

# Comparing YOLOv8 and Mask RCNN for object segmentation in complex orchard environments

Ranjan Sapkota, Dawood Ahmed and Manoj Karkee

*Center for Precision & Automated Agricultural Systems, Washington State University, 24106 N Bunn Rd, Prosser, 99350, Washington, USA*

---

## Abstract

Instance segmentation, an important image processing operation for automation in agriculture, is used to precisely delineate individual objects of interest within images, which provides foundational information for various automated or robotic tasks such as selective harvesting and precision pruning. This study compares the one-stage YOLOv8 and the two-stage Mask R-CNN machine learning models for instance segmentation under varying orchard conditions across two datasets. Dataset 1, collected in dormant season, includes images of dormant apple trees, which were used to train multi-object segmentation models delineating tree branches and trunks. Dataset 2, collected in the early growing season, includes images of apple tree canopies with green foliage and immature (green) apples (also called fruitlet), which were used to train single-object segmentation models delineating only immature green apples. The results showed that YOLOv8 performed better than Mask R-CNN, achieving good precision and near-perfect recall across both datasets at a confidence threshold of 0.5. Specifically, for Dataset 1, YOLOv8 achieved a precision of 0.90 and a recall of 0.95 for all classes. In comparison, Mask R-CNN demonstrated a precision of 0.81 and a recall of 0.81 for the same dataset. With Dataset 2, YOLOv8 achieved a precision of 0.93 and a recall of 0.97. Mask R-CNN, in this single-class scenario, achieved a precision of 0.85 and a recall of 0.88. Additionally, the inference times for YOLOv8 were 10.9 ms for multi-class segmentation (Dataset 1) and 7.8 ms for single-class segmentation (Dataset 2), compared to 15.6 ms and 12.8 ms achieved by Mask R-CNN's, respectively. These findings show YOLOv8's superior accuracy and efficiency in machine learning applications compared to two-stage models, specifically Mast-RCNN, which suggests its suitability in developing smart and automated orchard operations, particularly when real-time applications are necessary in such cases as robotic harvesting and robotic immature green fruit thinning.

**Keywords:** *Machine Learning, Deep learning, YOLOv8, Mask R-CNN, Automation, Robotics, Artificial intelligence, Machine Vision*

---

## 1. Introduction

Instance segmentation is a powerful computer vision technique that combines the benefits of both object detection and semantic segmentation [1]. One of the key benefits of instance segmentation in agricultural applications is its ability to accurately quantify plant and crop structures [2], which can provide valuable information about plant growth, disease identification, and yield estimation, and can provide a foundation for various key areas of research and development such as robotic green (immature) fruit thinning [3]. Instance segmentation can provide precise measurements of plant features, such as leaf area, stem length, and plant height, with a high level of accuracy and efficiency [4], [5]. The traditional methods of instance segmentation in agricultural images were mostly based on hand-crafted features and classical image processing techniques such as Watershed Transform [6], Graph-based Segmentation [7], Active Contours (or Snakes) [8], [9], level set [10]–[12], Region Growing [10]–[12], Morphological Operation [13], [14] and Clustering-based methods [15], [16]. However, these methods require a lot of manual setup and refinements, making them time-consuming and less reliable [17]. Additionally, these methods couldn't easily learn from new data, making them less flexible and difficult to adapt to different scenarios. Moreover, these methods involved multiple disjointed image processing stages such as noise removal, contrast adjustment, image enhancement, refinement and manually defining and/or extracting specific features such as edge, texture, or colors.

Machine Learning (ML) and, more notably, Deep Learning (DL) have revolutionized the domain of instance segmentation in image processing including those in agricultural applications, offering more reliable and robust techniques compared to the traditional instant segmentation techniques [18]. DL models, particularly convolutional neural networks (CNNs), have shown remarkable ability to automatically learn features from vast datasets without the necessity for manually extracting features [19]. This self-learning capability makes them adaptive, which is essential to develop models generalized across diverse agricultural scenarios and conditions. With these capabilities, [20][21][22] have been widely used in recent years to perform various tasks in agricultural productions such as plant disease identification [23] and yield prediction [24], [25].

More specifically, DL network architectures, including U-Net [20], Mask R-CNN [21], and YOLO [22], are increasingly utilized for a range of applications in agriculture. A key advantage of these DL techniques is their end-to-end learning approach, which enables direct mapping of raw images to segmentation results, thus enhancing consistency and reliability. Furthermore, transfer learning techniques allow for the adaptation of models pre-trained

on extensive datasets to specific agricultural tasks, reducing both training times and data requirements. Utilizing these features of DL models, various agricultural applications have been investigated including plant disease identification [23], yield prediction [24], [25], pest detection [26], [27], soil health assessment [28], crop maturity analysis [29], and site-specific weed control application [30]–[32] showcasing their versatility and efficiency in modern agricultural practices.

As mentioned before, instance segmentation techniques have been widely applied to crop disease management [33]. Early detection of plant diseases is crucial for maintaining crop yield and quality. Utilizing instance segmentation, researchers can quantify symptoms such as leaf spots and discoloration and monitor the progression of these diseases over time [34]. This capability is instrumental in developing effective disease management strategies, including targeted treatments and breeding for disease-resistant cultivars. Instance segmentation has also been proved pivotal for precise crop yield estimation. Accurate yield estimation is essential for growers and breeders to make informed decisions about crop management and to select traits for breeding new cultivars. Instance segmentation techniques can be used to accurately counting and sizing individual fruits or other canopy objects from images. Such information facilitates a precise yield estimation and provides key insights into cultivar characteristics [33]. Past studies have demonstrated the effectiveness of these techniques in various relevant applications such as the segmentation of apple flowers [34], segmentation and localization of strawberry fruits for harvesting segmentation and counting of cranberries, and the segmentation of guava fruits and branches [35]. The data derived from these studies assist in optimizing crop management strategies, including optimal application of water and fertilizers, and identifying high-yielding cultivars [34] as mentioned before.

In addition, instance segmentation has been applied extensively to develop machine vision systems for agricultural robots because it provides capabilities for robots to detect, delineate, and track individual objects of interest in agricultural fields using images or videos, such as fruits, branches, flowers, vegetables, and livestock [36]. Detecting and tracking plant parameters such as leaves, stem, trunk, branch, flower, and fruit is necessary for a robot to automatically perform various tasks such as harvesting, and canopy, and crop-load management operations. In the last few years, several studies have implemented the use of deep learning-based instance segmentation techniques for developing robotic solutions for various agricultural applications such as tree pruning in dormant season [37] picking fruits and vegetables [38]–[40], thinning flowers [41], [42] and fruitlet [43] and identifying and killing weeds [44], [45] among others.

Among the broad applications of deep learning techniques in agriculture, there has been a focus on the use of two specific architectures: YOLO (You Only Look Once) and Mask-R-CNN. These models, known for their effectiveness in instance segmentation, have been pivotal in advancing tasks such as crop detection, pest and disease management, weed identification, tree canopy segmentation, and canopy object (e.g., branch and fruit) detection. These tasks, critical in precision and automated agriculture, benefit immensely from the capabilities of these two deep learning models. As mentioned before, many recent studies conducted in agricultural applications used Mask-RCNN-based [46] instance segmentation for tasks such as crop detection [47], [48], pest and disease detection [49]–[51], weed detection [52], [53], tree canopy segmentation [54], [55], and tree branch detection [54]–[56]. Concurrently, the YOLO family of models has been used widely in object detection because of its ability to handle tasks like object detection, image classification, and instance segmentation simultaneously with one-stage networks. Unlike Mask R-CNN, a two-stage model suitable for segmentation tasks [57], YOLO optimizes the overall processing ensuring speed and efficiency crucial for real-time applications in agriculture such as robotic pruning [58], thinning [59], and pesticide application [60]. Both models are extensively studied, as discussed above and as shown in Table 1, highlighting 23 publications in the last 3 years focusing specifically on analyzing images of modern apple tree canopies.

**Table 1: Highlighting the studies conducted in the last three years on YOLO and Mask RCNN during different apple orchard environments.**

References	Year	DL model	Objectives
[61], [62]	2021	YOLO-V4	Apple detection in a complex scene
[63], [64]	2021	Mask R-CNN	Deep learning-based apple detection
[63], [65]	2021	YOLO-V3	Green fruit detection (apples, mangoes)
[65], [66]	2021	YOLO-V5	Apple fruitlet detection for fruitlet thinning
[67]	2022	Mask R-CNN	Branch identification and junction points localization in apple trees; Trunk identification and segmentation
[68], [69]	2022	YOLO-V4	Apple detection, counting, and tree trunk tracking in modern orchards

[70]	2022	YOLO-V4	Immature/mature apple detection on dense-foliage tree architectures for early crop-load estimation
[71]	2022	YOLO-V5	Identification method for the apple growth pattern in the orchard
[72]	2022	YOLO-V5	Tree trunk and obstacle detection in apple orchards
[73], [74]	2022	Mask R-CNN	Ripe and green apple segmentation in orchards
[75], [76]	2022	Mask R-CNN	Tree and tree crown segmentation in orchards
[77]	2023	YOLO-V3	Apple fruit quality detection
[78]	2023	YOLO-V8	Tree trunk and branch detection
[79]	2023	YOLO-V7	Detection and counting of small target apples
[59], [74]	2023	Mask R-CNN	Green apple segmentation

Building upon this background of widespread application of YOLOv8 and Mask R-CNN models, the primary goal of this study is to systematically compare and evaluate the performance of these two models (YOLOv8 and Mask R-CNN) for instance segmentation tasks in modern, commercial apple orchards. Through this comprehensive comparison, this research aims to provide insights into the suitability, efficiency, and potential challenges associated with implementing each model in agricultural automation applications. To achieve this goal, the following specific objectives will be pursued in this study:

- To compare the performances of YOLOv8 and Mask R-CNN models in segmenting **single-class objects**, specifically green apples (fruitlets), in images collected from variable orchard environments in the early growing season; and
- To evaluate the capabilities of these two models in segmenting **multi-class objects**, specifically primary branches and tree trunks of apple trees in images collected from a model apple orchard during the dormant season.

## 2. Background

### 2.1 Mask RCNN

Mask R-CNN is a deep learning model designed for object detection and instance segmentation, renowned for its accuracy and efficiency. Its strength lies in its ability to precisely identify and delineate each object in an image, making it highly effective for complex image analysis tasks. The model was developed by researchers at Facebook AI Research in 2017 and builds on top of the Faster R-CNN object detection model by adding a branch for predicting object masks in parallel with the existing branch for bounding box detection [46]. The architecture of Mask R-CNN consists of three main components: a backbone network, a region proposal network (RPN), and two parallel branches for bounding box detection and mask prediction as shown in Figure 1. The backbone network is typically a convolutional neural network (CNN) that extracts features from the input images and is shared by both branches. The RPN generates a set of region proposals that are likely to contain objects, based on the feature maps generated by the backbone network. The bounding box branch predicts the class label and bounding box coordinates for each region proposal, while the mask branch predicts a binary mask for each object instance within the bounding box.

