Turkish Journal of Remote Sensing, 2025; 7(2); 184-199 Research Article





Deep learning based citrus tree detection from low resolution satellite images: A case study of Tarsus

Semih Kahveci*[1], Mehmet Özgür Çelik[2], Ramazan Akkurt[3], Özmen Kahveci[4]

[1] *Mersin University, Engineering Faculty, Department of Computer Engineering, Mersin, Türkiye;* semihkahveci@mersin.edu.tr
[2] *Mersin University, Engineering Faculty, Department of Geomatics Engineering, Mersin, Türkiye;* mozgurcelik33@gmail.com
[3] *Mersin University, Engineering Faculty, Department of Engineering, Mersin, Türkiye;* rakkurt@mersin.edu.tr
[4] *GIBTU, Institute of Graduate Studies, Department of Computer Engineering, Gaziantep, Türkiye; ozmenkahveci.ce*@hotmail.com

Check for updates

Article History:
Received: 3 April 2025
Revised: 11 April 2025
Accepted: 25 May 2025
Published: 30 December 2025

Copyright: © 2025 by the authors.
This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY-SA) license. https://creativecommons.org/licenses/by-sa/4.0/

Abstract: The escalating global population, industrialization, and climate change are increasing pressure on agricultural lands. In this context, sustainable agricultural land management is critically important, particularly for high-value crops such as citrus, which plays critical role in economic and food security. Accurate detection and enumeration of citrus trees are essential for ensuring the sustainability and effective monitoring of citrus cultivation. This study employs deep learning methods for object detection of citrus trees in the Tarsus district of Mersin, comparing the performance of Mask R-CNN, YOLOv8, and YOLO11 models using low-resolution satellite imagery. Additionally, the impact of super-resolution (SR) techniques on model accuracy is examined. Results demonstrate that integrating SR techniques significantly improves object detection accuracy, with the YOLO11 model achieving the highest performance. This research underscores the potential of deep learning-based approaches to enhance citrus tree monitoring, yield estimation, and agricultural management practices, offering actionable insights for sustainable agriculture.

Keywords: Deep learning; satellite image; citrus tree; object-detection

Citation: Kahveci, S., Çelik, M. Ö., Akkurt, R., & Kahveci, Ö. (2025). Deep learning based citrus tree detection from low resolution satellite images: A case study of Tarsus. *Turk. J. Remote Sens.*, 7(2), 184-199. https://doi.org/10.51489/tuzal.1669616

1. Introduction

As the global population grows and industrialization progresses, the increasing demands on land become more pronounced each day. This issue is exacerbated by disasters such as droughts, floods, and wildfires, which are primarily driven by climate change and represent some of the most pressing challenges of our time, posing significant threats to land resources (Uyar, 2024). Agricultural areas, essential for ensuring food security and sustaining life, are particularly vulnerable to these threats. Additionally, these lands often suffer from the adverse effects of development initiatives undertaken by countries that prioritize projects in construction, industry, technology, and other sectors. Poor planning and unregulated land use are key factors in the decline of agricultural land. Globally, agricultural lands decreased from 4.872 billion hectares (ha) in 2000 to 4.781 billion ha by 2022, reflecting a decline of approximately 1.87% (FAO, 2025a). Türkiye has experienced a similar trend, with agricultural land diminishing from 26.38 million ha in 2000 to 23.88 million ha by 2022, representing a decline of 9.48% (TÜİK, 2025). This reduction in agricultural land inevitably leads to decreased yields of agricultural products, jeopardizing the sustainability of agriculture as a whole.

The protection and management of existing agricultural areas are of utmost importance. From this perspective, citrus trees, which have significantly proliferated in the Mediterranean basin and hold substantial economic value, deserve particular attention (Orhan, 2021). As of 2023, global citrus production totaled 169,388,809.85 tons, comprising oranges (69.85 million tons), tangerines (52.56 million tons), lemons (23.64 million tons),

grapefruits (9.93 million tons), and bitter oranges (13.41 million tons) (FAO, 2025b). This agricultural activity utilized 10,554,846 ha of land, with China, Mexico, India, and Brazil being the leading producers. Türkiye also contributes significantly to citrus production, ranking third in mandarin harvest at 2,952,775 tons over 69,261.3 ha, following India. Additionally, Türkiye ranks fourth globally in lemon production, with 2,325,726 tons harvested (56,438.9 ha). The country stands seventh for grapefruit production, yielding 284,565 tons (4,945 ha), and is eighth for orange production with a total of 2,311,335 tons (49,481.7 ha) (FAO, 2025b; TÜİK, 2025).

The number of trees in cultivated areas must be assessed before citrus can be analyzed from a sustainable agriculture standpoint. This assessment allows for reliable estimations of the yield and efficiency of this product while also facilitating the monitoring of plant health. Numerous studies in the literature have employed various methods to estimate tree counts (Table 1). Zortea et al. (2018) explored the detection of citrus trees in Brazil using unmanned aerial vehicle (UAV) imagery and a convolutional neural network (CNN) algorithm. Similarly, Ozdarici-Ok (2015) utilized a fast radial symmetry (FRS) transformation method with high-resolution satellite images to automatically detect citrus trees, selecting Mersin as the study area. In a different approach, Li et al. (2016) employed the local maximum algorithm with LiDAR and UAV data to detect trees in the Huailai-Yanqing basin near Beijing. El Akrouchi et al. (2025) also applied the mask region-based CNN (Mask R-CNN) technique to determine the number of citrus fruits in Marrakesh, Morocco, using close-up images for their analysis.

Table 1. Research on tree detection in the literature

Reference	Method	Data	Tree	Study area
Donmez et al. (2021)	CCL	UAV	Citrus	Adana/Çukurova U.
Cheang et al. (2017)	CNN	Satellite	Palm	World
Fan et al. (2018)	CNN	UAV	Tobacco	Chongqing/China
Csillik et al. (2018)	CNN	UAV	Citrus	Tulare/USA
Liao et al. (2025)	YOLO-MECD	Camera	Citrus	Nanchang/China
Al Riza et al. (2025)	YOLOv7	Camera	Citrus	Batu/ Indonesia
Toosi et al. (2022)	Time series	Satellite, Google Earth	Citrus	Juybar/Iran
Tian et al. (2022)	YOLOv5	UAV	Citrus	Sihui/China
Yu et al. (2023)	Time series	Satellite	Citrus	Xunwu/China
Yang et al. (2022)	Mask R-CNN	Google Earth	Tree	Central Park/USA

As the literature states, tree detection is important, especially in regions where citrus cultivation occurs. Previous deep learning (DL) methods for citrus tree detection typically rely on high-resolution images from UAV or satellite data, focusing on relatively small areas (Csillik et al., 2018; Hunt et al., 2018; Kouvaras, & Petropoulos, (2024). These approaches often utilize only one or two models without thorough benchmarking and seldom explore the advantages of image enhancement (Yu et al., 2022). In contrast, this study leverages readily available low-resolution satellite images from Google Earth and employs the Real-ESRGAN super-resolution algorithm to improve visual detail.

This research was conducted in the Tarsus district of Mersin, an area characterized by a dominant Mediterranean climate and intensive citrus farming. In 2023, Mersin province produced 1,661,600 tons of citrus fruit, with Tarsus contributing 39.23%, or 651,793 tons (TÜİK, 2025). This highlights Tarsus's significant role in the cultivation of these species. Deep learning (DL) methods, such as Mask R-CNN, You Only Look Once (YOLOv8), and YOLO11, were employed to estimate the number of citrus trees in the selected study area. This study was conducted due to the insufficient detailed research concerning citrus trees in Tarsus.

The research notably highlights its unique value by offering a time- and cost-efficient approach that utilizes deep learning methods to accurately detect citrus trees within the study area and map citrus orchards across extensive regions with limited samples. Furthermore, it contributes to the existing literature by effectively employing various models for object detection.

