

**EFFECTS OF IMAGING ANGLE AND FIELD OF VIEW ON DETECTION  
AND TRACKING OF BAREROOT LOBLOLLY PINE SEEDLINGS USING  
COMPUTER VISION**

by

Ashish Reddy Mulaka

A thesis submitted to the Faculty of the University of Delaware in partial  
fulfillment of the requirements for the degree of Master of Science in Data Science

Winter 2025

© 2025 Ashish Reddy Mulaka  
All Rights Reserved

**EFFECTS OF IMAGING ANGLE AND FIELD OF VIEW ON DETECTION  
AND TRACKING OF BAREROOT LOBLOLLY PINE SEEDLINGS USING  
COMPUTER VISION**

by

Ashish Reddy Mulaka

Approved: \_\_\_\_\_  
Yin Bao, Ph.D.  
Professor in charge of thesis on behalf of the Advisory Committee

Approved: \_\_\_\_\_  
Tobin Driscoll, Ph.D.  
Director of the Data Science Program

Approved: \_\_\_\_\_  
Louis F. Rossi, Ph.D.  
Vice Provost for Graduate and Professional Education and  
Dean of the Graduate College

## **ACKNOWLEDGMENTS**

I would like to express my deepest gratitude to Dr. Yin Bao, my mentor and research advisor in the Department of Plant and Soil Sciences at the University of Delaware, for allowing me to do research and providing invaluable guidance throughout this research. His dynamism, vision, sincerity, and motivation have deeply inspired me. He has taught me the methodology to carry out the research and present the research works as clearly as possible. It was a great privilege and honor to work and study with his guidance. I am incredibly grateful for what he has offered me. I would also like to thank him for his friendship, empathy, and support.

I would like to extend my heartfelt thanks to the thesis committee members, Dr. Benjamin E. Bagozzi and Dr. Jarrod O. Miller, for graciously accepting the request to serve on the committee.

I am also grateful to my parents and my brother for their love, prayers, caring, and sacrifices in educating and preparing me for my future. I would also like to extend my gratitude to my roommates for their endless support. Finally, my thanks go to all the people who have supported me in completing the research work directly or indirectly.

## TABLE OF CONTENTS

LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
ABSTRACT .....	x

### Chapter

1	INTRODUCTION .....	1
1.1	Background.....	1
1.2	Problem Statement.....	3
1.3	Objectives .....	4
1.4	Thesis Structure .....	4
2	PINE SEEDLING DETECTION .....	6
2.1	Introduction .....	6
2.2	Literature Review .....	7
2.2.1	<i>Computer Vision in Agriculture</i> .....	7
2.2.1.1	Overview of Object Detection.....	9
2.2.2	<i>Seedling Detection and Counting</i> .....	13
2.2.3	<i>Camera Imaging (Viewing) Angles</i> .....	15
2.3	Methodology.....	16
2.3.1	<i>Data Collection</i> .....	16
2.3.2	<i>Data Preprocessing</i> .....	17
2.3.3	<i>Model Implementation</i> .....	18
2.3.3.1	Model Architectures .....	19
2.3.3.2	Model Training.....	23
2.3.3.3	Training Environments.....	23
2.3.4	<i>Evaluation Metrics</i> .....	24
2.3.4.1	Intersection over Union (IoU) .....	24