### 2.2 YOLOv8

The YOLO (You Only Look Once) family of object detection and instance segmentation models have evolved rapidly over the last several years, with each new iteration introducing improvements in accuracy and/or speed. YOLOv8 (Figure 2), the latest one-stage model, was built on the foundations provided by previous YOLO models, such as YOLOv3 and YOLOv5. Compared to two-stage models, YOLOv8 directly predicts bounding boxes and class probabilities without the need for a separate region proposal network, streamlining the object detection process. One key innovation in YOLOv8 is the adoption of an anchor-free, center-based approach for object detection, which offers several advantages over traditional anchor-based methods such as YOLOv5, YOLOv6, and YOLOv7. YOLOv8 implements Pseudo Ensemble or Pseudo Supervision (PS), a method that involves training multiple models with distinct configurations on the same dataset to generate a more diverse set of predictions, improving the accuracy and robustness of the final prediction. Additionally, YOLOv8 leverages the Darknet-53 architecture, a 53-layer deep convolutional neural network optimized for feature extraction and object detection. One significant architectural change in YOLOv8 is the replacement of the C3 module with the C2F module. The C3 module, also known as the convolutional module, processes input data through a series of convolutional operations. The C2F module, an improved version of the C3 module, enhances accuracy and processing times compared to previous models. Furthermore, YOLOv8 substitutes the 6x6 Convolutional (Conv) layer with a 3x3 Conv layer in the model backbone,

reducing the number of parameters and creating a more compact, computationally efficient network. YOLOv8 also employs a decoupled head, which separates the tasks of predicting object presence and classifying object types, thereby improving both accuracy and processing speed. This refinement positions YOLOv8 as an effective solution for both object detection and instance segmentation in computer vision.

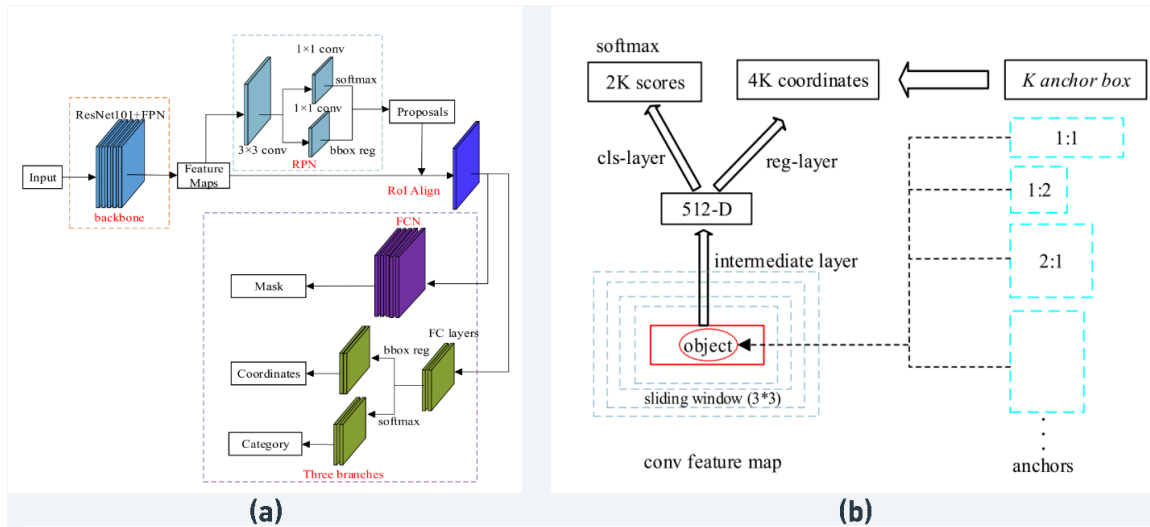


Figure 1: Mask-RCNN architecture with; (a) structure diagram, highlighting the backbone network, RPN, bounding box, and mask prediction branches; and (b) detailed view of the Region Proposal Network (RPN).

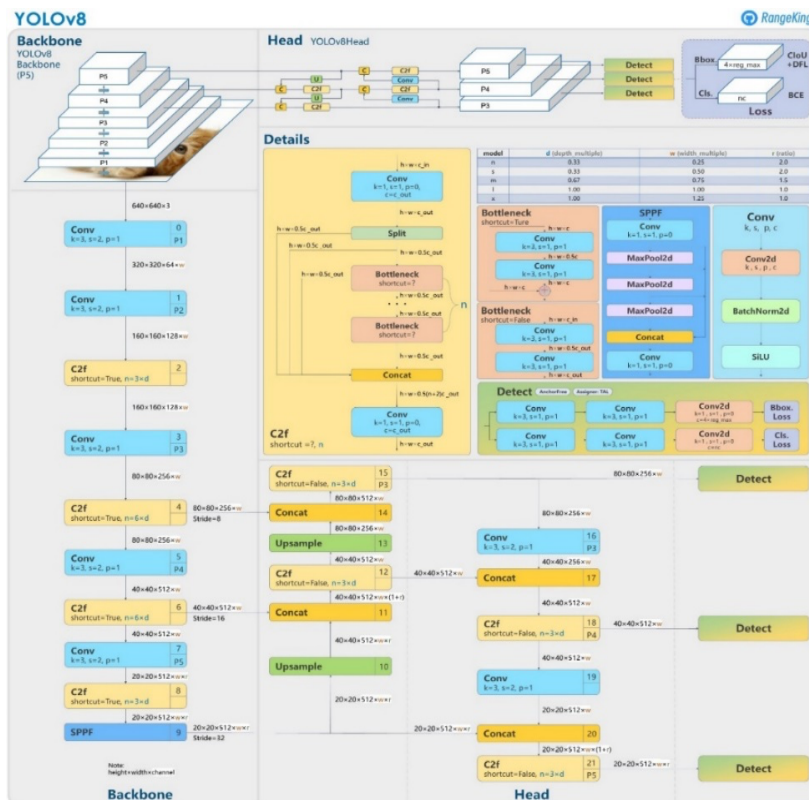
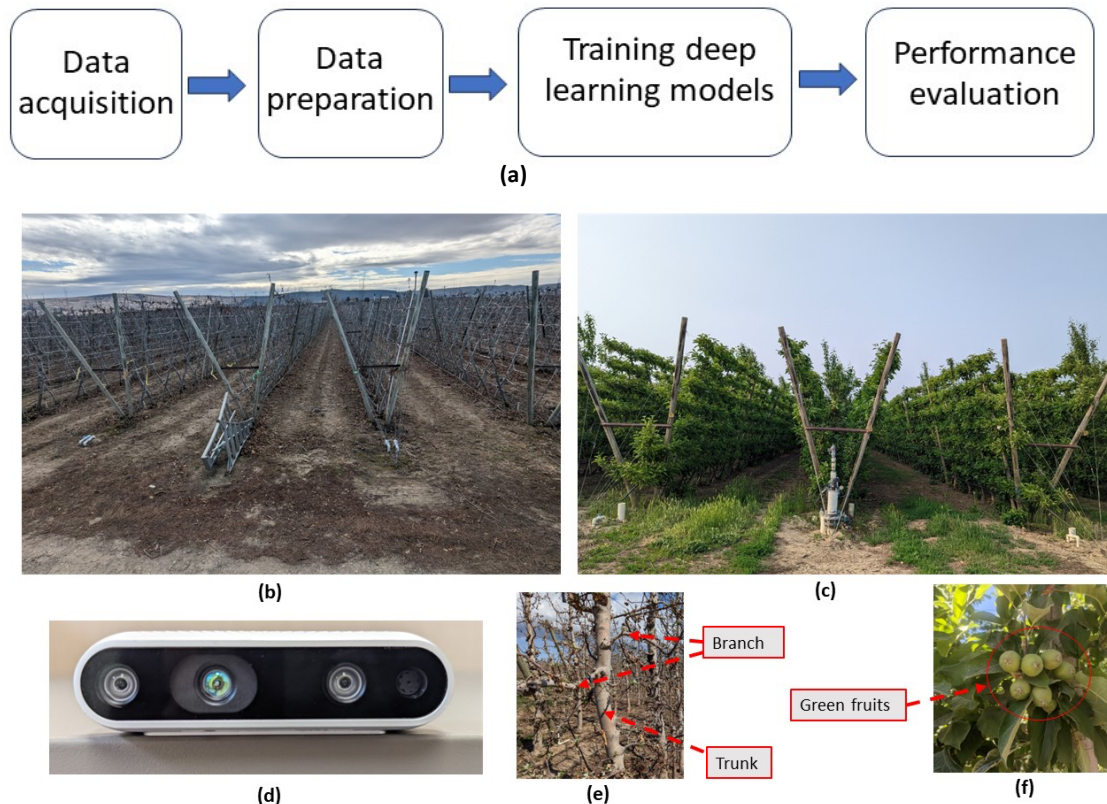


Figure 2: YOLOv8 architecture from RangeKing's GitHub repository, showcasing its innovative design for object detection and segmentation (<https://github.com/RangeKing>).

### 3. Materials and Methods

This study consisted of four major steps as outlined in Figure 3a beginning with RGB images acquisition from commercial orchards in two distinct seasons (Figure 3b as dormant season and 3c as early growing season). These images, captured under varying environmental conditions such as bright and cloudy days, were then manually annotated to create the training and testing datasets. The training dataset was subsequently used to train the two deep learning models mentioned previously, and their performance in instance segmentation was evaluated using the test dataset.



**Figure 3:** (a) Overall workflow diagram used in this research; (b) An example image of an Apple orchard during the dormant season (November 22, 2022); (c) An example image of an apple orchard during early fruit growing season (June 18, 2023); (d) IntelRealSense 435i camera used to acquire images to train and test the instance segmentation models; (e) Example trunks and branches used to annotate the dormant season images; and (f) Example immature green fruits (fruitlets) used to annotate early growing season images.

#### 3.1 Study site and data acquisition

This study was conducted in a commercial apple orchard (Figures 3b and 3c) owned and operated by Allan Brothers Fruit Company, located at Prosser, Washington State, USA. The orchard was planted in 2009 with a Scilate apple cultivar with a row spacing of 9.0 ft, and a plant spacing of 3.0 f, and was trained to a V-trellis architecture. Two sets of RGB images were acquired using IntelRealSense 435i (Intel Corporation, California, USA); one in November 2022 creating dormant season dataset as shown in Figures 3b and 3e, while the other set of images was acquired in June 2023 (just before manual fruitlet thinning) which provided the dataset for early growing season as illustrated in Figure 3c and 3f. The Intel RealSense camera was selected for capturing RGB images due to its SDK's ability to adjust parameters and capture high-quality images.

### 3.2 Data preparation

Two kinds of datasets comprising 1,553 RGB images, capturing a variety in orchard lighting conditions were prepared for analysis of the deep learning model. Dataset 1 comprised 474 images from the dormant season, which were annotated manually to represent multi-class objects: the tree trunk and primary branches growing out from the trunks (Figure 4). Altogether, 1,141 annotations for the tree trunk and 2,369 annotations for the tree branches were generated manually by creating the polygon over desired objects in these images using the image labeling software Labelbox. Likewise, the dataset comprised 1,079 images from the green fruit growing season in which 5,921 annotations of immature green apples were generated. During the image preprocessing stage using the label box software, all these annotations were formatted in accordance with the COCO dataset specification, which meets the requirement of both the YOLOv8 and Mask RCNN model for image segmentation. Furthermore, to facilitate model training and validation, both datasets were resized to 640 x 640 pixels, and both datasets were divided randomly into training, validation, and test subsets, following an 8:1:1 distribution ratio for each object class.

### 3.3 Deep Learning Model Implementation

Both the YOLOv8 and Mask R-CNN models were trained on a workstation with an Intel Xeon® W-2155 CPU @ 3.30 GHz x20 processor, NVIDIA TITAN Xp Collector's Edition/PCIe/SSE2 graphics card, 31.1 GiB memory, and Ubuntu 16.04 LTS 64-bit operating system. The backend framework for the model implementation was Pytorch, operating on a Linux system. To optimize performance, the learning rate used was 0.001, the batch size used was 32, and the dropout rate used was 0.5 to mitigate overfitting. The training was conducted over 1,000 iterations. The model training was stopping before reaching 1000 epochs, if the model performance did not improve for 20 consecutive epochs over the validation dataset, which was useful to minimize model overfitting to the training dataset and improving generality. Images were resized to dimensions of 640x640 pixels and batched into groups of 16 images. An initial learning rate of 0.01 was used in training both models, whereas the momentum and weight decay used were 0.937 and 0.0005 respectively for the two models. These parameter settings were chosen to optimize the speed of the training process while minimizing the chances of overfitting the model to the training dataset. During the initial three epochs, a warm-up phase was employed, using a momentum of 0.8 and a bias learning rate of 0.1, to stabilize the model's optimization and mitigate the risk of being stuck at a poor local minimum.