Throughout the study, several crucial steps were taken to drive our research forward:

- Study area selection,
- Data acquisition and preparation,
- Image processing,
- Image augmentation,
- Implementation of deep learning models
- Performance evaluation,
- Comparative model analysis,
- Visual comparison of detection results. Furthermore, a flowchart was presented in "2.2. data collection" chapter.

2. Materials and Methods

2.1. Study area

Tarsus, which was chosen as the study area, is located in the Mediterranean region between 36°43'-37°23' North latitude and 34°33'-35°9' East longitude. Tarsus is the most populous district of Mersin, with a population of 356,937 (TÜİK, 2025). In addition, it is the third largest district of the province with a surface area of 1993.21 km^2. While the altitude in the city center of Tarsus varies between 10-65 m, this value reaches up to 3500 m in the Taurus Mountains located in the north.

In the study area where the effects of the Mediterranean climate are seen, the annual average temperature is around 20°C (MGM, 2025). The average temperature rises to 28°C in the summer and around 11°C in winter (Weather Spark, 2025). The total annual average precipitation is 468.7 mm. The average precipitation is 6.9 mm in summer and 77 mm in winter (Weather Spark, 2025). However, predominantly alluvial, red-brown mediterranean, and brown forest soils are seen.

Tarsus is characterized by its geographical location, soil composition, and climate, sustaining active agricultural practices (Figure 1). Approximately 50.38% of the total area (1,004.22 km^2) is dedicated to agriculture, with 142.55 km^2 (14.20%) specifically allocated for citrus production. As a result of these activities, it is known that Tarsus provides 39.23% of citrus production in Mersin. Given these factors, the study was conducted in Tarsus, where three areas of citrus cultivation were selected for analysis.

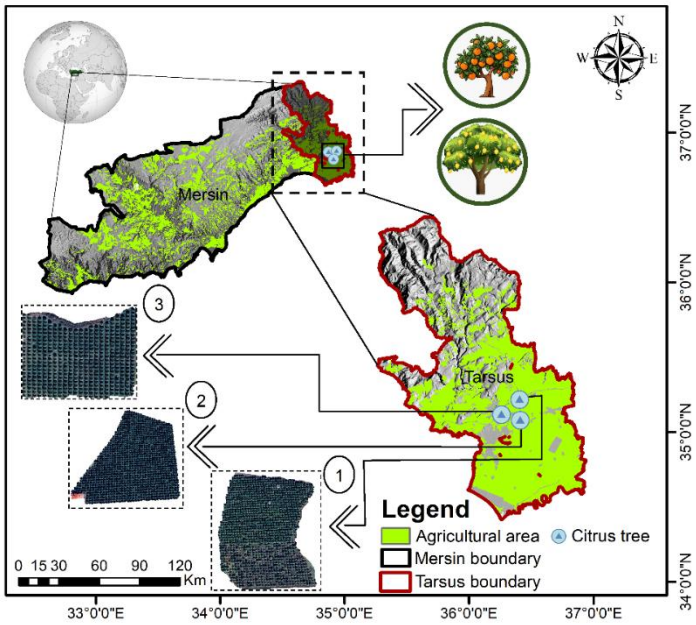


Figure 1. Study area

2.2. Data collection

In this study, we obtained high-resolution RGB satellite images of the study area, measuring 8192×4540 pixels, from various altitudes using Google Earth to facilitate object detection and segmentation processes for citrus trees (Figure 2). The analysis focused exclusively on regions containing citrus orchards, which were designated as the study area. To enhance the training of object detection algorithms and ensure greater accuracy in the labeling process, these images were subdivided into smaller segments of 640×640 pixels, yielding a total of 150 images. The segmented images were then uploaded to the Roboflow platform for labeling, where the polygon-based labeling method was chosen to improve detection performance by clearly defining the boundaries of the trees (Akkurt et al., 2024).

In order to increase the performance of the model trained with labeled data and to strengthen its generalization ability, data augmentation techniques are used to increase the efficiency of the training process by generating new examples from existing data. Among the common data augmentation techniques, the number of images was increased to 260 by using geometric transformations (rotation, scaling), color space adjustments (brightness, contrast), adding noise, and mixing methods. These techniques increase the diversity of the data, making it possible for models to be more reliable and robust in real-world scenarios (Shorten & Khoshgoftaar, 2019). The obtained data set was divided into three subsets as 70% training, 20% validation, and 10% test in order to evaluate the overall performance of the model.

In object detection and segmentation applications, denoising images and making objects more distinct is a critical step that directly affects the learning capacity of the model (Wang et al., 2022). Therefore, a deep learning-based super-resolution (SR) algorithm was used to improve the learning process of the model by increasing the quality and clarity of the images. SR is a sub-heading of the image enhancement field and aims to obtain high-resolution representations from low-resolution images. This process not only scales the image to a higher resolution, but also reduces noise, making the details of the objects more distinct (Nasrollahi & Moeslund, 2014). In this context, SR technologies are considered as an important pre-processing step to increase the success of applications that have gained popularity in recent years, such as object detection and segmentation.

In this study, SR is applied as a preprocessing step and Real-ESRGAN (Wang et al., 2021) method is used for SR. Real-ESRGAN is a Generative Adversarial Network (GAN) based approach developed for the problem of "blind super-resolution" in real-world images and trained with synthetic data. Instead of estimating complex degradations (noise, blur, etc.) in low-quality or corrupted images, this method produces clearer and more distinct images for object detection and segmentation models by reconstructing high-resolution details through pure synthetic data. Real-ESRGAN, which is preferred due to its superior generalization ability especially in practical applications, is applied in this study as a critical preprocessing step to increase the accuracy of the model developed for the detection and counting of citrus trees (Wang et al., 2021). Figure 3 visually compares the raw image and the image enhanced with Real-ESRGAN.

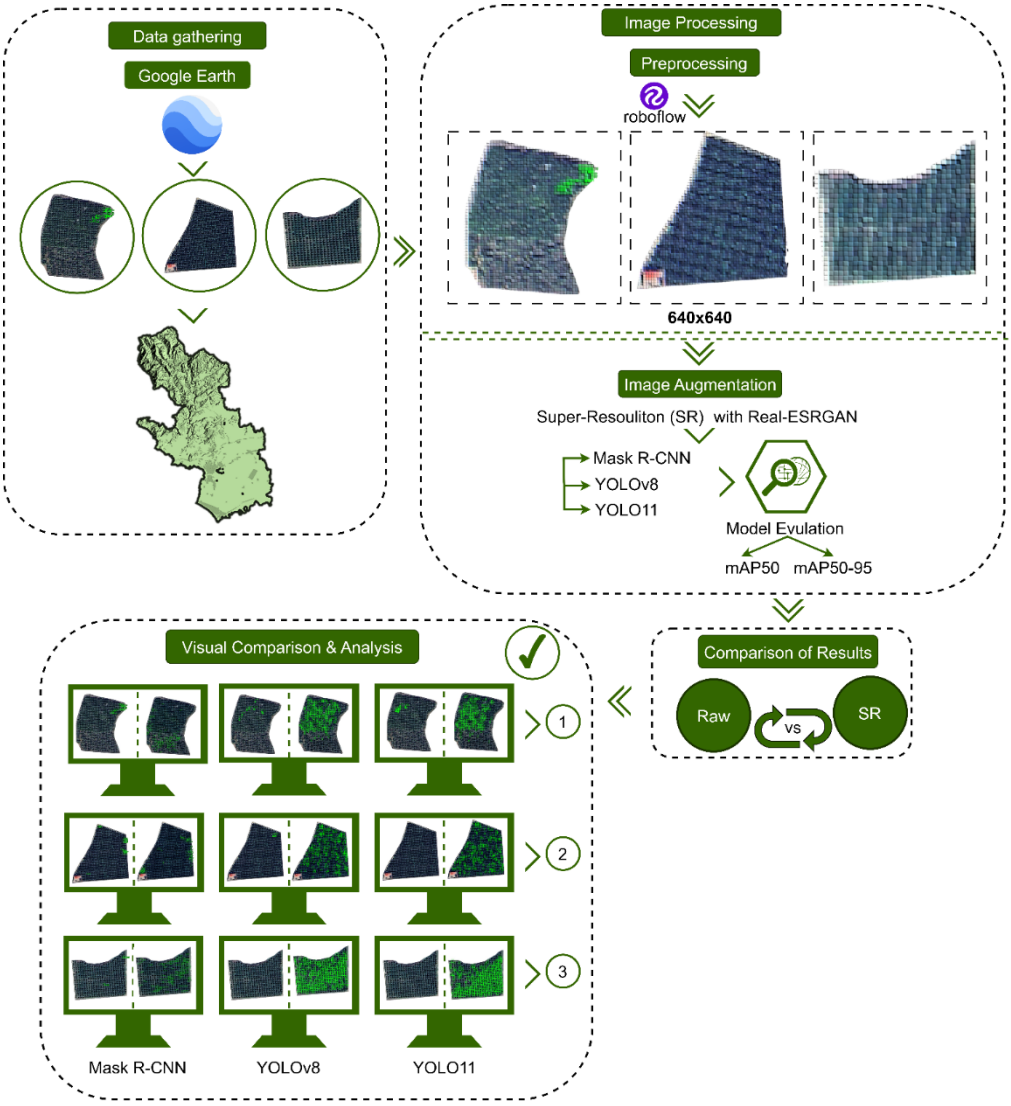
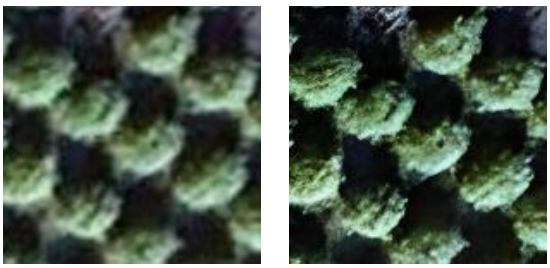


