement (m)
- $h$ = average tree height (m)
- $H$ = effective drone altitude (m)

The corrected displacement components are given by:

$$\alpha_{cx} = \alpha_x - \frac{\alpha_x}{\alpha} \times \alpha_c \tag{9}$$

$$\alpha_{cy} = \alpha_y - \frac{\alpha_y}{\alpha} \times \alpha_c \tag{10}$$

3.3.7. *Final Geolocation of Trees.* The corrected displacements are added to the drone's GPS position to obtain the final geographic coordinates of each detected tree:

$$\text{Lat}_{\text{tree}} = \text{Lat}_{\text{center}} + \alpha_{cy} \tag{11}$$

$$\text{Long}_{\text{tree}} = \text{Long}_{\text{center}} + \frac{\alpha_{cx}}{\cos(\text{Lat}_{\text{center}})} \tag{12}$$

where:

- $\text{Lat}_{\text{tree}}$, $\text{Long}_{\text{tree}}$ = latitude and longitude of the detected tree
- $\text{Lat}_{\text{center}}$, $\text{Long}_{\text{center}}$ = GPS coordinates of the image center

3.3.8. *Removal of Repeated Tree Detections.* Due to overlap between consecutive aerial images, the same tree may be detected multiple times. To eliminate duplicate detections, trees detected in sequential images were spatially matched using their real-world coordinates. If the distance between two detections was less than 5 m, they were considered to represent the same tree, and the duplicate detection was removed.

A sensitivity analysis was conducted by varying the spatial deduplication threshold between 1 m and 9 m. Thresholds below 3 m resulted in over-counting due to GPS positional jitter (±2.5 m) of consumer-grade UAV sensors. Thresholds above 7 m led to excessive merging of neighboring trees, reducing recall in dense vegetation areas. A threshold of 5 m was identified as the optimal inflection point from Table 1. At this distance, the framework achieved the highest F1-score (0.92), balancing duplicate removal and accurate individual tree detection.

Table 1. Sensitivity analysis of the spatial deduplication threshold (D) on tree counting accuracy

| Threshold (D) | Estimated Count | True Count | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 1 m | 1,420 | 849 | 0.59 | 0.98 | 0.74 |
| 3 m | 1,055 | 849 | 0.80 | 0.95 | 0.87 |
| **5 m** | **876** | **849** | **0.91** | **0.92** | **0.92** |
| 7 m | 720 | 849 | 0.94 | 0.79 | 0.86 |
| 9 m | 580 | 849 | 0.96 | 0.65 | 0.77 |

The duplication and the following elimination of redundant detections are shown in Figure 7. It depicts tree detections of Image 1 with a yellow dot, and Image 2 with a