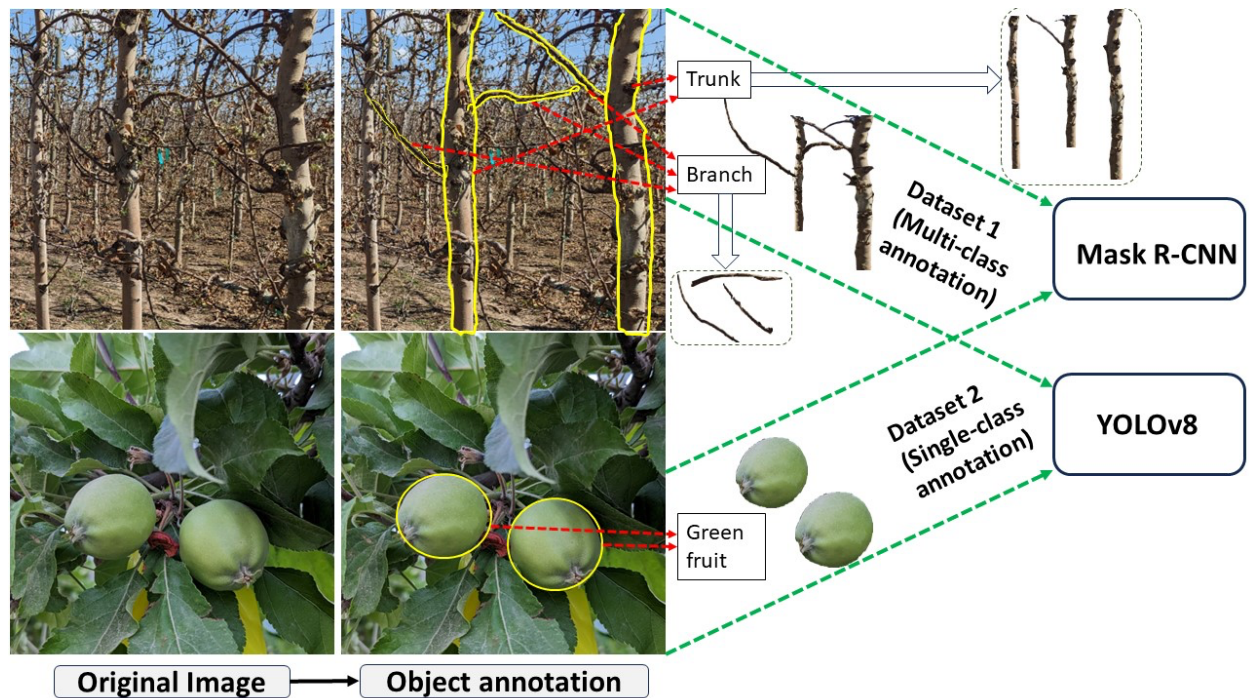


Figure 4: Workflow diagram showing the two types of datasets used in the study; Dataset 1 included the dormant season apple trees with multi-class objects (Trunk and branch) and Dataset 2 included growing season apple tree canopies with immature green fruits.

During the training process, various augmentation techniques were applied to enhance model robustness and generalization such as hue augmentation (0.015), saturation augmentation (0.7), value augmentation (0.4), translation adjustments (0.1), scaling variations (0.5), and a 50% probability for left-right flips. Additionally, a mosaic augmentation was applied with a probability of 1.0. After the model training was completed, the model outputs were converted to TorchScript format to simplify further processing to evaluate the performances of both YOLOv8 and Mask R-CNN models in terms of precision, recall, mean average precision (mAP), and area under curve (AUC) as discussed below.

### 3.4 Performance Evaluation

To evaluate the instance segmentation capabilities of the Mask R-CNN and YOLOv8 models, five distinct metrics were used: Precision, Recall, mean Average Precision (AP) at 0.5 intersection over union (mAP@0.5 IOU), Area Under the receiver operating characteristic Curve (AUC), and Inference speed. Precision is defined as the proportion of correctly identified positive instances to the total predicted positive instances, as depicted by equation 1. Similarly, recall, depicted by equation 2, quantified the proportion of correctly identified positive instances out of all actual instances of the target objects. Furthermore, the mean average precision (mAP), represented as the average of the AP across k categories (equation 4), was crucial in evaluating the model's precision at a threshold of 50% overlap between predicted and true object boundaries/bounding boxes. The area under the curve (AUC), defined by equation 5, assessed the model's classification efficacy across all possible thresholds. The model's efficiency in processing and delivering predictions was measured by the inference speed and was inversely related to the time taken per image analysis. These metrics are calculated as follows:

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

*Equation 1*

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

*Equation 2*

$$IoU = \frac{Area\ Overlap}{Area\ Union} = \frac{TP}{FP + TP + FN}$$

*Equation 3*

$$mAP = \left(\frac{1}{K}\right) \sum_{i=0}^k (AP)_i$$

*Equation 4*

$$AUC = \int_0^1 TPR(FPR)^{-1}(u) du$$

*Equation 5*

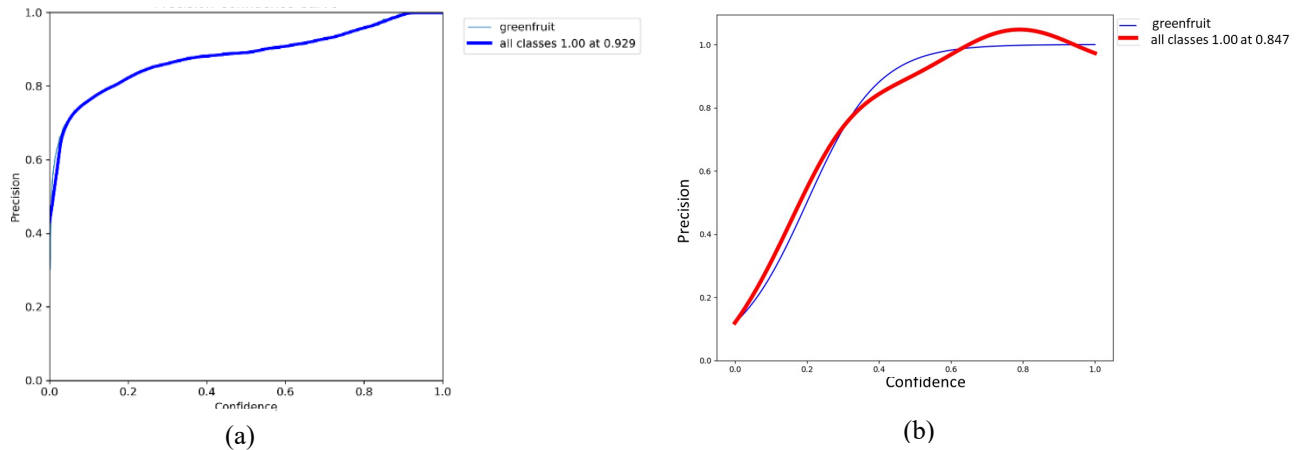
where TP, FP, and FN represent true positive, false positive, and false negative object instances respectively. Variable 'k' represents the total number of object classes, and  $(AP)_i$  refers to the average precision calculated for the  $i^{th}$  class among these k classes. AP is the area under the precision-recall curve for a given class. TPR represents the true positive rate, FPR is the false positive rate, and t indicates the time taken for the model to infer results for a given (single) image.

## 4. Results and Discussion

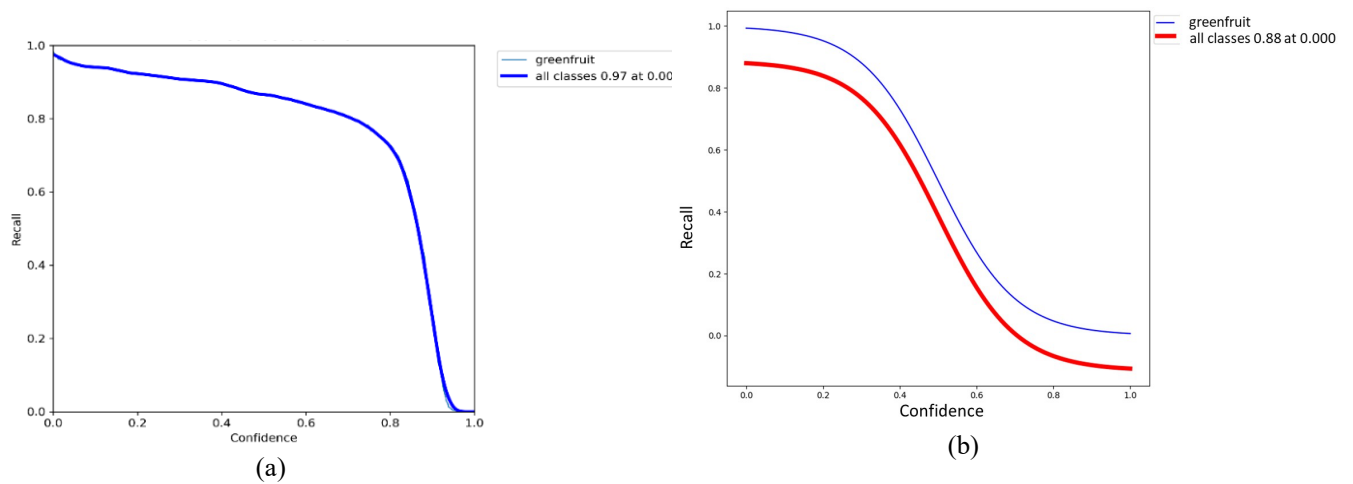
### 4.1 Single-class Object Segmentation of Immature Green Apples (Fruitlets)

For single-class segmentation for immature green fruits, the Precision-Confidence curves depicted in Figure 5, revealed that the YOLOv8 model achieved a maximum precision of 1.00 when the confidence threshold was 0.929 (Figure 5a). Correspondingly, Recall-Confidence curves for the respective models are presented in Figure 6, which

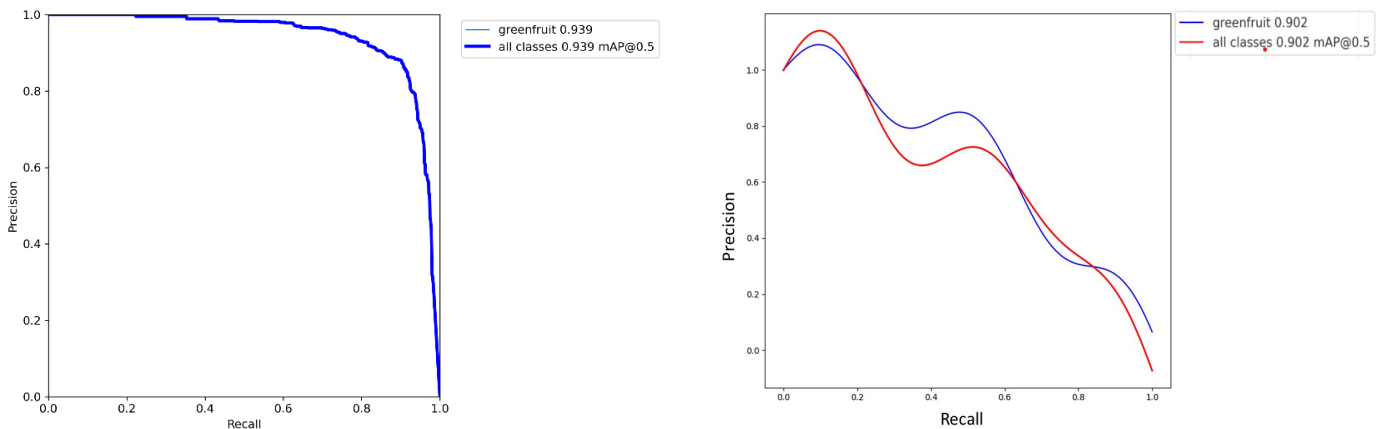
showed that YOLOv8's recall reached 0.97 at the minimum confidence threshold of 0.000. This high recall rate, or sensitivity, indicates the model's ability to correctly identify a high percentage of actual objects, which showed models effectiveness in segmenting green fruits even at the lowest confidence levels. Additionally, YOLOv8 outperformed Mask R-CNN in terms of mean average precision (mAP), achieving 0.939 at a 0.5 IoU threshold for green fruits and overall categories, compared to mAP of 0.902 achieved with Mask R-CNN (Figure 7).