2.3.4.2	Precision, Recall and F1-Score.....	25
2.3.4.3	Average Precision (AP).....	27
2.3.4.4	Mean Average Precision (mAP).....	27
2.3.5	<i>Camera Imaging Angles</i> .....	28
2.4	Results .....	30
2.4.1	<i>Latency Comparisons of Different Model Architectures</i> .....	30
2.4.2	<i>Detection Performance</i> .....	31
2.4.3	<i>Effect of Viewing Angles</i> .....	34
2.5	Discussion and Conclusion.....	37
2.5.1	<i>Discussion</i> .....	37
2.5.1.1	Model Performance .....	37
2.5.1.2	Impact of Viewing Angles.....	38
2.5.2	<i>Limitations</i> .....	38
2.5.3	<i>Recommendations</i> .....	39
2.5.4	<i>Conclusion</i> .....	40
3	PINE SEEDLING TRACKING AND COUNTING .....	42
3.1	Introduction .....	42
3.2	Literature Review .....	43
3.2.1	<i>Multi-Object Tracking</i> .....	43
3.2.2	<i>Tracking Algorithms for Seedling Counting</i> .....	44
3.2.2.1	Simple Online Realtime Tracking (SORT) .....	45
3.2.2.2	ByteTrack .....	46
3.2.2.3	BoT-SORT (Boosted Tracking with SORT).....	47
3.2.3	<i>Research Gap in Existing Research</i> .....	48
3.3	Methodology.....	49
3.3.1	<i>Experimental Setup</i> .....	49
3.3.2	<i>Implementation of Tracking Algorithms</i> .....	49
3.3.3	<i>Counting Pine Seedlings</i> .....	51
3.3.4	<i>Impact of Horizontal Viewing Angles with Change in VFoV on Tracking and Counting</i> .....	52
3.3.5	<i>Evaluation Metrics</i> .....	55

3.3.5.1	MOTA (Multi-Object Tracking Accuracy) .....	55
3.3.5.2	IDF1 .....	56
3.3.5.3	MOTP (Multi-Object Tracking Precision) .....	56
3.3.5.4	Additional MOT Metrics .....	57
3.3.6	<i>Counting Accuracy</i> .....	58
3.4	Results .....	58
3.4.1	<i>Multi-Object Tracking Results</i> .....	59
3.4.2	<i>Effect of Imaging Angles and Field of View on Counting Accuracy</i> .....	63
3.5	Discussion and Conclusion.....	65
3.5.1	<i>Discussion</i> .....	65
3.5.1.1	Performance of MOT Algorithms .....	66
3.5.1.2	Effect of Field of View on Tracking and Counting Accuracy .....	67
3.5.2	<i>Limitations and Future Research</i> .....	68
3.5.3	<i>Conclusion</i> .....	68
4	CONCLUSION AND FUTURE WORK.....	70
4.1	Conclusion.....	70
4.2	Implications for Nursery Management.....	71
4.3	Limitations and Challenges .....	71
4.4	Recommendations For Future Research.....	72
	REFERENCES .....	73

## LIST OF TABLES

Table 2.1: Details of RGB video samples recorded at IFCO BARERROOT Nursery and used in this study. ....	17
Table 2.2: Overview of the YOLOv8, YOLOv9, and YOLOv10 model variants, which were trained on the COCO dataset, including 80 pre-trained classes. ....	22
Table 2.3: Latency and similar architectural comparison of the models. ....	30
Table 2.4: Detection metrics of all the models. ....	31
Table 2.5: Pearson Correlation (r) between mean IoU and the number of crops. ....	34
Table 2.6: Statistical test for significance of HVA and VVA against IoU. ....	36
Table 3.1: Ground truth count of seedlings in each video sample used for evaluation. ....	58
Table 3.2: Comparison of optimal detection models by architecture based on MOTA with three tracking algorithms. ....	61

## LIST OF FIGURES

Figure 2.1: Bar chart representations of the number of research papers found in ScienceDirect when the four research keywords mentioned were introduced. ....	8
Figure 2.2: Two-stage and one-stage object detector. ....	10
Figure 2.3: Architecture of YOLO ....	11
Figure 2.4: Detection Transformer [DETR] Architecture ....	12
Figure 2.5: Illustration of different viewing perspectives of the object located at the nadir and at an oblique viewing angle. ....	15
Figure 2.6: YOLOv8 Architecture. ....	20
Figure 2.7: PGI and related network architectures. ....	21
Figure 2.8: Consistent dual assignment for NMS-free training. ....	22
Figure 2.9: Intersection over Union. ....	25
Figure 2.10: Horizontal ( $\theta h$ ) and vertical ( $\theta v$ ) viewing angles of the object from the camera. ....	29
Figure 2.11: Detections by YOLOv10-balanced model (label: object, confidence). ...	32
Figure 2.12: Spatial distribution of IoU scores. ....	33
Figure 2.13: Mean IoU of all the rows in central and bottom segments. ....	34
Figure 2.14: Distribution of IoU scores at different horizontal viewing angles. ....	35
Figure 2.15: Distribution of IoU scores at different vertical viewing angles. ....	36
Figure 3.1: Tracking-by-detection paradigm. ....	44
Figure 3.2: SORT: Simple Online and Realtime Tracking. ....	45
Figure 3.3: ByteTrack: Multi-Object Tracking by associating every detection box. ...	47