Figure 2. Workflow of the methodology



(a) (b)

Figure 3. (a) Raw images belong to study area, (b) Enhanced image with SR belong to study area

2.3. Object detection models

Object detection algorithms, which play an important role in terms of productivity and sustainability in agricultural areas, are widely used in applications such as monitoring and counting of fruit trees (Sapkota et al., 2024). In this study, deep learning-based Mask R-CNN, YOLOv8 and YOLO11 three different object detection algorithms were used for the detection and counting of citrus trees. In this section, the architectural structures and object detection processes of these three models are discussed in detail.

2.3.1. Mask R-CNN

Mask R-CNN is a deep learning-based model developed by Facebook AI Research in 2017 that can solve object detection and instance segmentation problems with high accuracy.

This model is an improved version of the Faster R-CNN architecture, which has the ability to segment objects at the pixel level along with the detection of objects with the mask estimation branch method. Mask R-CNN has a Backbone that provides meaningful feature extraction from images and uses architectures such as ResNet-50 or ResNet-101, a Region Proposal Network (RPN) that determines potential regions of objects in images from these extracted features, and a head structure with a bounding box branch, one of which estimates the class label and bounding box coordinates for each proposed region, and the other one, which produces a pixel-based segmentation mask for each object (He et al., 2017).

2.3.2. YOLOv8

Object detection, a fundamental computer vision task, involves identifying and localizing objects within images or videos by classifying and enclosing them in bounding boxes. The YOLO series has gained prominence for real-time performance, achieved by unifying object localization and classification into a single neural network forward pass. YOLOv8 developed by Ultralytics in 2023, represents a milestone in object detection research (Ultralytics, 2023). Building upon the advancements of predecessors, YOLOv8 introduces optimized architectural designs, training methodologies, and computational efficiencies. The model integrates a hybrid architecture that combines CNNs for local feature extraction with Transformer-based components to capture global contextual dependencies, enhancing robustness across diverse environmental conditions. Its superior generalization capabilities stem from training on large-scale, diverse datasets, enabling reliable performance under varying lighting, scales, and occlusion scenarios. Key advantages of YOLOv8 include its modular architecture, computational efficiency (low latency and high frames-per-second processing), and seamless integration into real-world applications via Ultralytics' open-source ecosystem and pre-trained models. The versatile structure of YOLOv8 covers areas such as precision agriculture, autonomous systems, surveillance, medical diagnostics and retail analytics, providing great advantages for use in academia and industry.

2.3.3. YOLO11

YOLO11, introduced in September 2024, represents a groundbreaking advancement in object detection and establishes a new benchmark in computer vision. YOLO11 builds on the real-time performance and accuracy of predecessor like YOLOv8, while achieving notable breakthroughs in speed, precision, and adaptability through cutting-edge deep learning innovations and architectural optimizations (Khanam & Hussain, 2024). The model employs an enhanced hybrid architecture, synergizing CNNs for local feature extraction with transformer-based components to model global contextual dependencies. This dual approach ensures robust detection of fine-grained details and contextual coherence, even in complex scenarios such as densely populated object environments or scenes with dynamic lighting variations. Furthermore, training on large-scale, diverse datasets has strengthened generalization capabilities, mitigated overfitting risks and ensuring reliability across heterogeneous conditions.

Key strengths of YOLO11 include:

- Modular architecture for seamless adaptation to diverse hardware and application-specific requirements
- Computational efficiency, achieving high frames-per-second (FPS) with minimal resource consumption
- Ease of deployment in industrial and academic domains, including healthcare, precision agriculture, smart cities, robotics, and industrial automation.

These attributes establish YOLO11 as a cutting-edge solution for object detection, seamlessly connecting academic research with real-world applications.

3. Results

3.1. Evaluation metrics

In order to comprehensively evaluate the performance of Mask R-CNN, YOLOv8, and YOLO11 in object detection tasks, standardized evaluation metrics widely adopted in computer vision were employed. These include Precision (P), Recall (R), mAP50, and mAP50-95. These metrics collectively enable rigorous comparison of model performance, balancing speed, precision, and robustness for real-world applications such as citrus tree monitoring. Precision quantifies the proportion of true positive detections (correctly identified objects) relative to all instances classified as positive by the model. It reflects the algorithm's ability to minimize false positives (FP). For instance, in citrus tree detection, high precision ensures that non-target objects (e.g., other tree species or background features) are rarely misclassified as citrus trees. These metrics are critical for minimizing false alarms (e.g., misclassifying non-target objects in agriculture) (Zhang et al., 2016). High Recall ensures no trees are missed, which is vital for accurate yield estimation (Weinstein et al., 2019). Recall can be calculated using the formula in Equation 1.

$$\text{Recall} = \text{TP} / \text{TP} + \text{FN} \tag{1}$$

Recall measures the fraction of true positives correctly identified out of all actual positives in the dataset, emphasizing the model's capacity to reduce false negatives (FN). In an agricultural context, high recall is critical to ensure that no citrus trees are missed during detection or monitoring. It can be calculated using Equation 2).

$$\text{Precision} = \text{TP} / \text{TP} + \text{FP} \tag{2}$$

mAP50 (mean Average Precision at 50% Intersection-over-Union) provide comprehensive assessments of detection accuracy. While mAP50 prioritizes localization tolerance, mAP50-95 enforces stricter spatial alignment, offering a robust benchmark for model reliability under varying object scales and occlusion conditions. In order to derivate mAP value, it needs to calculate average precision. Average precision (AP) is the area under the precision and recall curve. The mAP metric is robust for handling class imbalance and localization accuracy (Everingham et al., 2010). It can be calculated using Equation 3.

$$AP = \int_0^1 p(r)dr \tag{3}$$

mAP50-95 (Average Precision in the IoU range of 0.50–0.95) is the average of the AP values calculated from the IoU threshold of 0.50 to 0.95 (in 0.05 increments). This metric is especially critical in tasks that require precise localization, such as robotic grasping. mAP50-95 provides a stricter benchmark for precise localization tasks compared to mAP50. The mAP50-95 metric can be excessively stringent for small objects, such as trees in aerial imagery, potentially resulting in an underestimation of model performance. (Padilla et al., 2021). mAP can be calculated using Equation 4.