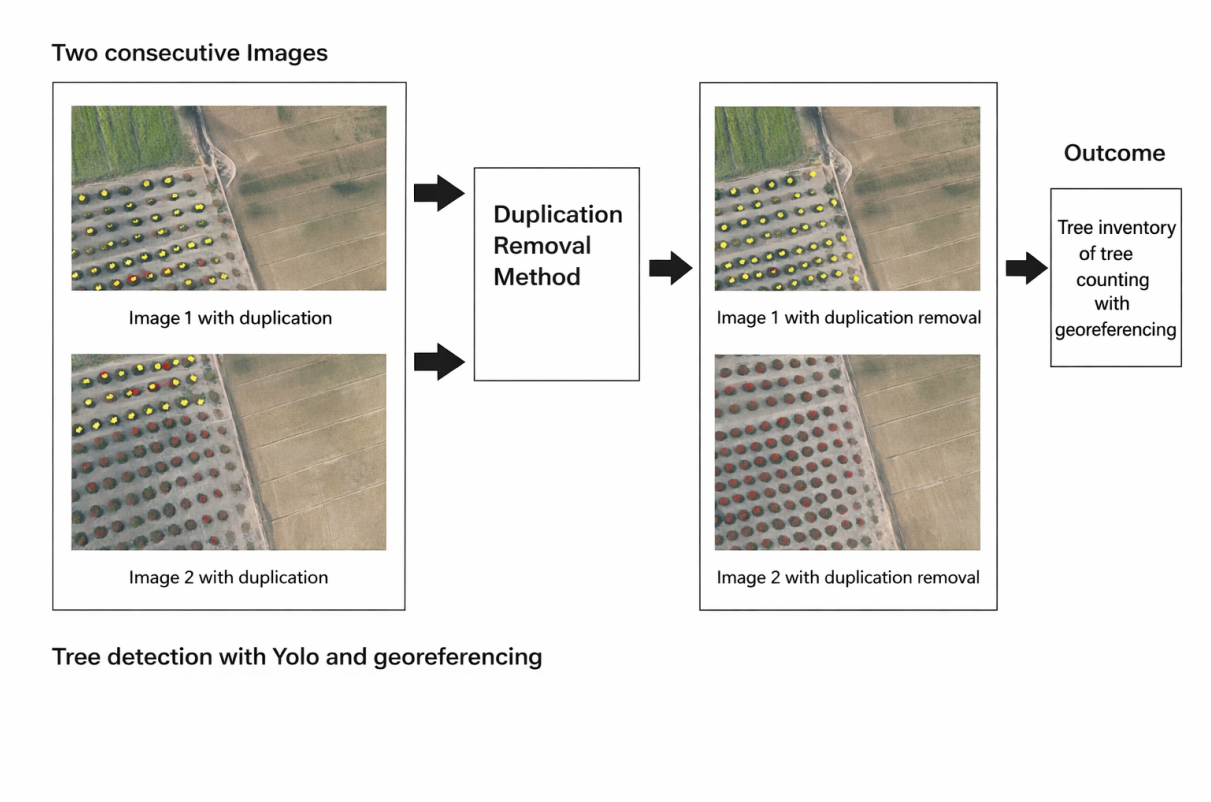


FIGURE 7. Duplication of detections of trees and removal of duplicated detection of trees

red dot. The overlapped area also displays trees that are detected in both images with both yellow dots and red dots, which is a duplicate detection, and the 5-meter threshold is used to eliminate them.

4. **Results and Discussion.** The following section shows the empirical findings based on the implementation of the proposed automated framework of tree detection and georeferencing to the survey area of 35000 $m^2$ of Chandrala district in Gujarat. The dataset is made up of total 478 images and frequency of data collection was fixed at 60 % overlapping areas. The drone was set at 90m and using WGS coordinate. Figure 8 indicates surveyed area which was chosen among the sight.

The finding from the collected data using presented methodology is organized sequentially, from the performance of the detection model to the generation of the final georeferenced tree inventory.

4.1. **Model Performance and Geolocation Validation.** The quality of the ITC detection model was strictly measured with respect to the classical measures of accuracy, precision, recall (sensitivity), and the f1-score. These measures present a full evaluation of the model in terms of its ability to classify trees correctly and reducing false positives (finding non-trees) and false negatives (false and missing trees). ArcGIS Pro was used to verify the final tree inventory that was geolocated. The resulting CSV file was added as point feature layer with the XY Table to Point tool. The point map generated was then examined visually by overlaying it against the original UAV imagery to ensure that the spatial accuracy, and the correspondence of the detected tree locations to their real-world locations was verified. This last step will confirm the integrity of the end-to-end of the suggested structure. The 478 images were further classified into three subsets in

FIGURE 8. Image of ArcGIS software of survey area

approximate proportion 8:1:1, i.e. 382 images in training, 48 images in validation and 48 images in testing. The object detection model was developed and fine-tuned on the training and validation subsets, which consisted of 430 images. The training of the model was carried out on the basis of the YOLOv11 framework with the number of epochs set to 50. In this stage, the training data was utilized in optimization of the model parameters and the validation data was utilized to check the model performance and avoidance of overfitting. Once the training process was done, the saved weights were used in future evaluation. The Quality of performance of the model was measured quantitatively on the independent test data with evaluation metrics, such as Precision, Recall (Sensitivity), F1 Score, and Accuracy. On the test data, the model was found to have a mean accuracy in detection of 92.11%. It shows visual representations of the model output in Figure 9, which shows the instances of the tree that were correctly detected and areas where background features were correctly ignored by model at confidence threshold of 0.7.

4.2. **Georeferencing Parameter Calculation.** After the detection step, important parameters required by the georeferencing pipeline were calculated on a pair of images. The yaw angle, ($\theta$), was obtained as roughly 287.90. The conversion factor of the pixel distance to the real-world geographic distance was estimated as $1.49 \times 10^{-7}$ degrees/pixel which is known as the effective pixel width.
The framework was used to a series of all the sequential images of a particular field area. The first step of the georeferencing algorithm gave geolocated points of all the trees identified in each image. As Figure 6 shows, the survey area portion is common because of which the trees picked in the two successive images are identical. A common artifact of UAV surveys with large image overlap was the same tree showing up and being geolocated by several of the overlapping images.