**Figure 5: Precision-Confidence curve for single class segmentation of immature green apples (fruitlets) using; (a) YOLOv8; and (b) Mask R-CNN.**



**Figure 6: Recall-Confidence curve for single class segmentation of green apple fruits using; (a) YOLOv8; and (b) Mask R-CNN**



**Figure 7: Precision-Recall curve for single class segmentation of green apple fruits at mAP@0.5; (a) YOLOv8; and (b) Mask R-CNN**



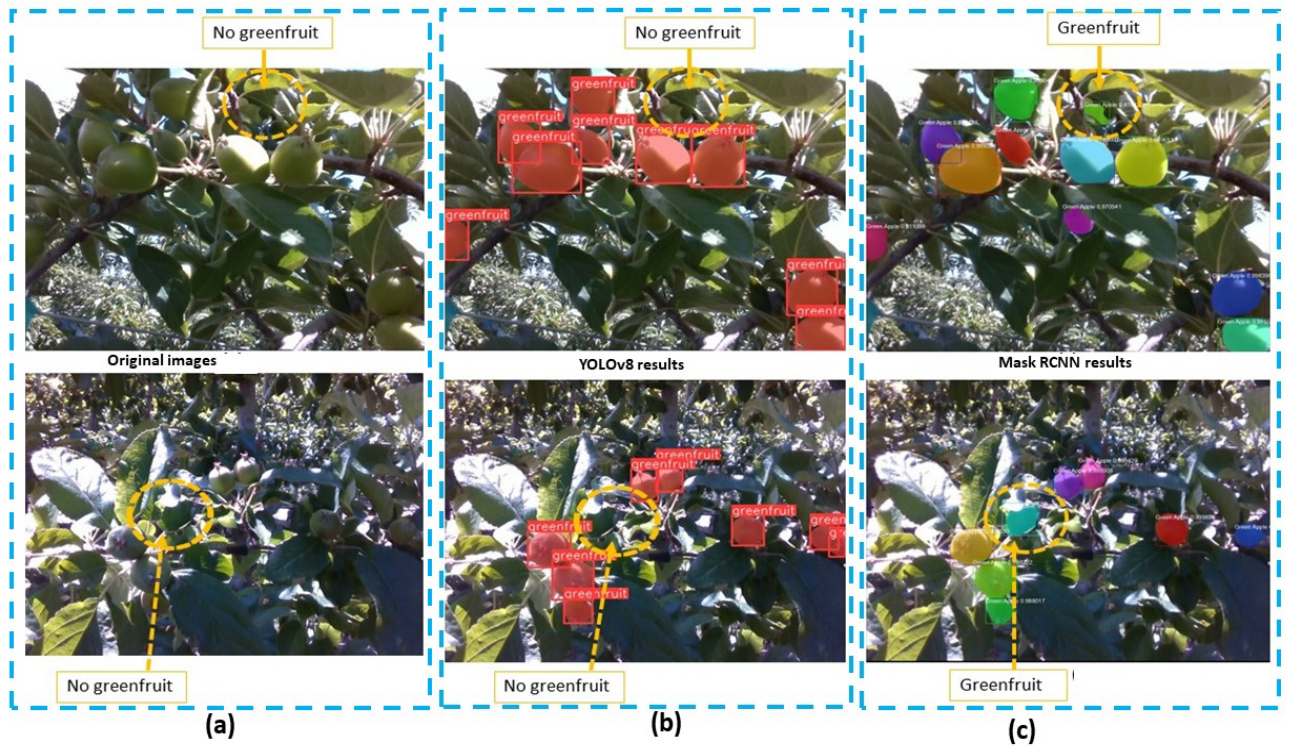


Figure 8: Example images showing the performance of two methods in segmenting immature green fruit in orchard condition; (a) Original images; (b) Instance segmentation results of YOLOv8; and (c) Instance segmentation results of Mask R-CNN. It is noted that some problematic regions in the canopy images (yellow circles) were incorrectly segmented as green fruit by Mask R-CNN but were correctly left as background by YOLOv8.

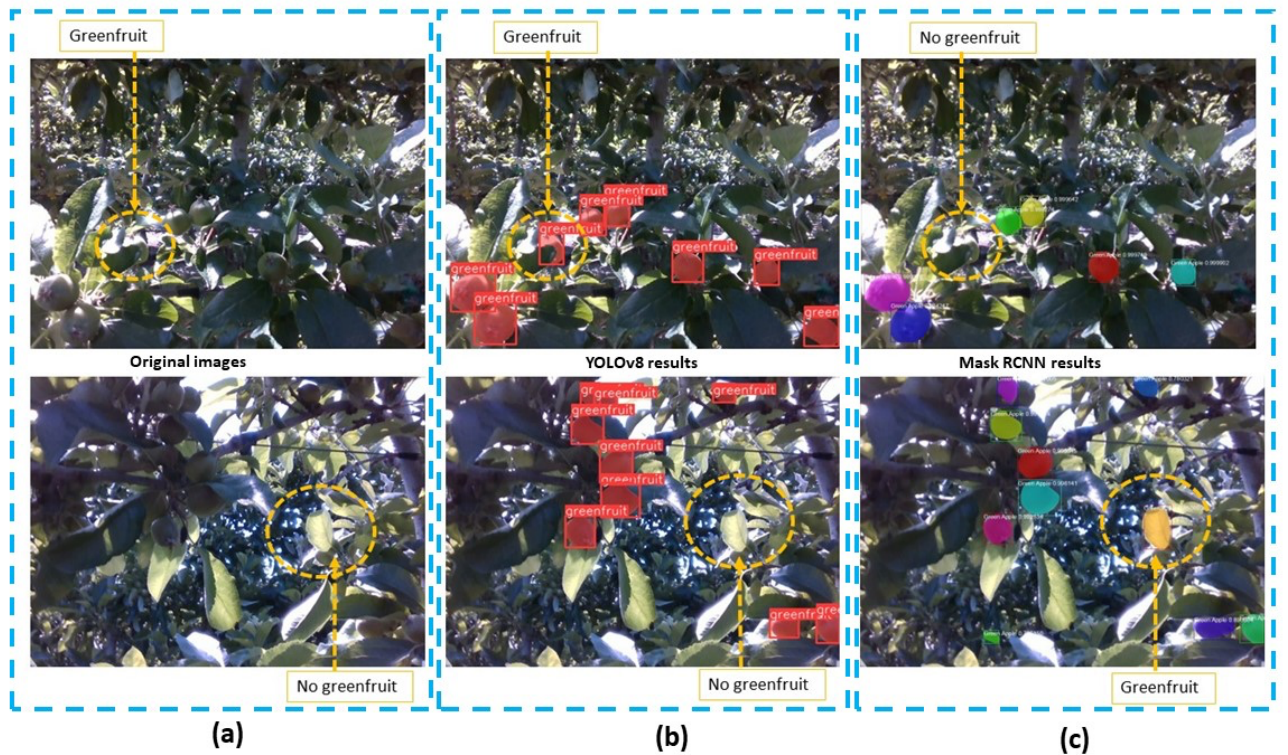


Figure 9: Figure illustrating wrong detection during the growing season orchard condition, yellow region includes the focus area (a) Original image 1; (b) YOLOv8 identification and segmentation; (c) Mask R-CNN failed to identify; (d) Original image2; (e) YOLOv8 segmentations; and (f) Mask-RCNN segmentation where it identified leaf section as a greenfruit

The performance differences between YOLOv8 and Mask R-CNN generally reflected the distinct nature of their architectures and the way they process images. YOLOv8, being a one-stage detector, is designed for speed and accuracy, making it capable of excluding similar non-target areas, as observed in the segmentation tasks (Figure 8b). Its direct approach to object detection avoids the region proposal step, leading to fewer false positives in areas of the canopy that resembled the target fruit in color. Mask R-CNN, on the other hand, uses a two-stage process, which involved generating region proposals before classifying and segmenting objects. This can sometimes result in the inclusion of non-target areas, such as leaves and stems being misclassified as fruits (Figure 8c). Moreover, its performance appears to be more sensitive to lighting variations, which can lead to errors in object identification under the extreme sides of lighting situation such as bright, direct sun-light and dark shadows (Figure 9c). Despite these differences, there are specific situations where Mask R-CNN could still be the preferred choice. Its two-stage process, particularly the region proposal step, can be advantageous in complex segmentation tasks where precision is critical, and objects are densely packed or partially obscured. In the past, green fruit segmentation has been investigated using various approaches. Wei et al.'s D2D framework [80], GHFormer's focus on night-time detection [81], Liu et al.'s FCOS model for obscured fruits [82], Jia et al.'s ResNet-based FoveaMask [83], and Sun et al.'s combination of GrabCut and Ncut algorithms [84] each offered solutions to specific segmentation challenges such as lower accuracy and higher computation cost. Some studies also explored semi-automated models [85]. However, the performance of the YOLOv8 model in this study exceeded those of the reviewed past studies. Likewise, the performance of the Mask R-CNN model in segmenting immature green fruits, while not as high as YOLOv8's, still surpassed many recent approaches [70], [77], [81], [82], [85], [86].

## 4.2 Multi-class Object Segmentation in Images of Dormant Apple Trees

Similar to single class object segmentation discussed above, YOLOv8 performed better than Mask R-CNN in segmenting dormant apple tree images into multiple object classes (trunks and branches). YOLOv8 achieved a precision of 1.00 at a confidence threshold of 0.906, as shown in Figure 10. Similarly, Figure 11 shows that the recall for YOLOv8 reached 0.95 at the minimal confidence threshold, indicating a high degree of accuracy in segmenting these complex structures of dormant tree canopies. Mask R-CNN reached a precision of 1.00 at a lower confidence threshold of 0.813, suggesting a strong ability to correctly detecting target objects at this level of confidence (Figure 10b). Additionally, the recall of Mask R-CNN, as depicted in Figure 11, achieved 0.837 at the lowest confidence threshold, indicating slightly higher rate of false negatives compared to YOLOv8. Similarly, precision-recall curve (Figure 12a) showed that YOLOv8 achieved a mean average precision (mAP) of 0.845 over all object classes at an intersection over union (IoU) of 0.5, which for the trunk and branch classes were 0.971 and 0.719, respectively. Mask R-CNN achieved relatively lower performance in multi-class segmentation tasks as well. As seen in Figure 12b (precision-recall curve) the model achieved an all-class mAP of 0.748 at an IoU of 0.5, with individual mAP of 0.828 for trunk segmentation and 0.673 for branch segmentation.