Figure 3.4: Overview of BoT-SORT pipeline (CMC and ReID).....	48
Figure 3.5: Horizontal (h) and vertical (v) field of view.....	52
Figure 3.6: Rows (1-8) division based on HVA.....	53
Figure 3.7: Increasing VFoV of a crop row. ....	54
Figure 3.8: Four cases illustrating tracker-to-target assignments.....	55
Figure 3.9: MOTA scores across trackers with varying detection models.....	60
Figure 3.10: BoT-SORT’s performance in the central row across multiple frames. The green points indicate the continuity of the track ID in subsequent frames. ....	62
Figure 3.11: BoT-SORT is switching ID from (a) to (c) after being lost in (b) in consecutive frames, where red color denotes the switching or re- assigning of a track ID, green denotes maintaining continuity, and black refers to a missing track. ....	62
Figure 3.12: ByteTrack loses tracks in (a) and (b) and assigns an existing ID to a new track in (c) in consecutive frames, where red color denotes the switching or re-assigning of a track ID, green denotes maintaining continuity, and black refers to a missing track. ....	63
Figure 3.13: SORT switching and re-assigning IDs in consecutive frames, where red color denotes the switching or re-assigning of a track ID, green denotes maintaining continuity, and black refers to a missing track.....	63
Figure 3.14: Counting accuracy at various VFoV across 8 crop rows using YOLOv9medium and the three trackers. ....	64
Figure 3.15: Counting accuracy at various VFoV across 8 crop rows using YOLOv8nano and the three trackers. ....	65

## **ABSTRACT**

Developing an automated inventory system for forest nurseries is essential for efficiently counting bareroot loblolly pine seedlings, which is crucial for effective nursery management and large-scale reforestation efforts. Traditional manual sampling-based counting methods are labor-intensive, time-consuming, and error-prone for production field operations. This study aimed to develop an automated inventory system during spring, using a tracking-by-detection paradigm to detect, track, and count seedlings, focusing on how camera imaging angles and field of view affect detection and counting accuracy in RGB imagery. Detection models were trained on a robust dataset of 480 images collected from a commercial forest nursery, and performance was assessed on a test dataset containing eight unique videos comprising 480 frames. The fine-tuned YOLOv10-balanced detection model achieved a mean Average Precision of 95.5%, and the BoT-SORT tracking algorithm reached a Multi-Object Tracking Accuracy of 85.33%. Notably, seedlings in peripheral regions exhibited a significant decrease in detection accuracy compared to the center ones. The results demonstrate that detection accuracy is maximized at the nadir and decreases as the horizontal and vertical viewing angles become more oblique. An increase in the vertical field of view improved counting accuracy up to a certain point, after which it plateaued. Conversely, increasing the horizontal viewing angle resulted in decreased counting accuracy due to factors like viewing perspectives. This study provides valuable insights into selecting the optimal camera field of view and imaging angles for the automated counting of pine seedlings during spring inventory.

# Chapter 1

## INTRODUCTION

### 1.1 Background

Forests are indispensable ecosystems covering vast land areas, predominantly populated with trees and undergrowth. They act as major carbon sinks by absorbing carbon dioxide (CO<sub>2</sub>) from the atmosphere, thereby reducing greenhouse gas concentrations and helping to regulate the global climate (Pan et al., 2011). Additionally, forests preserve biodiversity by providing habitats for countless species and maintaining water cycles and soil health. Economically and socially, forests supply timber and non-timber products, support the livelihoods of millions of people, especially Indigenous communities, and offer recreational and tourism opportunities that enhance human well-being (Fargione et al., 2021). Implementing effective reforestation strategies can ensure the sustainability and resilience of forests for future generations.