$$mAP = \frac{1}{k}\sum_i^k AP_i \tag{4}$$

Intersection over Union (IoU) is a fundamental metric for evaluating the accuracy of object detection models by measuring the spatial overlap between predicted and ground-truth bounding boxes. IoU provides a straightforward and interpretable measure of localization accuracy, making it widely adopted in object detection tasks (Rezatofighi et al., 2019). Despite its widespread use, IoU has notable drawbacks. It can be overly sensitive to small objects or irregular shapes, leading to unstable evaluations when objects are tiny or highly asymmetric (Zheng et al., 2021). It is calculated as the ratio of the area of intersection to the area of union of the two bounding boxes using Equation 5.

$$IoU = \frac{Intersection\ Area}{Union\ Area} \quad (5)$$

3.2. Experimental results of models

In this section, the performance of different object detection models such as Mask R-CNN, YOLOv8, and YOLO11 for the detection of citrus trees is investigated. In the scope of the study, the effect of the SR algorithm on the detection performance of the models is evaluated in detail. Model performances are compared with metrics such as mAP50 and mAP50-95 using the SR applied dataset and the raw dataset.

3.2.1. Mask R-CNN

The detection performance of the Mask R-CNN object detection model on the SR applied dataset (SR-Mask R-CNN) and the raw (Mask R-CNN) dataset is compared and presented in Figure 4.

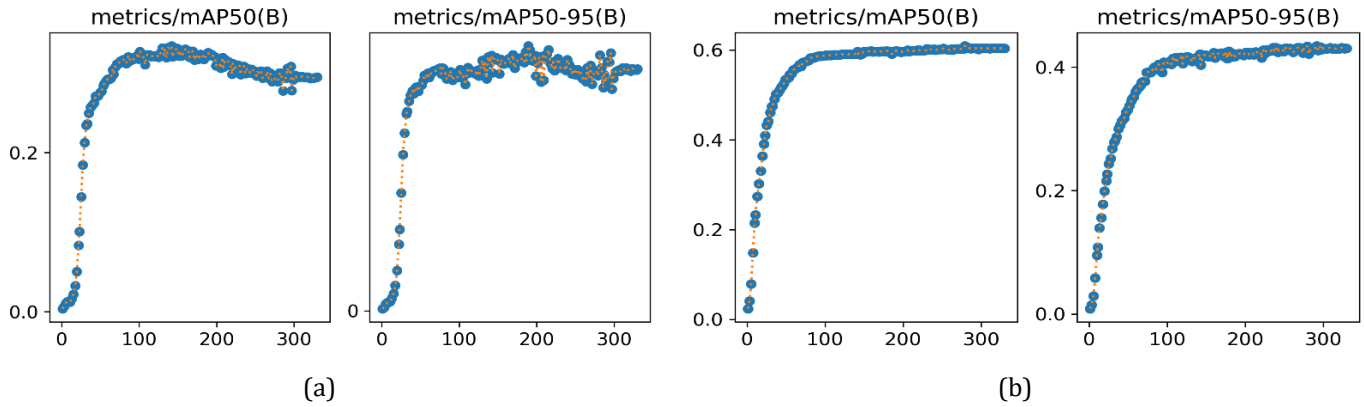


Figure 4. Training graphs for iteration-based mAP50 and mAP50-95 values of the Mask R-CNN model trained on the raw and SR applied dataset

Performance analysis revealed that integrating SR preprocessing significantly enhanced Mask R-CNN's object detection capabilities. The SR-augmented model achieved a mean Average Precision (mAP50-95) of %40, surpassing the baseline model's %13.5, while also outperforming it in mAP50 (%60 vs. %30). This improvement shows SR's efficiency in refining localization accuracy, particularly under stringent Intersection-over-Union (IoU) thresholds. Additionally, training with SR-enhanced data accelerated convergence, enabling faster attainment of high Average Precision (AP) values and earlier stabilization of loss curves. These results emphasis the advantages of SR by enhancing detection precision through resolution-related improvements and optimizing the training process, making it a practical solution for improving object detection in resource-constrained applications such as agricultural monitoring.

3.2.2. YOLOv8

The training results for the Yolov8 model on the raw dataset and SR-dataset are presented in Figure 5.

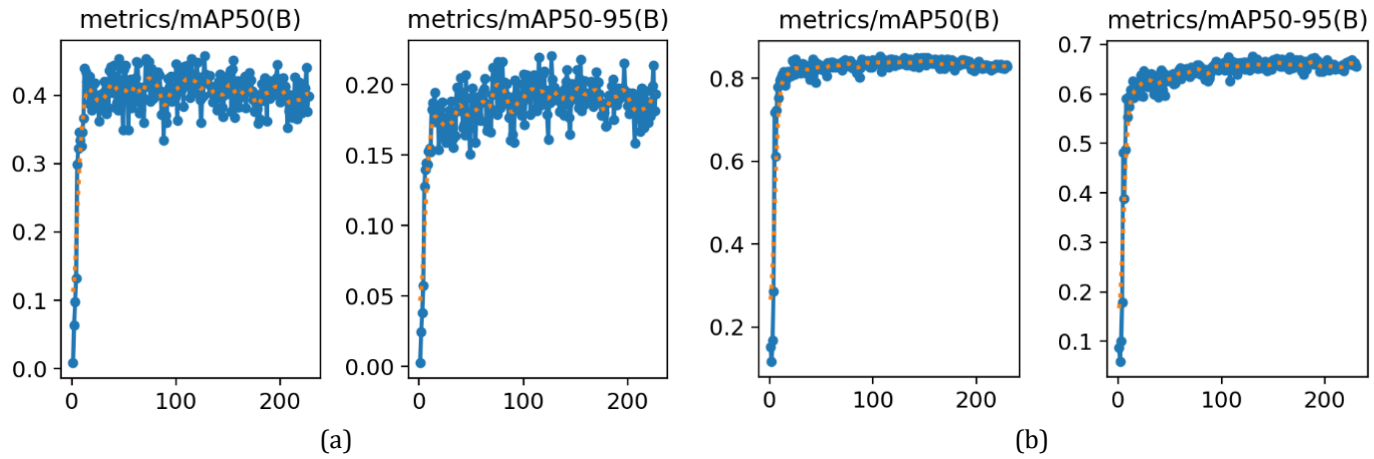


Figure 5. Training graphs for iteration-based mAP50 and mAP50-95 values of the YOLOv8 model trained on the raw and SR applied dataset

Analysis of the training process graphs revealed that the SR-YOLOv8 model, trained on a super-resolution (SR)-enhanced dataset, demonstrated superior performance compared to the baseline YOLOv8 model trained on raw data. Evaluated using mAP50 and mAP50-95 metrics, SR-YOLOv8 achieved higher accuracy levels and faster convergence during training, indicating enhanced robustness in object localization across varying IoU thresholds. Conversely, the baseline YOLOv8 model exhibited lower mAP values and marked fluctuations in performance, reflecting challenges in generalization, particularly for low-resolution or detail-degraded imagery. These findings highlight that super-resolution preprocessing not only improves detection accuracy for visually complex inputs but also stabilizes and accelerates the training process, reinforcing its value in optimizing YOLOv8 for real-world applications with limited image quality.

3.2.3. YOLO11

The training results for the Yolo11 model on the raw dataset and SR-dataset are presented in Figure 6.