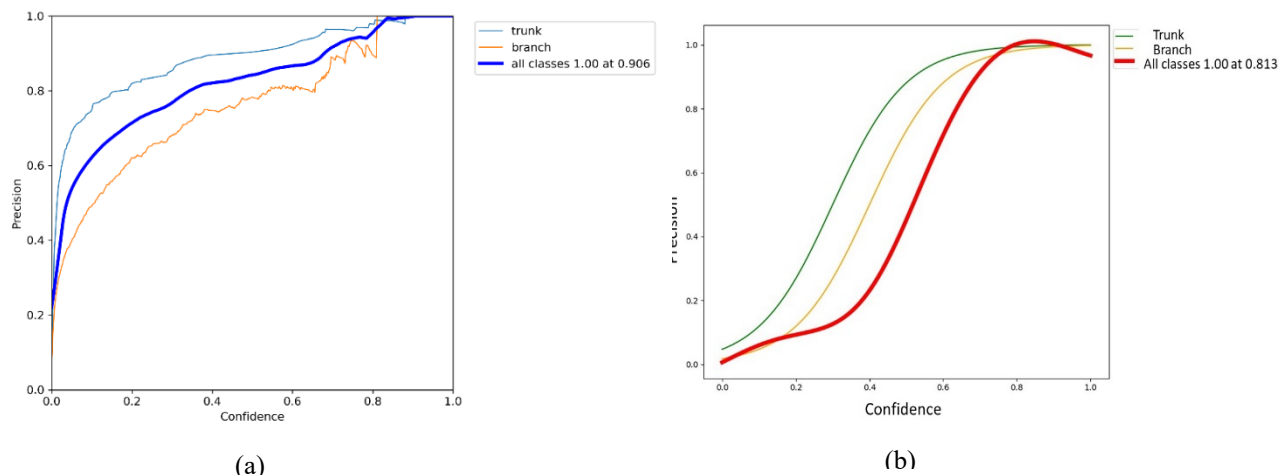


Figure 10: Precision-Confidence curve for multi-class segmentation of Trunk and Branch ; (a) YoloV8 , (b) Mask R-CNN

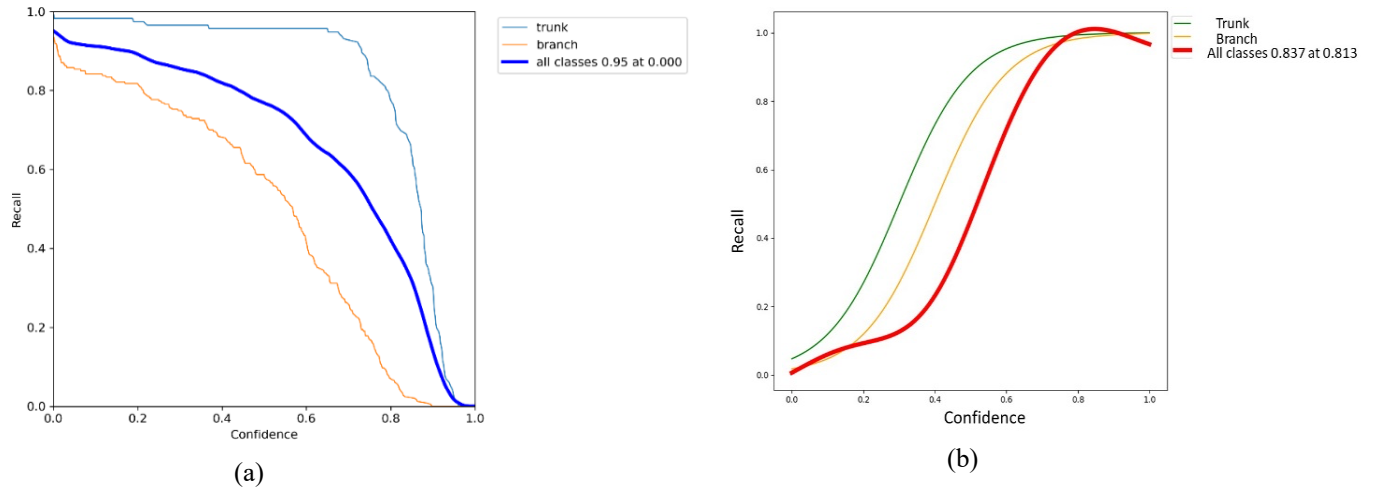


Figure 11: Recall-confidence curve for multi-class segmentation of trunks and branches achieved with; (a) YOLOv8; and (b) Mask R-CNN.

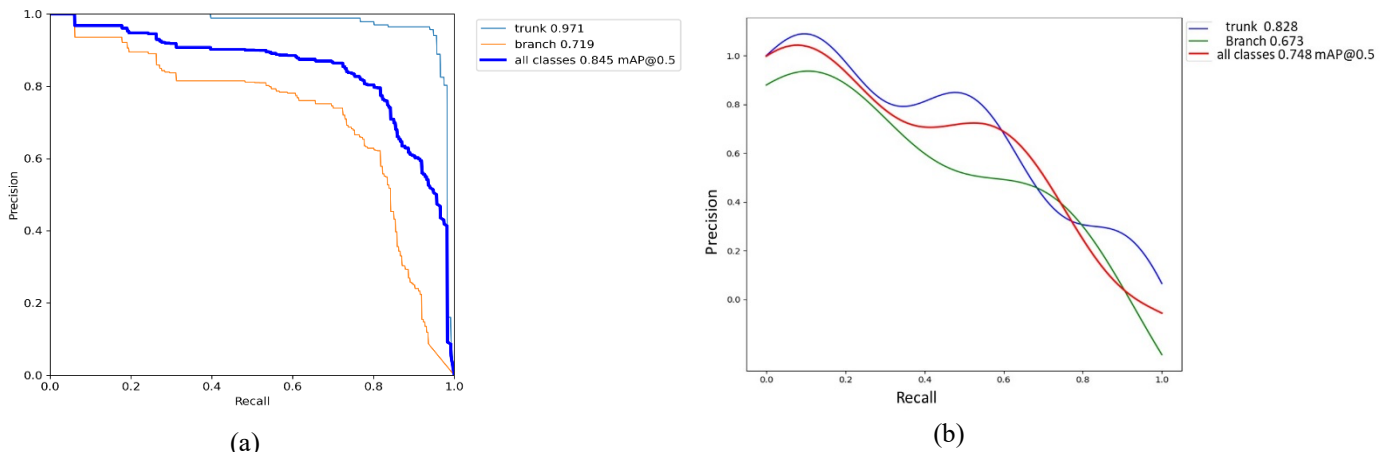
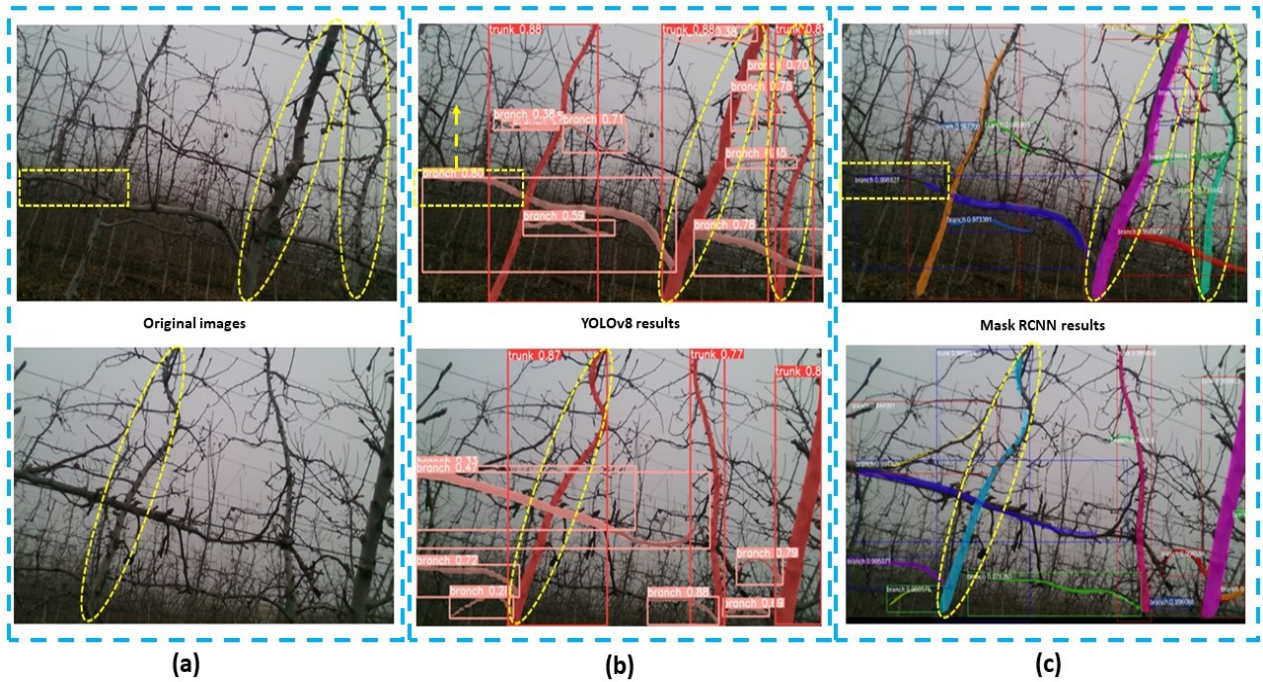


Figure 12: Precision-Recall curve for multi-class segmentation of trunks and branches of dormant apple trees at mAP@0.5; (a) with YOLOv8; and (b) with Mask R-CNN.

Example images demonstrating comparative successes and failures of these models (YOLOv8 and Mask R-CNN) in segmenting trunks and branches are depicted in Figures 13 and 14. As shown before with mAP and other measures, trunks were segmented with higher accuracy by YOLOv8 compared to Mask R-CNN, which are indicated by sample cases shown in shown in Figure 13b and 13c respectively. Specifically, the branch highlighted within the yellow dotted rectangle (Figure 13 a, b and c) was successfully detected by YOLOv8 but not by Mask R-CNN, showing YOLOv8's better performance in low light conditions compared to Mask R-CNN. The example in Figure 13 shows that YOLOv8 was more effective in segmenting trunks. Similarly, Figure 14 presented examples of successful and failed segmentations in both trunk and branches, which showed that YOLOv8 was more precise (less false detection) than Mask R-CNN, particularly in area with challenging lighting and complex backgrounds (e.g., a rectangular box in Figure 14b). Comparatively, Mask R-CNN exhibited lower performance under these conditions, with the limitations being more apparent in poorly lit areas with complex backgrounds (e.g., Figure 14c). The segmentation of the branch within the yellow rectangle (Figure 14d) also highlighted YOLOv8's ability to detect features despite variable lighting conditions created by shadows and hue variations, an area where Mask R-CNN was less robust in segmenting desired objects (Figure 14e).

A number of recent studies focused on the segmentation of tree trunks and branches, employing various deep learning approaches. For example, [87], [88] used deep learning for automatic branch detection in jujube trees and [89] used Regions-Convolutional Neural Networks (R-CNN) models alongside depth features for branch detection in fruiting wall apple trees. Segmenting plant canopy parts in dormant grapevines has also been studied widely using different

deep learning techniques (e.g., [90].[100]). Other models like ViNet [91] have emerged, providing deep learning solutions for estimating grapevine structures. Further advancements include the application of deep learning and geometric constraints for obscured branch segmentation and three-dimensional reconstruction [92], as well as the use of space colonization algorithms for dormant pruning in jujubee plants [88]. A deep learning-based sensing system (called SPGnet) for jujubee plant by Baojian et al. [87], Zhang et al.'s branch detection in apple trees using R-CNN [93], and Lin et al.'s tiny Mask R-CNN for guava branch reconstruction [94] are other recent studies in this field. Additionally, Aguiar et al. [95] explored trunk segmentation using a semantic segmentation-based deep learning approach with a Single Shot Multibox Detector (SSD). In comparison with the performance measures reported by these latest, innovative methodologies available in the literature, YOLOv8 model presented in this study performed better in segmenting tree trunks in terms of both precision (0.95), recall (0.97) and mAP@0.5(0.74). Furthermore, while the Mask R-CNN model achieved relatively lower performance relative to YOLOv8, its performance was comparable or better with many recent studies on trunk and branch detection including [89]–[92], [96], [97].