In reforestation programs, tree species are selected based on ecological suitability, growth rates, and economic value. In the United States, forest nurseries produced more than 1.4 billion tree seedlings, with 70% of those produced as bare-root stock (Carolyn et al., 2023). To achieve the ambitious goal of reforesting 131 million acres by 2030, the U.S. would need an additional 1.7 billion seedlings each year, a 2.4-fold increase over current nursery production levels. To meet the substantial demand for reforestation, the Southeastern U.S. alone produces over a

billion seedlings annually, with Alabama as one of the leading producers, contributing approximately 91 million seedlings (Fargione et al., 2021).

Pine species, particularly in the United States, play a prominent role due to their rapid growth, adaptability, and commercial importance in the timber industry. Loblolly pine (*Pinus taeda*) is extensively planted in southern U.S. forests, accounting for 84% (16 million hectares) of the total planted forests (Oswalt et al., 2019). Recent advancements in seedling genetics have enhanced seedling value, with prices ranging from \$50 to \$435 per thousand seedlings, depending on the pollination techniques used (Rousseau et al., 2012). Accurate seedling inventories are crucial for ensuring the quality and availability of seedlings to meet market demand, optimize nursery operations, and support reforestation efforts.

Traditional methods for conducting seedling inventories involve manual sampling-based counting, which is labor-intensive, time-consuming, and prone to human error. A typical forest nursery can have millions of seedlings (Carolyn et al., 2023), making counting every seedling impractical. Consequently, there is a growing interest in developing automated methods for seedling inventory management. Advances in computer vision and deep learning offer promising solutions for automating the detection, tracking, and counting of seedlings in nursery environments (Kamilaris & Prenafeta-Boldú, 2018).

While significant progress has been made in applying computer vision techniques for object detection and counting in agricultural settings, there is a lack of research specifically focused on pine seedling inventory in bareroot forest nurseries. Moreover, the effects of camera imaging angles and fields of view on the accuracy of detection and counting remain underexplored. Understanding how these factors

influence performance is essential for designing effective automated inventory systems.

## **1.2 Problem Statement**

The manual sampling-based counting of bareroot pine seedlings in forest nurseries is labor-intensive, time-consuming, and prone to errors due to the high density of seedlings in a huge forest nursery. Traditional inventories are conducted biannually in spring and fall. The spring inventory helps managers adjust bed densities and assess early losses due to wildlife or environmental factors, while the fall inventory is critical for estimating the number of market-ready seedlings (Wakeley, 1954). With nurseries containing millions of seedlings, manual counting methods are inefficient and impractical.

Adopting automated inventory technologies could significantly reduce labor costs, assess early loss, and improve operational efficiency. This would help nursery farm managers manage the spring seedling inventory effectively. However, a computer vision system capable of accurately detecting, tracking, and counting pine seedlings in nursery environments is needed. Additionally, the influence of camera imaging angles and fields of view on detection and counting accuracy is unexplored, presenting a challenge in designing such systems.

Therefore, this thesis aims to develop and evaluate a computer vision system for automating the counting of pine seedlings in bareroot forest nurseries and to investigate how imaging angles and fields of view affect detection and counting accuracy. By addressing these challenges, the research contributes to the advancement of automated forest nursery management practices.

### **1.3 Objectives**

The primary objective of this thesis is to develop and evaluate a computer vision system for automating the counting of bareroot loblolly pine seedlings for spring inventory with a focus on understanding how camera imaging angles and fields of view affect counting accuracy. The specific objectives are as follows: Objective 1 is to detect pine seedlings using various state-of-the-art (SoTA) object detection models and analyze the detection quality across varying horizontal and vertical viewing angles. This involves testing different models under various conditions to determine their robustness and suitability for practical applications. Objective 2 is to evaluate the counting of pine seedlings by tracking them across video sequences using Multi-Object Tracking (MOT) algorithms and to assess how the camera's field of view affects the counting accuracy of seedling inventory. Objective 3 is to provide a detailed analysis and insights into how the camera's imaging angle and various fields of view influence the counting accuracy of pine seedlings when using different detection model configurations and MOT tracking algorithms.