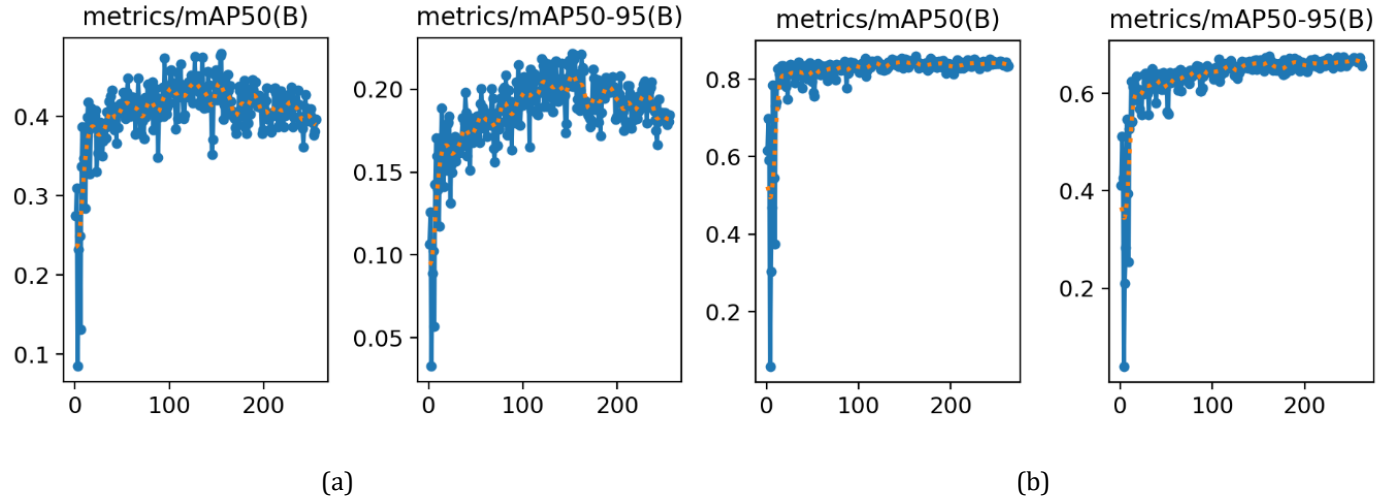


Figure 6. Training graphs for iteration-based mAP50 and mAP50-95 values of the YOLO11 model trained on the raw and SR applied dataset

Analysis of the training graphs demonstrated that the SR-YOLO11 model, trained on a SR-enhanced dataset, achieved markedly superior performance compared to the baseline YOLO11 model trained on raw data. Evaluated using mAP50 and mAP50-95 metrics, SR-YOLO11 stabilized at 80% (mAP50) and 65% (mAP50-95), reflecting robust object localization across IoU thresholds, while the baseline YOLO11 model remains at 40%

(mAP50) and 20% (mAP50-95). Furthermore, SR-YOLO11 exhibited accelerated convergence and a smoother learning curve, whereas YOLO11 displayed marked fluctuations and training instability. These outcomes highlight the effectiveness of super-resolution preprocessing in improving detection accuracy for low-resolution or degraded images while enhancing model stability and reducing computational inefficiencies during training.

3.3. Comparative analysis of models

The object detection performances of the object detection models used in this study on super-resolution applied (SR-Dataset) and non-super-resolution applied (Raw Dataset) datasets are compared in terms of mAP50 and mAP50:95 metrics in Table 2.

Table 2. Comparison of object detection models on SR applied dataset and raw dataset

	Raw Dataset				Applied SR Dataset			
	mAP50	**mAP50-95**	**Precision**	**Recall**	**mAP50**	**mAP50-95**	**Precision**	**Recall**
Mask R-CNN	29.46	13.28	-	-	60.37	43.07	-	-
YOLO v8	45.73	22.05	0.42	0.65	84.90	67.59	0.81	0.71
YOLO v11	45.39	22.15	0.433	0.58	85.93	67.66	0.76	0.79

The application of super resolution techniques yielded substantial improvements in object detection performance across all evaluated models. For the Mask R-CNN model, mAP50 increased from 29.46% to 60.37%, and mAP50-95 increased from 13.29% to 43.08% when trained on SR-dataset. Similarly, the YOLOv8 model exhibited dramatic gains, with mAP50 increased from 45.7% to 84.9% and mAP50-95 improving from 22.1% to 67.6%. The YOLO11 model achieved the highest performance on the SR-dataset, reaching 85.93% (mAP50) and 67.66% (mAP50-95). These results demonstrate that SR preprocessing significantly enhances detection accuracy for low-resolution or noisy datasets, irrespective of the model architecture. Notably, YOLO11's superior performance on SR-enhanced data highlights the potential of integrating advanced SR techniques with state-of-the-art object detection frameworks. This analysis underscores that SR optimizes model learning by preserving critical visual details and mitigating noise, thereby enabling robust object detection in challenging real-world scenarios.

3.4. Comparative visualization of test

In this section, visual comparisons of object detection algorithms are presented. The performance of Mask R-CNN, YOLOv8 and YOLO11 models in identifying citrus trees is evaluated in two different data qualities (raw and super-resolution applied) and is presented in Figures 7, 8, and 9. In the figures, the results of Mask R-CNN are given in the left column, YOLOv8 in the middle column and YOLO11 in the right column. The first row shows the detection results of the raw dataset and the second row shows the detection results of the super-resolution applied dataset.

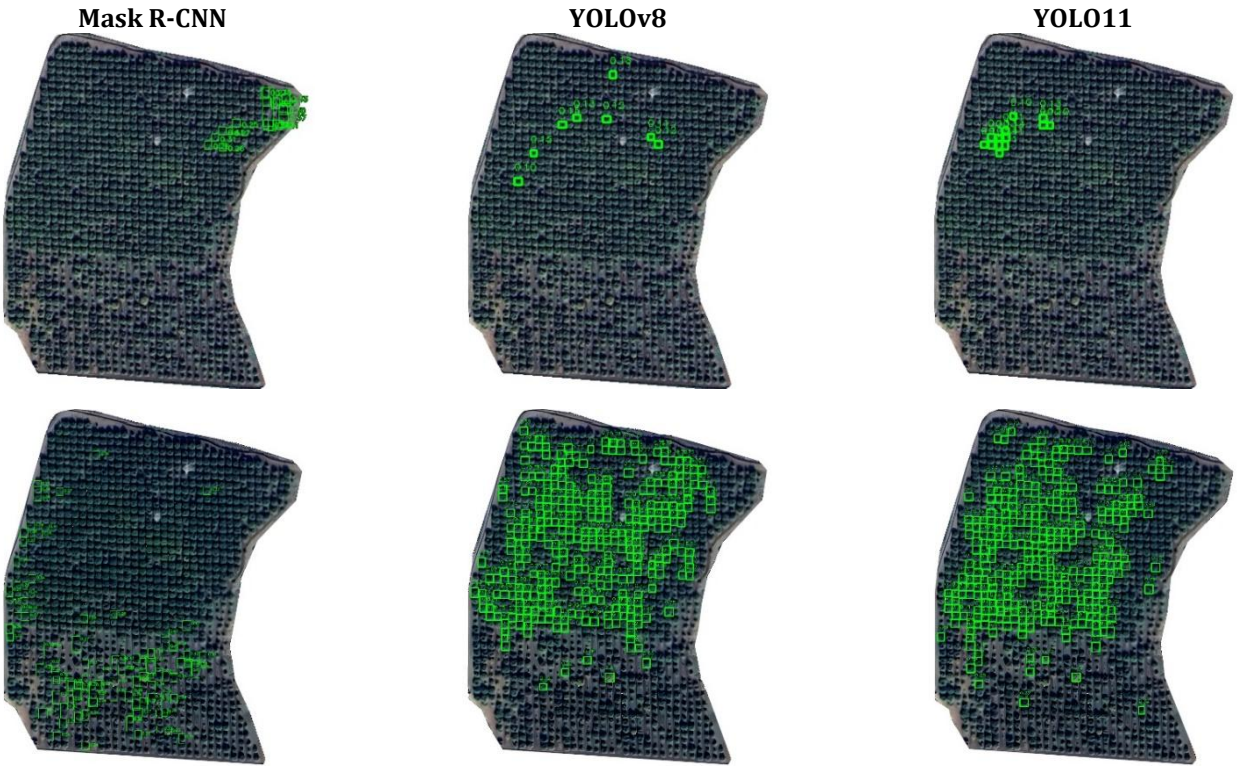


Figure 7. Mask R-CNN, Yolov8 and YOLO11 results for Area 1. (a) Results for the raw image (b) Results for the image enhanced with SR

In Figure 7, in the first row, Mask R-CNN exhibited more false positives (marking an object as present when it does not exist) in the raw image. At the same time, YOLOv8 and YOLO11 produced fewer predictions but with sharper boundaries. In the second row, the performance of all models increased with the application of SR: Mask R-CNN improved the segmentation quality while still showing false positive detections; YOLOv8 and YOLO11 provided more accurate results with fewer false positives and sharper boundaries due to the effect of SR.