**Figure 13: Example results for multiclass segmentation of trunks (yellow circle) and branches (yellow rectangle) in dormant season orchard images; (a) Original images ; (b) YOLOv8 segmentation results; and (c) Mask R-CNN segmentation results. This example showed slightly weaker segmentation performance of Mask R-CNN, qualitatively, compared to YOLOv8.**

Computational speed is one of the major performance major of these models, particularly when they are used for real-time field applications such as robotic pruning or thinning. The inference times (processing time per image during testing) required for segmenting green fruit and multi-class objects (trunks and branches) with YOLOv8 and Mask R-CNN models are presented in Table 2. It was found that YOLOv8 took only 7.8 ms to complete single-class segmentation and 10.9 ms for multi-class segmentation per test image using Intel Xeon® W-2155 CPU @ 3.30 GHz x20 processor, NVIDIA TITAN Xp Collector's edition/PCIe/SSE2 graphics card, 31.1 gigabyte memory, and Ubuntu 16.04 LTS 64-bit operating system. These inference times correspond to inference speeds of approximately 128 FPS and 92 FPS, respectively for single and multi-class segmentations. Comparatively, the inference times for Mask R-CNN was higher at 12.8 ms for single class segmentation, which translates to an inference speed of approximately 78 FPS. For multi-class segmentation, the inference time increased to 15.6 ms for Mask R-CNN, or roughly 64 FPS. This difference in processing time showed suitability of the YOLOv8 for both single and multi-object instance segmentation for real-time applications, particularly in agricultural settings where computational resources may be scarce.

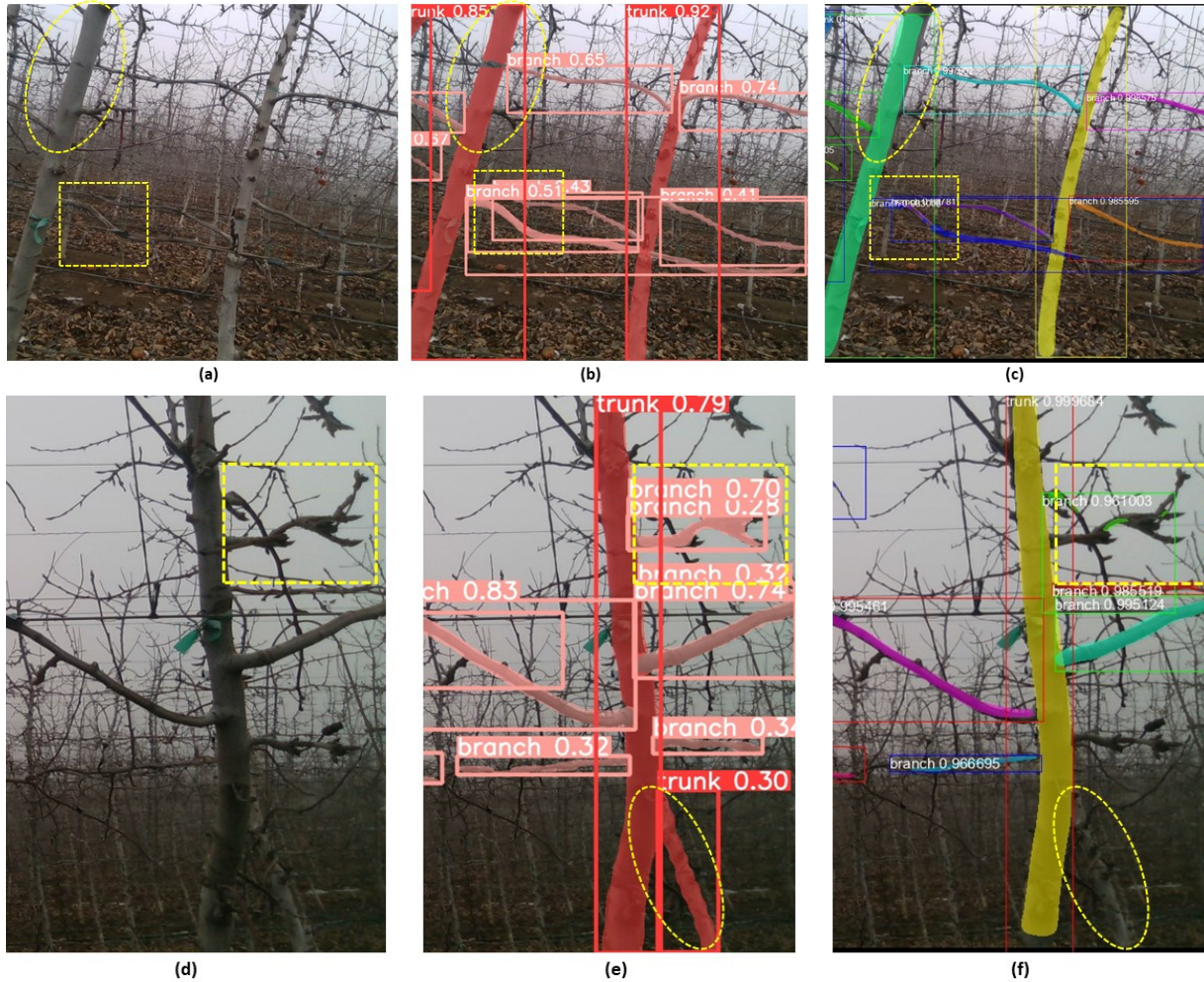


Figure 14: Figures illustrating multiclass segmentation (a) Original Image1 ; (b) YOLOv8 segmentation (c) Mask R-CNN segmentation; (d) Original image 2; (e) YOLOv8 segmentation (f) Mask R-CNN segmentation.

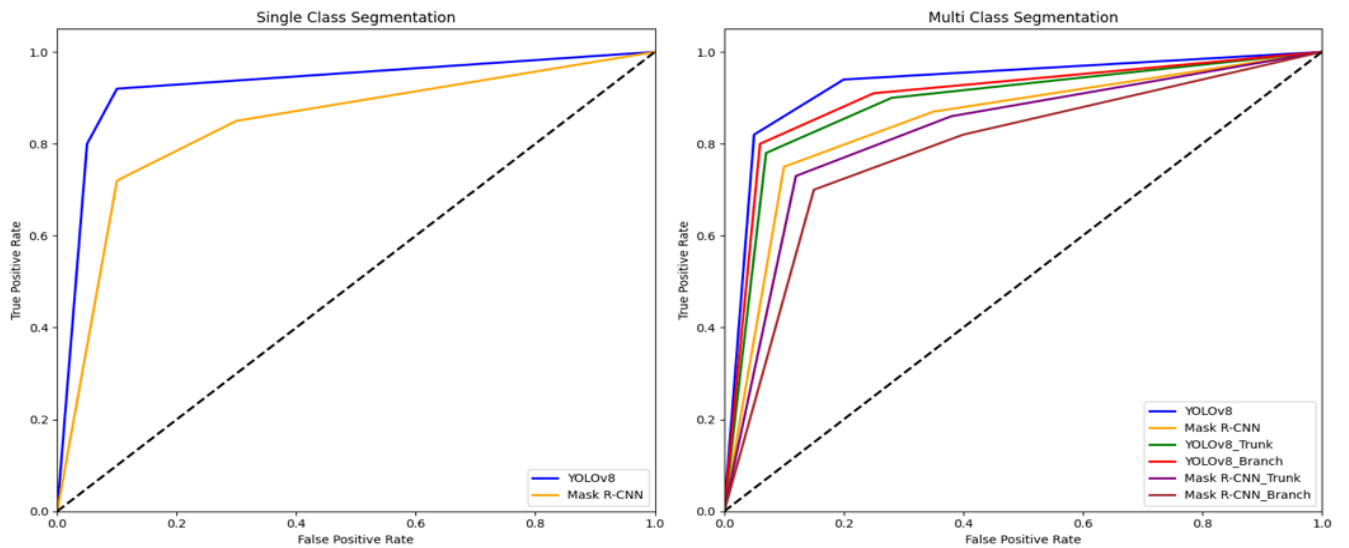


Figure 15: Area under curve (AUC) for the segmentation results of both datasets : (a) Immature green fruit (Apple) dataset; (b) Dormant season orchard dataset

**Table 2: Summary of the performance metrics of YOLOv8 and Mask R-CNN models including precision, recall, mAP@0.5, inference times, and FPS for single and multi-class object segmentation tasks in this study.**

Model	Precision	Recall	mAP@0.5	Inference Time (ms)	Frames Per Second (FPS)
YOLOv8 (Single-class)	92.9	97	0.902	7.8	128.21
Mask R-CNN (Single-class)	84.7	88	0.85	12.8	78.13
YOLOv8 (Multi-class)	90.6	95	0.74	10.9	91.74
Mask R-CNN (Multi-class)	81.3	83.7	0.700	15.6	64.10

According to a most recent study on comparative study of immature green apple detection using machine learning and deep learning models[98], Fully Convolutional One-Stage (FCOS) with a ResNet101 RFPN backbone achieved a precision of 81.2%. SSD, utilizing VGG16, had a precision of 69%, while YOLOv3 with Darknet-53 reached 71.3%. Faster-RCNN and RetinaNet, both employing ResNet101-FPN, achieved precisions of 72.1% and 76.6%, respectively. Lastly, CenterNet, using the Hourglass-104 backbone, recorded a precision of 71.2%. Compared to this result, the precision recorded by both Yolov8 and Mask RCNN (92.9 and 84.7) in this study is higher. Likewise, other recent studies conducted to detect and segment branch in apple trees such as Unet++[99] with an accuracy of 72%. Furthermore, studies[100], [101] also aimed at similar segmentation tasks, yet their outcomes fall short of the results obtained in our study, where YOLOv8 and Mask R-CNN demonstrated higher precision rates of 90.6% and 81.3% respectively for branch and trunk segmentation.

## 5. Conclusion

In recent years, there has been increased research, development and adoption of sensing, precision, automation and robotics technologies in agricultural operations, driven by the need to minimize farming inputs including labor and increasing crop yield and quality. This study, through a comprehensive experiment in commercial orchards, provided comparative performance measures of two latest, and most widely used machine learning or deep-learning models (YOLOv8 and Mask R-CNN) for instance segmentation as it relates to their applicability to various crop monitoring and automated canopy and crop-load management tasks (e.g., automated pruning and immature green fruit thinning). Based on the results found, the following specific conclusions could be made.

1. Both YOLOv8 and Mask R-CNN models can provide practically application segmentation results for apple tree canopy images acquired in dormant and early growing seasons with YOLOv8 achieving slightly better performance particularly under similar color feature (between objects and background) and varying light intensity. For single class immature green fruit segmentation, YOLOv8 achieved a high precision of 0.92 and recall of 0.97, which was slightly lower for Mask R-CNN (0.84 and 0.88 respectively). Similarly, for multi-class trunk and branch detection, YOLOv8 demonstrated superior precision and recall metrics (0.90 and 0.95) contrasting with the slightly lower performance of Mask R-CNN (0.81 and 0.83) in these parameters.
2. YOLOv8, with its faster inference rate of 128.21 FPS for single-class and 91.74 FPS for multi-class segmentation, demonstrates superior suitability for time-sensitive agricultural tasks like automated pruning, particularly in low-light conditions. Conversely, Mask R-CNN's slower inference speed, at 78.13 FPS for single-class and 64.10 FPS for multi-class scenarios, suggests potential constraints in applications requiring rapid response times.