### **1.4 Thesis Structure**

The thesis is structured as follows: Chapter 2 reviews relevant related work on object detection in computer vision within the agricultural domain. It describes the methodology for implementing and evaluating detection models and assessing the influence of camera imaging angles. The chapter then presents the results and concludes with a discussion of how viewing angles affect the detection accuracy. Chapter 3 explores MOT algorithms and outlines the methodology for integrating tracking with detection. It presents counting results and analyzes how tracking and counting are affected by the camera's field of view. Finally, Chapter 4 summarizes the

key findings of the research and discusses the contributions to the field of automated forestry management. It acknowledges the limitations encountered during the study and suggests directions for future research, including potential improvements to the system and its scalability to other types of seedling applications.

## Chapter 2

### PINE SEEDLING DETECTION

#### 2.1 Introduction

Accurate counting of bareroot loblolly pine seedlings is essential for effective inventory management in forest nurseries during spring. Precise seedling counts enable nursery managers to assess growth rates, predict future yields, and make informed decisions regarding resource allocation and market supply. Traditionally, seedling inventories are conducted manually—a process that is labor-intensive, time-consuming, and prone to human error due to the high density of seedlings and the extensive area of nursery beds.

To address these challenges, an automated counting system is needed to accurately count the bareroot loblolly pine seedlings. Recent advancements in computer vision and deep learning have opened new avenues for developing automated systems capable of accurately detecting and counting objects in complex environments like forest nurseries. Object detection models, particularly those based on convolutional neural networks (CNNs), have demonstrated remarkable performance in various applications within agriculture (Farjon et al., 2023).

However, the effectiveness of these models in detecting pine seedlings in nursery settings has not been thoroughly investigated, especially concerning the influence of imaging angles on the viewing perspective of the seedlings. The camera's imaging angle can significantly affect the appearance of seedlings in captured images, leading to variations in scale, occlusion, and viewing perspective challenges. These

factors can impact the performance of detection models, making it crucial to understand how different viewing angles influence the detection of pine seedlings.

Therefore, this chapter aims to evaluate and compare the performance of various state-of-the-art object detection models in detecting pine seedlings from red-green-blue (RGB) videos, with a specific focus on investigating the influence of camera imaging angles. By analyzing how different imaging angles affect detection accuracy, the research seeks to optimize the deployment of detection systems in real-world nursery environments and improve their robustness.

## **2.2 Literature Review**

### **2.2.1 *Computer Vision in Agriculture***

In recent years, the integration of computer vision technologies into agriculture has revolutionized traditional farming practices. The rise in research papers on computer vision applications can be seen in Figure 2.1. The advent of advanced image processing, imaging sensors, and increased computational power has facilitated the adoption of computer vision techniques for various agricultural tasks, including crop monitoring, disease detection, yield estimation, and automated harvesting (Dhanya et al., 2022).

One of the primary motivations for employing computer vision in agriculture is the need for precision farming. Precision agriculture aims to optimize field-level management concerning crop farming by observing, measuring, and responding to inter and intra-field variability in crops (Monteiro et al., 2021). Computer vision systems contribute to this by providing detailed spatial information about crop

conditions, enabling farmers to make informed decisions on irrigation, fertilization, and pest control (Hu et al., 2022).