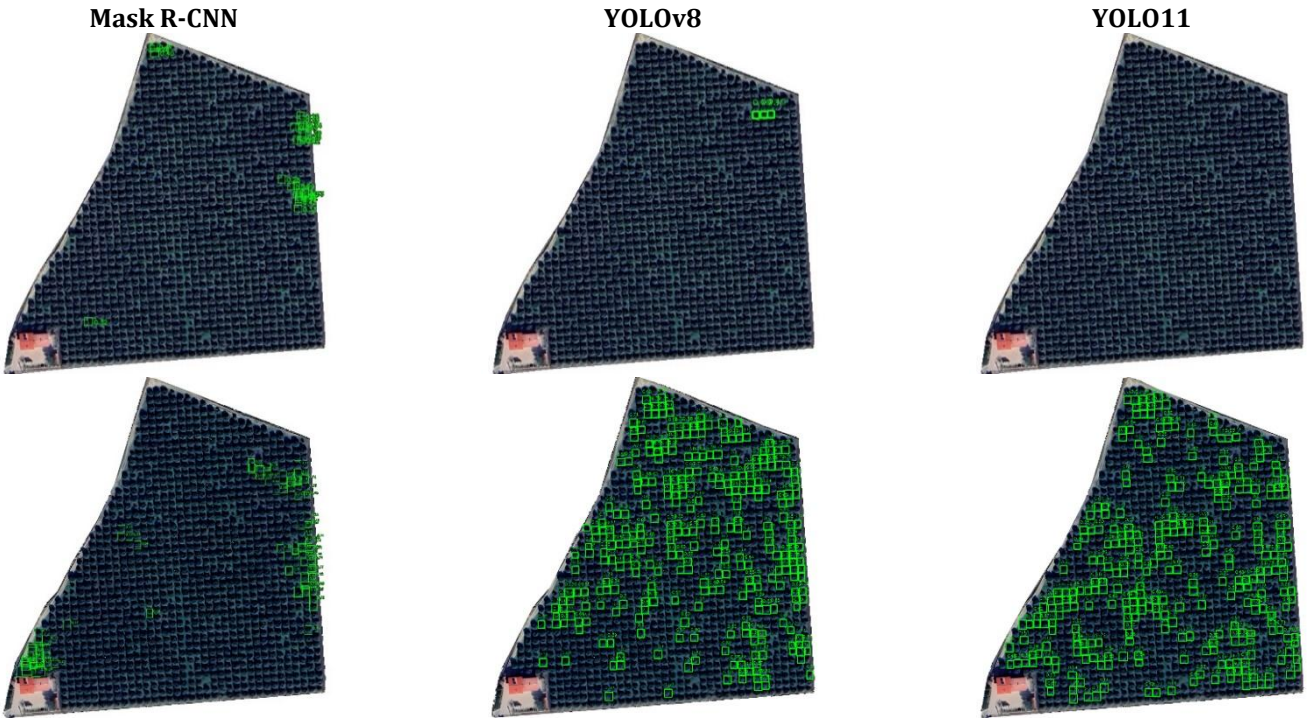


Figure 8. Mask R-CNN, Yolov8 and YOLO11 results for Area 2. (a) Results for the raw image (b) Results for the image enhanced with SR

In the first row of Figure 8, Mask R-CNN detects more objects but contains false positives. In contrast, YOLOv8 and YOLO11 could not make any predictions. When super-resolution is applied, the performance of all models improves: While the segmentation quality of Mask R-CNN increases, the deficiencies in object detection continue. On the other hand, YOLOv8 and YOLO11 produced more distinct boundaries, fewer false positives and more accurate results due to the effect of super-resolution.

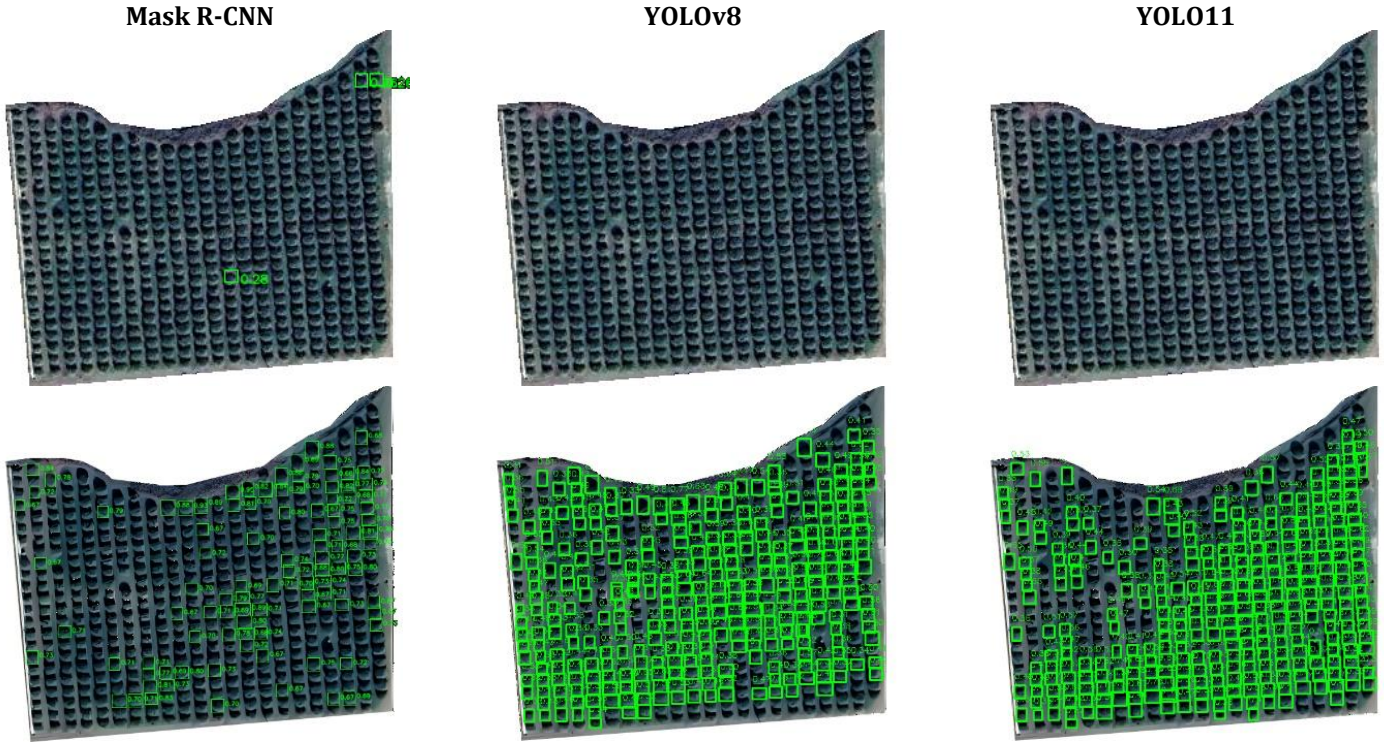


Figure 9. MaskR-CNN, Yolov8 and YOLO11 results for Area 3. (a) Results for the raw image (b) Results for the image enhanced with SR

In the raw image in the first row of Figure 9, YOLOv8 and YOLO11 could not make any predictions. When super-resolution is applied, the improvement in the performance of YOLOv8 and YOLO11 becomes visually more obvious compared to Mask R-CNN.