The duplicate removal methodology is used to correct this duplicate counting with threshold of 5m as discussed in section 3.3(b). The argument to choose the distance

FIGURE 9. Tree detections from YOLO11 model showing successful tree identification within survey area

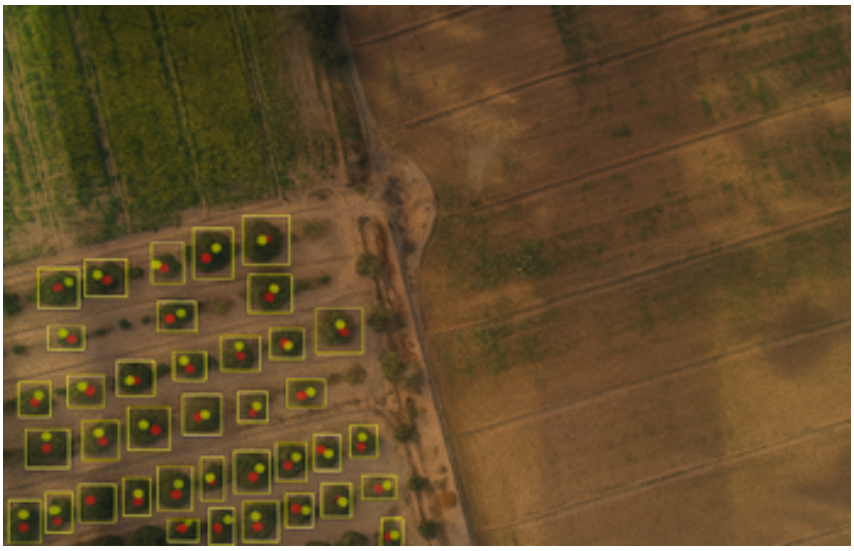

FIGURE 10. Geolocation of trees (i) Yellow points represent the actual locations of trees obtained using ArcGIS, and (ii) Red points represent the predicted locations generated by the proposed model.

threshold value of 5 meters displays the highest value of position error in georeferencing procedure. It is also proved through the comparison of actual number of trees and estimated number of trees of the proposed method. Figure 7 illustrates the resultant set of points after elimination of reduplication in geo-referenced trees which give a geometrically correct picture of the end result of the tree inventory.

As Figure 10 reveals, the locations of the predicted trees have an average positional error to the actual locations of about one meter. The ArcGIS software was used to calculate the

FIGURE 11. Tree detection results in another survey area (Dashela, Gandhinagar district, Gujarat), where yellow dots represent detections from the proposed method and red dots indicate ArcGIS derived locations.

ground-truth geolocations (yellow points), whereas the red points were used to determine the predicted locations (red points).

Table 2 gives a quantitative summary of this process of a sample of the dataset. Although the initial images had 23, 22, 29 or more detections, the duplicate removal technique applied to the whole sample data produced the final and unique count trees of 849 real trees.

The ultimate result of the suggested procedure and the respective tree inventory in the CSV files are accrued in 4th column of Table 2 listing the anticipated position of the tree which is to be used by the proposed approach. The geolocation of the trees to the ground using ArcGIS software is given as ground-truth in 3rd column of Table 2 to be compared.

TABLE 2. Tree counting and geo-referencing resultsfor the surveyed area (Chandrala , Gandhinagar district, Gujarat )

| **Image ID** | **Tree count using YOLO11** | **Actual Geo-reference of trees using ArcGIS software** | **Estimated Geo-reference of trees using proposed method** | **Estimated cumulative count** |
|---|---|---|---|---|
| 1 | 23 | (23.34517877, 72.7872980), (23.34504964, 72.7871706), (23.34507549, 72.78733089), ... | (23.34517977, 72.78729906), (23.34504864, 72.78717164), (23.34507539, 72.78733079), ... | 23 |
| 2 | 22 | (23.34487306, 72.78775734), (23.34488863, 72.7877036), (23.34490694, 72.78762995), ... | (23.34487206, 72.78775634), (23.34488763, 72.7877136), (23.34490684, 72.78762985), ... | 45 |
| 3 | 29 | (23.34455446, 72.78800066), (23.34458454, 72.78800182), (23.34465457, 72.7880461), ... | (23.34455456, 72.78800076), (23.34458354, 72.78800282), (23.34465467, 72.7880471), ... | 74 |
| | | ... | | |

**Mean error:** $1.25 \pm 0.35$ m
**Estimated trees in surveyed area (proposed method):** 776
**Actual tree count (ground truth):** 849