These findings showed that the two models evaluated in this study could be an effective and efficient tool for developing various precision and automated agricultural tools, with potential applications extending to various crops beyond apples, which will play a crucial role in enhancing crop management and improving crop yield and quality through machine learning. Particularly, YOLOv8 showed good adaptability across different orchard conditions, which is a critical benefit in advancing robust machine learning-based solutions for future innovations in smart farming. The incorporation of machine learning is a key to meet global agricultural sustainability and food security needs.

## Acknowledgement

This research is funded by the National Science Foundation and United States Department of Agriculture, National

Institute of Food and Agriculture through the “AI Institute for Agriculture” Program (Award No.AWD003473). The authors gratefully acknowledge Dave Allan (Allan Bros., Inc.) for providing access to the orchards during the data collection and field evaluation.

### Author’s Contribution

**Ranjan Sapkota:** Conceptualization, Investigation, Visualization, Methodology, Writing – original draft, review & editing. **Dawood Ahmed:** Investigation, Visualization & Methodology. **Manoj Karkee:** Supervision, writing – review & editing.

**Declaration of generative AI and AI-assisted technologies in the writing process:** During the preparation of this work, the author used chatGPT in order to correct grammar and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

### REFERENCES

- [1] A. M. Hafiz and G. M. Bhat, ‘A survey on instance segmentation: state of the art’, *Int J Multimed Inf Retr*, vol. 9, no. 3, pp. 171–189, 2020.
- [2] Q. Zhang, Y. Liu, C. Gong, Y. Chen, and H. Yu, ‘Applications of deep learning for dense scenes analysis in agriculture: A review’, *Sensors*, vol. 20, no. 5, p. 1520, 2020.
- [3] J. Champ, A. Mora-Fallas, H. Goëau, E. Mata-Montero, P. Bonnet, and A. Joly, ‘Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots’, *Appl Plant Sci*, vol. 8, no. 7, p. e11373, 2020.
- [4] Y. Chen, S. Baireddy, E. Cai, C. Yang, and E. J. Delp, ‘Leaf segmentation by functional modeling’, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, p. 0.
- [5] N. Lüling, D. Reiser, and H. W. Griepentrog, ‘Volume and leaf area calculation of cabbage with a neural network-based instance segmentation’, in *Precision agriculture ’21*, Wageningen Academic Publishers, 2021, pp. 2719–2745.
- [6] C. Niu, H. Li, Y. Niu, Z. Zhou, Y. Bu, and W. Zheng, ‘Segmentation of cotton leaves based on improved watershed algorithm’, in *Computer and Computing Technologies in Agriculture IX: 9th IFIP WG 5.14 International Conference, CCTA 2015, Beijing, China, September 27-30, 2015, Revised Selected Papers, Part I 9*, Springer, 2016, pp. 425–436.
- [7] V. H. Pham and B. R. Lee, ‘An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm’, *Vietnam Journal of Computer Science*, vol. 2, pp. 25–33, 2015.
- [8] M. G. S. Jayanthi and D. R. Shashikumar, ‘Leaf disease segmentation from agricultural images via hybridization of active contour model and OFA’, *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 35–52, 2017.
- [9] J. Clement, N. Novas, J.-A. Gazquez, and F. Manzano-Agugliaro, ‘An active contour computer algorithm for the classification of cucumbers’, *Comput Electron Agric*, vol. 92, pp. 75–81, 2013.
- [10] Y. A. N. Gao, J. F. Mas, N. Kerle, and J. A. Navarrete Pacheco, ‘Optimal region growing segmentation and its effect on classification accuracy’, *Int J Remote Sens*, vol. 32, no. 13, pp. 3747–3763, 2011.
- [11] N. Jothiaruna, K. J. A. Sundar, and B. Karthikeyan, ‘A segmentation method for disease spot images incorporating chrominance in comprehensive color feature and region growing’, *Comput Electron Agric*, vol. 165, p. 104934, 2019.
- [12] J. Ma, K. Du, L. Zhang, F. Zheng, J. Chu, and Z. Sun, ‘A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing’, *Comput Electron Agric*, vol. 142, pp. 110–117, 2017.
- [13] V. Gupta, N. Sengar, M. K. Dutta, C. M. Travieso, and J. B. Alonso, ‘Automated segmentation of powdery mildew disease from cherry leaves using image processing’, in *2017 International Conference and Workshop on Bioinspired Intelligence (IWOB)*, IEEE, 2017, pp. 1–4.
- [14] S. D. Khirade and A. B. Patil, ‘Plant disease detection using image processing’, in *2015 International conference on computing communication control and automation*, IEEE, 2015, pp. 768–771.
- [15] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, ‘Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm’, *Comput Electron Agric*, vol. 165, p. 104962, 2019.
- [16] T. Arsan and M. M. N. Hameez, ‘A clustering-based approach for improving the accuracy of UWB sensor-

- based indoor positioning system', *Mobile Information Systems*, vol. 2019, pp. 1–13, 2019.
- [17] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, 'Recent advances in image processing techniques for automated leaf pest and disease recognition—A review', *Information processing in agriculture*, vol. 8, no. 1, pp. 27–51, 2021.
- [18] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, 'Deep learning for precision agriculture: A bibliometric analysis', *Intelligent Systems with Applications*, vol. 16, p. 200102, 2022.
- [19] S. A. Singh and K. A. Desai, 'Automated surface defect detection framework using machine vision and convolutional neural networks', *J Intell Manuf*, vol. 34, no. 4, pp. 1995–2011, 2023.
- [20] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, 'U-net and its variants for medical image segmentation: A review of theory and applications', *Ieee Access*, vol. 9, pp. 82031–82057, 2021.
- [21] K. He, G. Gkioxari, P. Dollár, and R. Girshick, 'Mask r-cnn', in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, 'You only look once: Unified, real-time object detection', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [23] J. Rashid, I. Khan, G. Ali, F. Alturise, and T. Alkhalifah, 'Real-Time Multiple Guava Leaf Disease Detection from a Single Leaf Using Hybrid Deep Learning Technique.', *Computers, Materials & Continua*, vol. 74, no. 1, 2023.
- [24] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, 'Instance segmentation of apple flowers using the improved mask R-CNN model', *Biosyst Eng*, vol. 193, pp. 264–278, 2020.
- [25] A. K. Maji, S. Marwaha, S. Kumar, A. Arora, V. Chinnusamy, and S. Islam, 'SlypNet: Spikelet-based yield prediction of wheat using advanced plant phenotyping and computer vision techniques', *Front Plant Sci*, vol. 13, p. 889853, 2022.
- [26] J. Liu and X. Wang, 'Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network', *Front Plant Sci*, vol. 11, p. 898, 2020.
- [27] M. Lippi, N. Bonucci, R. F. Carpio, M. Contarini, S. Speranza, and A. Gasparri, 'A yolo-based pest detection system for precision agriculture', in *2021 29th Mediterranean Conference on Control and Automation (MED)*, IEEE, 2021, pp. 342–347.
- [28] X. Qu, J. Wang, X. Wang, Y. Hu, T. Zeng, and T. Tan, 'Gravelly soil uniformity identification based on the optimized Mask R-CNN model', *Expert Syst Appl*, vol. 212, p. 118837, 2023.
- [29] L. Zu, Y. Zhao, J. Liu, F. Su, Y. Zhang, and P. Liu, 'Detection and segmentation of mature green tomatoes based on mask R-CNN with automatic image acquisition approach', *Sensors*, vol. 21, no. 23, p. 7842, 2021.
- [30] Q. Wang, M. Cheng, S. Huang, Z. Cai, J. Zhang, and H. Yuan, 'A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed *Solanum rostratum* Dunal seedlings', *Comput Electron Agric*, vol. 199, p. 107194, 2022.
- [31] H. Li *et al.*, 'Design of field real-time target spraying system based on improved YOLOv5', *Front Plant Sci*, vol. 13, p. 1072631, 2022.
- [32] C. Hu, J. A. Thomasson, and M. V. Bagavathiannan, 'A powerful image synthesis and semi-supervised learning pipeline for site-specific weed detection', *Comput Electron Agric*, vol. 190, p. 106423, 2021.
- [33] S. Chen *et al.*, 'An approach for rice bacterial leaf streak disease segmentation and disease severity estimation', *Agriculture*, vol. 11, no. 5, p. 420, 2021.
- [34] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, 'Instance segmentation of apple flowers using the improved mask R-CNN model', *Biosyst Eng*, vol. 193, pp. 264–278, 2020.
- [35] G. Lin, Y. Tang, X. Zou, and C. Wang, 'Three-dimensional reconstruction of guava fruits and branches using instance segmentation and geometry analysis', *Comput Electron Agric*, vol. 184, p. 106107, 2021.
- [36] K. Jha, A. Doshi, P. Patel, and M. Shah, 'A comprehensive review on automation in agriculture using artificial intelligence', *Artificial Intelligence in Agriculture*, vol. 2, pp. 1–12, 2019.
- [37] A. You *et al.*, 'Semiautonomous Precision Pruning of Upright Fruiting Offshoot Orchard Systems: An Integrated Approach', *IEEE Robot Autom Mag*, 2023.
- [38] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng, 'Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot', *Comput Electron Agric*, vol. 172, p. 105380, 2020.
- [39] Y. Yu, K. Zhang, L. Yang, and D. Zhang, 'Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN', *Comput Electron Agric*, vol. 163, p. 104846, 2019.
- [40] L. Zu, Y. Zhao, J. Liu, F. Su, Y. Zhang, and P. Liu, 'Detection and segmentation of mature green tomatoes based on mask R-CNN with automatic image acquisition approach', *Sensors*, vol. 21, no. 23, p. 7842, 2021.