Visual analysis shows that the performance of object detection algorithms is significantly increased on the super-resolution dataset. It is observed that especially YOLOv8 and YOLO11 models show a significant improvement in terms of object detection accuracy. On the other hand, it is seen that Mask R-CNN model does not provide the expected benefit from super-resolution in more complex segmentation tasks, but still increases the number of detected objects. As a result, it is visually confirmed that super-resolution techniques significantly improve object detection performance by enhancing the details in low-resolution satellite images and make a significant contribution to image processing applications.

4. Discussion

The study focused on detecting citrus trees, leading to the development of an application in this area. For object detection, three different selections were used: Mask R-CNN, YOLOv8, and YOLO11; the models were evaluated on the raw dataset, and the super files (SR) were applied to the dataset. Mask R-CNN obtained 29.47 and 13.29 values in mAP50 and mAP50:95 metrics in the raw dataset, respectively, while these values increased to 60.37 and 43.07 in the SR-applied dataset. YOLOv8 reached 45.73 mAP50 and 22.05 mAP50-95 values in the raw dataset, while it got 84.90 and 67.59 scores in the SR-applied dataset. Similarly, in the YOLO11 model, mAP50 45.39 and mAP50-95 22.15 values were obtained in the raw dataset; In the SR applied dataset, these metrics were measured as 85.93 and 67.66. The

findings of the research can be articulated in this manner. However, it is crucial to compare this article with similar studies in the literature to ensure the reliability and consistency of academic research. In this context, it is known that there are many studies on tree detection in the literature. The studies conducted by Zortea et al. (2018), Ozdarici-Ok (2015), Donmez et al. (2021), and Al Riza et al. (2025) aimed to detect citrus trees. In this context, this research aligns well with existing literature.

High- and low-resolution satellite images and close-up photographs are used to detect citrus trees. Marín-Buzón et al. (2020) detected orange trees using Google Maps images. Yu et al. (2023) conducted a study on citrus trees using Sentinel-2 MSI and Landsat 8 TIRS images. Oussaoui et al. (2025) conducted a study on citrus and olive trees using Sentinel-2 satellite images. Low-resolution Google Earth images were preferred in this article. In the studies conducted by Osco et al. (2020), Hussain & Mehdi (2024), and Csillik et al. (2018), citrus trees were detected using UAV images. In this respect, it is possible to say that higher-resolution images were used in the studies mentioned, and the analyses made were more consistent. However, considering that satellite images were preferred in many studies and successful results were obtained, it seems that free data sources can also be used in the detection of citrus trees. In this context, the fact that the study carries out the application with open-source data shows that it is compatible with the literature.

The choice of detection methods and data sources for analyzing citrus trees is crucial, as it significantly impacts the accuracy of the results. In their research, Li et al. (2016), Yang et al. (2022), and Zhao et al. (2024) favored the Mask R-CNN method. Conversely, Liao et al. (2025), Tian et al. (2022), and Qu et al. (2025) opted for various YOLO models in their studies. This research aims to enhance the existing literature by employing Mask R-CNN, YOLOv8, and YOLO11 models.

5. Conclusions

This study employed deep learning models Mask R-CNN, YOLOv8, and YOLO11 for object-based detection and counting of citrus trees in the Tarsus district of Mersin, focusing on three distinct agricultural regions. Utilizing low-resolution satellite imagery (e.g., Google Earth) as the primary dataset, the research demonstrated the efficacy of deep learning algorithms in agricultural monitoring. It underscored the critical role of super-resolution techniques in enhancing object detection performance. Key findings revealed that the YOLO11 model achieved the highest accuracy, delivering robust and stable results, while SR preprocessing significantly improved performance metrics across all models. For instance, Mask R-CNN's mAP50 increased from 29.47% to 60.37%, YOLOv8's mAP50 rose from 45.7% to 84.9%, and YOLO11 attained 85.93% (mAP50) and 67.66% (mAP50-95) with SR-dataset. These outcomes highlight that refining image resolution through SR mitigates challenges posed by low-quality inputs, enabling more precise feature extraction and localization. Based on the findings and literature review, researchers undertaking similar studies are advised to:

Advancing Super-Resolution Techniques: The integration of advanced super-resolution (SR) methods, such as Real-ESRGAN, to enhance satellite image quality is critical for improving object detection accuracy in agricultural monitoring. Future studies should explore SR's potential to address challenges posed by low-resolution or degraded imagery, particularly in resource-limited settings.

Agricultural Applications: This study's framework can serve as a foundational tool for precision agriculture, enabling applications such as crop health monitoring, yield estimation, and policy formulation. For instance, detecting citrus tree distributions at scale could inform sustainable land management and irrigation strategies. Integrating developed models with GIS technology can be a game-changer for farmers, agricultural organizations, and decision-makers. This powerful combination provides valuable insights and tools to enhance decision-making, improve crop management, and foster more sustainable agricultural practices.

Incorporating Various Data Sources: such as LiDAR for 3D structural analysis and multispectral imagery for vegetation health indices could expand the scope of agricultural insights. This integration would enable comprehensive assessments of tree vitality, stress factors, and productivity forecasts.

Cross-Regional Validation:

The proposed methodology should be tested across geographically and ecologically diverse agricultural regions to ensure generalizability. Validating the model's adaptability to varying climatic conditions, tree densities, and soil types will strengthen its applicability for global agricultural monitoring.

Consequently, the study can constitute an important reference for future studies by revealing the power and potential of deep learning models in object-oriented agricultural analyses.

Author Contributions:

S. Kahveci: Conceptualization, Designed the research methodology, data curation, Software, Writing—original draft.

M. Ö. Çelik: Conceptualization, Designed the research methodology, Visualization, Writing—original draft preparation.

R. Akkurt: Conceptualization, Designed the research methodology, data curation, Software, Writing—original draft.

Ö. Kahveci: Software, Writing-reviewing and editing.

Research and publication ethics statement: In the study, the authors declare that there is no violation of research and publication ethics and that the study does not require ethics committee approval.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Akkurt, R., Kahveci, S., Yetgin, Z., Avaroğlu, E., & Keleş, H. (2024, September 21-22). Detecting red pine seedlings with YOLOv8: A labeling method comparison. [Paper presentation]. 8th International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Türkiye.

Al Riza, D. F., Musahada, L. C., Aufa, R. I., Hermanto, M. B., Nugroho, H., & Hendrawan, Y. (2025). Comparative study of citrus fruits (Citrus reticulata Blanco cv. Batu 55) detection and counting with single and double labels based on convolutional neural network using YOLOv7. *Smart Agricultural Technology*, 10. https://doi.org/10.1016/j.atech.2024.100763

Cheang, E. K., Cheang, T. K., & Tay, Y. H. (2017). Using convolutional neural networks to count palm trees in satellite images. *arXiv.* https://doi.org/10.48550/arXiv.1701.06462

Csillik, O., Cherbini, J., Johnson, R., Lyons, A., & Kelly, M. (2018). Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. *Drones*, 2(4), 39. https://doi.org/10.3390/drones2040039

Donmez, C., Villi, O., Berberoglu, S., & Cilek, A. (2021). Computer vision-based citrus tree detection in a cultivated environment using UAV imagery. *Computers and Electronics in Agriculture*, 187, 106273. https://doi.org/10.1016/j.compag.2021.106273

El Akrouchi, M., Mhada, M., Bayad, M., Hawkesford, M. J., & Gérard, B. (2025). AI-Based Framework for Early Detection and Segmentation of Green Citrus fruits in Orchards. *Smart Agricultural Technology*, 100834. https://doi.org/10.1016/j.atech.2025.100834

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88, 303-338.