TABLE 3. Tree counting and geo-referencing results for the surveyed area (Dashela, Gandhinagar district, Gujarat)

| Image ID | Tree Count (YOLO_11) | Actual Geo-referenced Coordinates (ArcGIS) | Estimated Geo-referenced Coordinates (Proposed Method) | Cumulative Count |
|---|---|---|---|---|
| 1 | 20 | (23.28437877, 72.7523990), (23.28406914, 72.7521566), (23.28403549, 72.75233159), ... | (23.28437967, 72.75239906), (23.28406864, 72.75217574), (23.28403529, 72.75233129), ... | 20 |
| 2 | 21 | (23.34485316, 72.75275614), (23.34484863, 72.7527036), (23.34493694, 72.75262895), ... | (23.28485206, 72.75275634), (23.28484753, 72.7527026), (23.28493664, 72.75262875), ... | 41 |
| 3 | 23 | (23.28435426, 72.75202436), (23.28438454, 72.75203182), (23.28453457, 72.7521151), ... | (23.28435456, 72.75202476), (23.28438344, 72.75203172), (23.28455437, 72.7521171), ... | 64 |
| | | ⋮ | | |

**Mean Error:** $1.25 \pm 0.40$ m

**Estimated Trees in Surveyed Area (Proposed Method):** 121

**Actual Tree Count (Ground Truth):** 128

The same experiments were conducted on another survey area, Dashela village in Gandhinagar district, Gujarat as shown in Figure 11, which is characterized by a dense forest region. This area contrasts with the previously surveyed Chandrala village, also in Gandhinagar district, Gujarat, which consists of a moderately dense, man-made tree environment. Table 3 presents a summary of geo-referenced tree detections along with the cumulative tree count estimated for Dashela village. The results obtained using ArcGIS software are compared with those generated by the proposed method, demonstrating the effectiveness and consistency of the proposed approach across varying geographic conditions.

5. **Conclusion.** This study proposes a UAV-based end-to-end framework for automated tree detection, counting, and geo-referencing using high-resolution RGB imagery. The framework integrates YOLOv11 with a novel metric-consistent geo-referencing pipeline that directly transforms image-space detections into accurate geographic coordinates, eliminating the need for orthomosaics or LiDAR data. Experimental results from a real-world semi-urban site demonstrate a mean detection accuracy of 92.11% and a positional error of $1.25 \pm 0.35$ m. A spatial deduplication strategy effectively eliminates redundant detections caused by image overlap, yielding a reliable tree inventory. The proposed approach offers a cost-effective, scalable, and computationally efficient alternative to conventional forest inventory techniques. These results highlight the framework's potential for large-scale environmental monitoring, urban forestry management, and sustainable planning applications.

Limitations: It is the reliance on manually annotated training data, which is time-consuming and may introduce human bias. The geo-referencing approach depends on

approximate pixel-to-ground transformations and UAV GPS data, which can lead to positional errors (≈1–1.5 m). The fixed threshold (5m) for duplicate removal may not generalize well in very dense or sparse vegetation.

## REFERENCES


[1] Y. Pan, R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, et al. A large and persistent carbon sink in the world's forests. *Science*, 333(6045):988–993, 2011.

[2] J. A. Foley, R. DeFries, G. P. Asner, C. Barford, and G. Bonan. Global consequences of land use. *Science*, 309(5734):570–574, 2005.

[3] G. B. Bonan. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science*, 320(5882):1444–1449, 2008.

[4] Zhenbang Zhang, Chongyang Han, Xinrong Wang, Haoxin Li, Jie Li, Jinbin Zeng, Si Sun, and Weibin Wu. Large field-of-view pine wilt disease tree detection based on improved yolo v4 model with uav images. *Frontiers in Plant Science*, 15:1381367, 2024.

[5] D. J. Nowak and J. F. Dwyer. Understanding the benefits and costs of urban forest ecosystems. In J. E. Kuser, editor, *Urban and Community Forestry in the Northeast*, pages 25–46. Springer, 2007.

[6] Qian Wang, Zhi Pu, Lei Luo, Lei Wang, and Jian Gao. A study on tree species recognition in uav remote sensing imagery based on an improved yolov11 model. *Applied Sciences*, 15(16):8779, 2025.

[7] Yunmei Huang, Botong Ou, Kexin Meng, Baijian Yang, Joshua Carpenter, Jinha Jung, and Songlin Fei. Tree species classification from uav canopy images with deep learning models. *Remote Sensing*, 16(20):3836, 2024.

[8] F. J. Escobedo, T. Kroeger, and J. E. Wagner. Urban forests and pollution mitigation: analyzing ecosystem services and disservices. *Environmental Pollution*, 159(8–9):2078–2087, 2011.

[9] M. Köhl, S. S. Magnussen, and M. Marchetti. *Sampling Methods, Remote Sensing and GIS Multi-Resource Forest Inventory*. Springer, 2006.