- [41] S. R. Khanal, R. Sapkota, D. Ahmed, U. Bhattarai, and M. Karkee, 'Machine Vision System for Early-stage Apple Flowers and Flower Clusters Detection for Precision Thinning and Pollination', *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 8914–8919, 2023, doi: <https://doi.org/10.1016/j.ifacol.2023.10.096>.
- [42] R. Sapkota *et al.*, 'Robotic Pollination of Apples in Commercial Orchards', *arXiv preprint arXiv:2311.10755*, 2023.
- [43] D. A. Ranjan Sapkota, M. Churuvija, and M. Karkee, 'Immature Green Apple Detection and Sizing in Commercial Orchards using YOLOv8 and Shape Fitting Techniques'.
- [44] S. Xie, C. Hu, M. Bagavathiannan, and D. Song, 'Toward robotic weed control: detection of nutsedge weed in bermudagrass turf using inaccurate and insufficient training data', *IEEE Robot Autom Lett*, vol. 6, no. 4, pp. 7365–7372, 2021.
- [45] J. Champ, A. Mora-Fallas, H. Goëau, E. Mata-Montero, P. Bonnet, and A. Joly, 'Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots', *Appl Plant Sci*, vol. 8, no. 7, p. e11373, 2020.
- [46] K. He, G. Gkioxari, P. Dollár, and R. Girshick, 'Mask r-cnn', in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [47] S. Wang, G. Sun, B. Zheng, and Y. Du, 'A crop image segmentation and extraction algorithm based on Mask RCNN', *Entropy*, vol. 23, no. 9, p. 1160, 2021.
- [48] P. Ganesh, K. Volle, T. F. Burks, and S. S. Mehta, 'Deep orange: Mask R-CNN based orange detection and segmentation', *IFAC-PapersOnLine*, vol. 52, no. 30, pp. 70–75, 2019.
- [49] U. Afzaal, B. Bhattarai, Y. R. Pandeya, and J. Lee, 'An instance segmentation model for strawberry diseases based on mask R-CNN', *Sensors*, vol. 21, no. 19, p. 6565, 2021.
- [50] T.-L. Lin, H.-Y. Chang, and K.-H. Chen, 'The pest and disease identification in the growth of sweet peppers using faster R-CNN and mask R-CNN', *Journal of Internet Technology*, vol. 21, no. 2, pp. 605–614, 2020.
- [51] Z. U. Rehman *et al.*, 'Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture', *IET Image Process*, vol. 15, no. 10, pp. 2157–2168, 2021.
- [52] G. H. Krishnan and T. Rajasenbagam, 'A Comprehensive Survey for Weed Classification and Detection in Agriculture Lands', *Journal of Information Technology*, vol. 3, no. 4, pp. 281–289, 2021.
- [53] K. Osorio, A. Puerto, C. Pedraza, D. Jamaica, and L. Rodríguez, 'A deep learning approach for weed detection in lettuce crops using multispectral images', *AgriEngineering*, vol. 2, no. 3, pp. 471–488, 2020.
- [54] T. Zhao, Y. Yang, H. Niu, D. Wang, and Y. Chen, 'Comparing U-Net convolutional network with mask R-CNN in the performances of pomegranate tree canopy segmentation', in *Multispectral, hyperspectral, and ultraspectral remote sensing technology, techniques and applications VII*, SPIE, 2018, pp. 210–218.
- [55] A. Safonova, E. Guirado, Y. Maglinets, D. Alcaraz-Segura, and S. Tabik, 'Olive tree biovolume from UAV multi-resolution image segmentation with mask R-CNN', *Sensors*, vol. 21, no. 5, p. 1617, 2021.
- [56] Y. Changhui, W. Zhuo, X. Longye, L. I. U. Yanping, K. Xilong, and Z. Wanhua, 'Identification and reconstruction of citrus branches under complex background based on Mask R-CNN', *Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery*, vol. 50, no. 8, 2019.
- [57] P. Soviany and R. T. Ionescu, 'Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction', in *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, IEEE, 2018, pp. 209–214.
- [58] A. You *et al.*, 'Semiautonomous Precision Pruning of Upright Fruiting Offshoot Orchard Systems: An Integrated Approach', *IEEE Robot Autom Mag*, 2023.
- [59] M. Hussain, L. He, J. Schupp, D. Lyons, and P. Heinemann, 'Green fruit segmentation and orientation estimation for robotic green fruit thinning of apples', *Comput Electron Agric*, vol. 207, p. 107734, 2023.
- [60] J. Seol, J. Kim, and H. Il Son, 'Field evaluations of a deep learning-based intelligent spraying robot with flow control for pear orchards', *Precis Agric*, vol. 23, no. 2, pp. 712–732, 2022.
- [61] L. Wu, J. Ma, Y. Zhao, and H. Liu, 'Apple detection in complex scene using the improved YOLOv4 model', *Agronomy*, vol. 11, no. 3, p. 476, 2021.
- [62] W. Chen, J. Zhang, B. Guo, Q. Wei, and Z. Zhu, 'An apple detection method based on des-YOLO v4 algorithm for harvesting robots in complex environment', *Math Probl Eng*, vol. 2021, pp. 1–12, 2021.
- [63] Z. Huang, P. Zhang, R. Liu, and D. Li, 'Immature apple detection method based on improved Yolov3', *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.
- [64] Y. Liu, G. Yang, Y. Huang, and Y. Yin, 'SE-Mask R-CNN: An improved Mask R-CNN for apple detection and segmentation', *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 6, pp. 6715–6725, 2021.
- [65] A. Kuznetsova, T. Maleva, and V. Soloviev, 'YOLOv5 versus YOLOv3 for apple detection', in *Cyber-*

- Physical Systems: Modelling and Intelligent Control*, Springer, 2021, pp. 349–358.
- [66] D. Wang and D. He, ‘Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning’, *Biosyst Eng*, vol. 210, pp. 271–281, 2021.
- [67] S. Tong, Y. Yue, W. Li, Y. Wang, F. Kang, and C. Feng, ‘Branch Identification and Junction Points Location for Apple Trees Based on Deep Learning’, *Remote Sens (Basel)*, vol. 14, no. 18, p. 4495, 2022.
- [68] F. Gao *et al.*, ‘A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard’, *Comput Electron Agric*, vol. 197, p. 107000, 2022.
- [69] C. Zhang, F. Kang, and Y. Wang, ‘An improved apple object detection method based on lightweight YOLOv4 in complex backgrounds’, *Remote Sens (Basel)*, vol. 14, no. 17, p. 4150, 2022.
- [70] S. Lu, W. Chen, X. Zhang, and M. Karkee, ‘Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation’, *Comput Electron Agric*, vol. 193, p. 106696, 2022.
- [71] J. Lv *et al.*, ‘A visual identification method for the apple growth forms in the orchard’, *Comput Electron Agric*, vol. 197, p. 106954, 2022.
- [72] F. Su *et al.*, ‘Tree Trunk and Obstacle Detection in Apple Orchard Based on Improved YOLOv5s Model’, *Agronomy*, vol. 12, no. 10, p. 2427, 2022.
- [73] D. Wang and D. He, ‘Fusion of Mask RCNN and attention mechanism for instance segmentation of apples under complex background’, *Comput Electron Agric*, vol. 196, p. 106864, 2022.
- [74] W. Jia *et al.*, ‘Accurate segmentation of green fruit based on optimized mask RCNN application in complex orchard’, *Front Plant Sci*, vol. 13, p. 955256, 2022.
- [75] P. Cong, J. Zhou, S. Li, K. Lv, and H. Feng, ‘Citrus Tree Crown Segmentation of Orchard Spraying Robot Based on RGB-D Image and Improved Mask R-CNN’, *Applied Sciences*, vol. 13, no. 1, p. 164, 2022.
- [76] C. Zhang *et al.*, ‘Multi-species individual tree segmentation and identification based on improved mask R-CNN and UAV imagery in mixed forests’, *Remote Sens (Basel)*, vol. 14, no. 4, p. 874, 2022.
- [77] M. Karthikeyan, T. S. Subashini, R. Srinivasan, C. Santhanakrishnan, and A. Ahilan, ‘YOLOAPPLE: Augment Yolov3 deep learning algorithm for apple fruit quality detection’, *Signal Image Video Process*, pp. 1–10, 2023.
- [78] D. Ahmed, R. Sapkota, M. Churuvija, and M. Karkee, ‘Machine Vision-Based Crop-Load Estimation Using YOLOv8’, *arXiv preprint arXiv:2304.13282*, 2023.
- [79] L. Ma, L. Zhao, Z. Wang, J. Zhang, and G. Chen, ‘Detection and Counting of Small Target Apples under Complicated Environments by Using Improved YOLOv7-tiny’, *Agronomy*, vol. 13, no. 5, p. 1419, 2023.
- [80] J. Wei, Y. Ding, J. Liu, M. Z. Ullah, X. Yin, and W. Jia, ‘Novel green-fruit detection algorithm based on D2D framework’, *International Journal of Agricultural and Biological Engineering*, vol. 15, no. 1, pp. 251–259, 2022.
- [81] M. Sun, L. Xu, R. Luo, Y. Lu, and W. Jia, ‘GHFormer-Net: Towards more accurate small green apple/begonia fruit detection in the nighttime’, *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 4421–4432, 2022.
- [82] M. Liu, W. Jia, Z. Wang, Y. Niu, X. Yang, and C. Ruan, ‘An accurate detection and segmentation model of obscured green fruits’, *Comput Electron Agric*, vol. 197, p. 106984, 2022.
- [83] W. Jia *et al.*, ‘FoveaMask: A fast and accurate deep learning model for green fruit instance segmentation’, *Comput Electron Agric*, vol. 191, p. 106488, 2021.
- [84] S. Sun, M. Jiang, D. He, Y. Long, and H. Song, ‘Recognition of green apples in an orchard environment by combining the GrabCut model and Ncut algorithm’, *Biosyst Eng*, vol. 187, pp. 201–213, 2019.
- [85] A. Prabhu and N. S. Rani, ‘Semiautomated Segmentation Model to Extract Fruit Images from Trees’, in *2021 International Conference on Intelligent Technologies (CONIT)*, IEEE, 2021, pp. 1–13.
- [86] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, ‘Apple detection during different growth stages in orchards using the improved YOLO-V3 model’, *Comput Electron Agric*, vol. 157, pp. 417–426, 2019.
- [87] B. Ma, J. Du, L. Wang, H. Jiang, and M. Zhou, ‘Automatic branch detection of jujube trees based on 3D reconstruction for dormant pruning using the deep learning-based method’, *Comput Electron Agric*, vol. 190, p. 106484, 2021.
- [88] Y. Fu *et al.*, ‘Skeleton extraction and pruning point identification of jujube tree for dormant pruning using space colonization algorithm’, *Front Plant Sci*, vol. 13, p. 1103794, 2023.
- [89] J. Zhang, L. He, M. Karkee, Q. Zhang, X. Zhang, and Z. Gao, ‘Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN)’, *Comput Electron Agric*, vol. 155, pp. 386–393, 2018.
- [90] P. Guadagna *et al.*, ‘Using deep learning for pruning region detection and plant organ segmentation in

- dormant spur-pruned grapevines’, *Precis Agric*, pp. 1–23, 2023.
- [91] T. Gentilhomme, M. Villamizar, J. Corre, and J.-M. Odobez, ‘Towards smart pruning: ViNet, a deep-learning approach for grapevine structure estimation’, *Comput Electron Agric*, vol. 207, p. 107736, 2023.
- [92] E. Kok, X. Wang, and C. Chen, ‘Obscured tree branches segmentation and 3D reconstruction using deep learning and geometrical constraints’, *Comput Electron Agric*, vol. 210, p. 107884, 2023.
- [93] J. Zhang, L. He, M. Karkee, Q. Zhang, X. Zhang, and Z. Gao, ‘Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN)’, *Comput Electron Agric*, vol. 155, pp. 386–393, 2018.
- [94] G. Lin, Y. Tang, X. Zou, and C. Wang, ‘Three-dimensional reconstruction of guava fruits and branches using instance segmentation and geometry analysis’, *Comput Electron Agric*, vol. 184, p. 106107, 2021.
- [95] A. S. Aguiar *et al.*, ‘Bringing semantics to the vineyard: An approach on deep learning-based vine trunk detection’, *Agriculture*, vol. 11, no. 2, p. 131, 2021.
- [96] S. Tong, Y. Yue, W. Li, Y. Wang, F. Kang, and C. Feng, ‘Branch Identification and Junction Points Location for Apple Trees Based on Deep Learning’, *Remote Sens (Basel)*, vol. 14, no. 18, p. 4495, 2022.
- [97] R. Xiang, M. Zhang, and J. Zhang, ‘Recognition for stems of tomato plants at night based on a hybrid joint neural network’, *Agriculture*, vol. 12, no. 6, p. 743, 2022.
- [98] M. Liu, W. Jia, Z. Wang, Y. Niu, X. Yang, and C. Ruan, ‘An accurate detection and segmentation model of obscured green fruits’, *Comput Electron Agric*, vol. 197, p. 106984, 2022.
- [99] E. Kok, X. Wang, and C. Chen, ‘Obscured tree branches segmentation and 3D reconstruction using deep learning and geometrical constraints’, *Comput Electron Agric*, vol. 210, p. 107884, 2023.
- [100] D.-H. Kim, C.-U. Ko, D.-G. Kim, J.-T. Kang, J.-M. Park, and H.-J. Cho, ‘Automated Segmentation of Individual Tree Structures Using Deep Learning over LiDAR Point Cloud Data’, *Forests*, vol. 14, no. 6, p. 1159, 2023.
- [101] S. Tong, Y. Yue, W. Li, Y. Wang, F. Kang, and C. Feng, ‘Branch Identification and Junction Points Location for Apple Trees Based on Deep Learning’, *Remote Sens (Basel)*, vol. 14, no. 18, p. 4495, 2022.