Fan, Z., Lu, J., Gong, M., Xie, H., & Goodman, E. D. (2018). Automatic tobacco plant detection in UAV images via deep neural networks. IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 11(3), pp. 876-887. https://doi.org/10.1109/10.1109/JSTARS.2018.2793849

FAO. (2025a). Retrieved March 10, 2025, from https://www.fao.org/faostat/en/#data/RL/visualize

FAO. (2025b). Retrieved March 10, 2025, from https://www.fao.org/faostat/en/#data/QC/visualize

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE international conference on computer vision, 2961-2969.

Hunt Jr, E. R., & Daughtry, C. S. (2018). What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture? *International journal of remote sensing*, 39(15-16), 5345-5376. https://doi.org/10.1080/01431161.2017.1410300

Hussain, M. H., & Mehdi, S. A. (2024, October 15-16). Spatial recognition of citrus trees, seedlings, and gaps: A deep learning and geometric approach on satellite-based RGB imagery [Paper presentation]. Horizons of Information Technology and Engineering (HITE), Lahore, Pakistan. https://doi.org/10.1109/HITE63532.2024.10777150

Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv*. https://doi.org/10.48550/arXiv.2410.17725

Kouvaras, L., & Petropoulos, G. P. (2024). A Novel technique based on machine learning for detecting and segmenting trees in very high resolution digital images from unmanned aerial vehicles. *Drones*, 8(2), 43. https://doi.org/10.3390/drones8020043

Li, D., Guo, H., Wang, C., Li, W., Chen, H., & Zuo, Z. (2016). Individual tree delineation in windbreaks using airborne-laser-scanning data and unmanned aerial vehicle stereo images. IEEE Geoscience and Remote Sensing Letters, 13(9), pp. 1330-1334. https://doi.org/10.1109/LGRS.2016.2584109

Liao, Y., Li, L., Xiao, H., Xu, F., Shan, B., & Yin, H. (2025). YOLO-MECD: Citrus detection algorithm based on YOLO11. *Agronomy*, 15(3), 687. https://doi.org/10.3390/agronomy15030687

Marín-Buzón, C., Pérez-Romero, A., Tucci-Álvarez, F., & Manzano-Agugliaro, F. (2020). Assessing the orange tree crown volumes using google maps as a low-cost photogrammetric alternative. *Agronomy*, 10(6), 893. https://doi.org/10.3390/agronomy10060893

MGM. (2025). Retrieved March 22, 2025, from https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx?k=A&m=MERSIN

Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: a comprehensive survey. *Machine vision and applications*, 25, pp. 1423-1468. https://doi.org/10.1007/s00138-014-0623-4

Orhan, O. (2021). Land suitability determination for citrus cultivation using a GIS-based multi-criteria analysis in Mersin, Turkey. *Computers and Electronics in Agriculture*, 190, 106433. https://doi.org/10.1016/j.compag.2021.106433

Osco, L. P., De Arruda, M. D. S., Junior, J. M., Da Silva, N. B., Ramos, A. P. M., Moryia, É. A. S., Imai, N. N., Pereira, D. R., Creste, É. A. S. Matsubara, E. T., Li, J., & Gonçalves, W. N. (2020). A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 97-106. https://doi.org/10.1016/j.isprsjprs.2019.12.010

Oussaoui, S., Boudhar, A., Hadri, A., Lebrini, Y., Houmma, I. H., Karaoui, I., El Khalki E. M., Ouzemou J. E., & Kinnard, C. (2025). Mapping drought severity impact on arboriculture systems over Tadla and lower Tassaout plains in Morocco using Sentinel-2 data and machine learning approaches. *Geocarto International*, 40(1), 2471104. https://doi.org/10.1080/10106049.2025.2471104

Ozdarici-Ok, A. S. L. I. (2015). Automatic detection and delineation of citrus trees from VHR satellite imagery. *International Journal of Remote Sensing*, 36(17), pp. 4275-4296. https://doi.org/10.1080/01431161.2015.1079663

Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., & Da Silva, E. A. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. Electronics, 10(3), 279.

Qu, H., Du, H., Tang, X., & Zhai, S. (2025). Citrus fruit diameter estimation in the field using monocular camera. *Biosystems Engineering*, 252, pp. 47-60. https://doi.org/10.1016/j.biosystemseng.2025.02.012

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 658-666).

Sapkota, R., Ahmed, D., & Karkee, M. (2024). Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. *Artificial Intelligence in Agriculture*, 13, pp. 84-99. https://doi.org/10.1016/j.aiia.2024.07.001

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), 1-48. https://doi.org/10.1186/s40537-019-0197-0

Tian, H., Fang, X., Lan, Y., Ma, C., Huang, H., Lu, X., Zhao, D., Liu, H., & Zhang, Y. (2022). Extraction of citrus trees from UAV remote sensing imagery using YOLOv5s and coordinate transformation. *Remote Sensing*, 14(17), 4208. https://doi.org/10.3390/rs14174208

Toosi, A., Javan, F. D., Samadzadegan, F., Mehravar, S., Kurban, A., & Azadi, H. (2022). Citrus orchard mapping in Juybar, Iran: Analysis of NDVI time series and feature fusion of multi-source satellite imageries. *Ecological Informatics*, 70, 101733. https://doi.org/10.1016/j.ecoinf.2022.101733

TÜİK. (2025). Retrieved March 10, 2025, from https://biruni.tuik.gov.tr/medas/?locale=tr

Ultralytics. (2023). Retrieved March 10, 2025, from https://github.com/ultralytics/ultralytics

Uyar, N. (2024). Monitoring and analysis of air quality in Zonguldak Province by remote sensing. *Turkish Journal of Remote Sensing*, 6(1), 57-67. https://doi.org/10.51489/tuzal.1484324

Wang, X., Xie, L., Dong, C., & Shan, Y. (2021, March 10). Real-esrgan: Training real-world blind super-resolution with pure synthetic data [Paper presentation]. International conference on computer vision (ICCV 2021).

Wang, Y., Bashir, S. M. A., Khan, M., Ullah, Q., Wang, R., Song, Y., Guo, Z., & Niu, Y. (2022). Remote sensing image super-resolution and object detection: Benchmark and state of the art. *Expert Systems with Applications,* 197, 116793. https://doi.org/10.1016/j.eswa.2022.116793

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E. (2019). Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. Remote Sensing, 11(11), 1309.

Weather Spark. (2025). Retrieved March 10, 2025, from https://tr.weatherspark.com/y/98265/Tarsus-T%C3%BCrkiye-Ortalama-Hava-Durumu-Y%C4%B1l-Boyunca

Yang, M., Mou, Y., Liu, S., Meng, Y., Liu, Z., Li, P., Xiang, W., Zhou, X., & Peng, C. (2022). Detecting and mapping tree crowns based on convolutional neural network and Google Earth images. *International Journal of Applied Earth Observation and Geoinformation*, 108, 102764. https://doi.org/10.1016/j.jag.2022.102764

Yu, K., Hao, Z., Post, C. J., Mikhailova, E. A., Lin, L., Zhao, G., Tian, S., & Liu, J. (2022). Comparison of classical methods and mask R-CNN for automatic tree detection and mapping using UAV imagery. *Remote Sensing*, 14(2), 295. https://doi.org/10.3390/rs14020295

Yu, G., Zhang, L., Luo, L., Liu, G., Chen, Z., & Xiong, S. (2023). Mapping Insect-Proof Screened Citrus Orchards Using Sentinel-2 MSl Time-Series Images. Remote Sensing, 15(11), 2867. https://doi.org/10.3390/rs15112867

Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and remote sensing magazine, 4(2), 22-40.

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020, April). Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 12993-13000).

Zortea, M., Macedo, M. M., Mattos, A. B., Ruga, B. C., & Gemignani, B. H. (2018, October 29, November 1). Automatic citrus tree detection from UAV images based on convolutional neural networks [Paper presentation]. Proceedings of the 31th Sibgrap/WIA—Conference on Graphics, Patterns and Images, SIBGRAPI. 18. Paraná, Brazil.