[10] J. Smith. A comparative analysis of forest inventory field methods. *Journal of Forestry*, 108(5):230–235, 2010.

[11] J. C. White, N. C. Coops, M. A. Wulder, M. Vastaranta, T. Hilker, and P. Tompalski. Remote sensing technologies for enhancing forest inventories: a review. *Canadian Journal of Remote Sensing*, 42(5):619–641, 2016.

[12] Baohua Chang, Fei Li, Yuncai Hu, Hang Yin, Zhenhua Feng, and Liang Zhao. Application of uav remote sensing for vegetation identification: a review and meta-analysis. *Frontiers in Plant Science*, 16:1452053, 2025.

[13] R. E. McRoberts and E. O. Tomppo. Remote sensing support for national forest inventories. *Remote Sensing of Environment*, 110(4):412–419, 2007.

[14] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, et al. High-resolution global maps of 21st century forest cover change. *Science*, 342(6160):850–853, 2013.

[15] M. Garcia and L. Martinez. A review on the use of very high resolution satellite imagery for individual tree crown detection. *Remote Sensing Reviews*, 34(1):1–25, 2015.

[16] J. Hyyppä, H. Hyyppä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29(5):1339–1366, 2008.

[17] Weijie Kuang, Hann Woei Ho, Ye Zhou, Shahrel Azmin Suandi, and Farzad Ismail. A comprehensive review on tree detection methods using point cloud and aerial imagery from unmanned aerial vehicles. *Computers and Electronics in Agriculture*, 227:109476, 2024.

[18] L. Wallace, A. Lucieer, Z. Malenovský, D. Turner, and P. Vopěnka. Assessment of forest structure using two uav techniques: a comparison of airborne laser scanning and structure from motion point clouds. *Forests*, 7(3):62, 2016.

[19] C. Torresan, A. Berton, F. Carotenuto, S. F. Di Gennaro, B. Gioli, and A. Matese. Forestry applications of uavs in europe: a review. *International Journal of Remote Sensing*, 38(8–10):2427–2447, 2017.

[20] B. T. Fraser and R. G. Congalton. Issues in unmanned aerial systems (uas) data collection of complex forest environments. *Remote Sensing*, 10(6):908, 2018.

[21] P. J. Zarco-Tejada, R. Diaz-Varela, R. Angulo-Jaramillo, and J. A. J. Berni. Tree height quantification using very high resolution imagery acquired from a uav and automatic 3d photo-reconstruction methods. *European Journal of Agronomy*, 55:89–99, 2014.

[22] Tien-Wen Sung, Zheng-Jiang Xiao, Qingjun Fang, You-Te Lu, and Thi-Minh-Phuong Ha. Super-resolution of uav images based on improved generative adversarial networks. *Journal of Network Intelligence*, 10(4), November 2025. Received April 23, 2024; revised December 9, 2024; accepted August 28, 2025.

[23] X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, et al. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017.

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: unified, real-time object detection. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.

[25] E. Sengun, S. Aksoy, E. Sertel, and J. E. Fransson. Comparative analysis of yolov8 and yolov11 on tree detection using uav rgb and laser scanning data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X:173–179, 2025.

[26] R. Yang, D. Yuan, M. Zhao, Z. Zhao, L. Zhang, Y. Fan, G. Liang, and Y. Zhou. Camellia oleifera tree detection and counting based on uav rgb image and yolov8. *Agriculture*, 14(10):1789, 2024.

[27] Hend A. Elsayed, Mustafa F. Hameed, and Mohammed M. El Sherbiny. Image classification using decision tree classifier and features extraction using hough transform and genetic algorithm. *Journal of Information Hiding and Multimedia Signal Processing*, 16(1):1–X, March 2025. ISSN 2073-4212.

[28] T. Rohe, B. Böhm, M. Kölle, J. Stein, R. Müller, and C. Linnhoff-Popien. Coconut palm tree counting on drone images with deep object detection and synthetic training data. *arXiv preprint arXiv:2412.11949*, 2024.

[29] Ş. N. Topgül, E. Sertel, S. Aksoy, C. Ünsalan, and J. E. Fransson. Vhrtrees: a new benchmark dataset for tree detection in satellite imagery and performance evaluation with yolo-based models. *Frontiers in Forests and Global Change*, 7:1495544, 2025.

[30] R. Abreu-Dias, J. M. Santos-Gago, F. Martín-Rodríguez, and L. M. Álvarez Sabucedo. Advances in the automated identification of individual tree species: A systematic review of drone- and ai-based methods in forest environments. *Technologies*, 13(5):187, 2025.