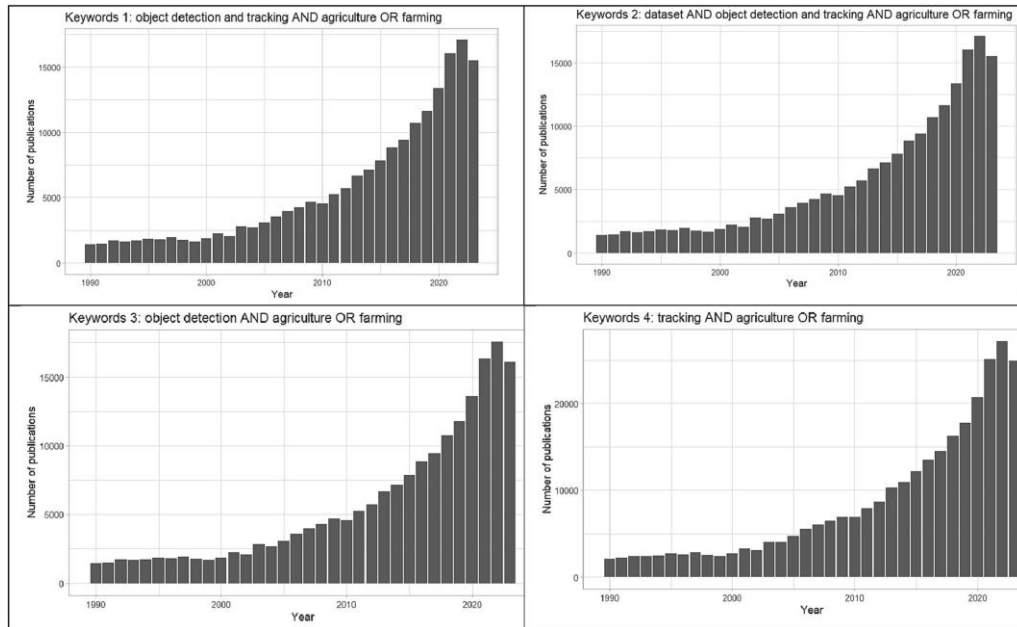


Figure 2.1: Bar chart representations of the number of research papers found in ScienceDirect when the four research keywords mentioned were introduced.

The integration of deep learning with computer vision has further enhanced the capability of these systems to handle complex agricultural environments. Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated remarkable performance in image recognition tasks, making them suitable for applications like plant phenotyping, weed detection, and fruit counting (Kamilaris & Prenafeta-Boldú, 2018). These models' ability to learn hierarchical

feature representations allows for a more accurate and robust analysis of agricultural imagery under varying conditions.

### **2.2.1.1 Overview of Object Detection**

Object detection is a fundamental task in computer vision, involving the identification and localization of objects within an image. It combines image classification and object localization to draw bounding boxes around detected objects and assign them class labels. Significant advancements have been made in object detection algorithms, primarily due to the rise of deep learning techniques (Agarwal et al., 2018).

#### **Two-Stage Detectors**

Early Object Detection methods relied on handcrafted features and traditional machine learning classifiers. The introduction of deep CNNs revolutionized deep learning, leading to models such as the Region-based Convolutional Neural Network (R-CNN) (Girshick et al., 2013). Subsequent improvements led to Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015) and R-FCN (Dai et al., 2016), which introduced Region Proposal Networks for accurate detection. For these methods, two-stage detectors generate region proposals in the first stage and classify objects in the second stage.

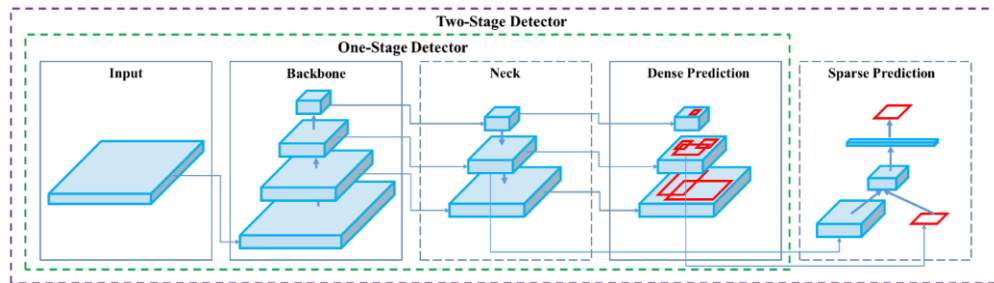


Figure 2.2: Two-stage and one-stage object detector.

While two-stage detectors established the groundwork for high-accuracy object detection, their complexity, and computational demands often constrained their practical deployment in resource-limited or real-time scenarios. The need for faster inference and simpler pipelines naturally led researchers to explore approaches that could bypass the separate region proposal step altogether. This shift in focus laid the foundation for single-stage detectors, which aim to streamline the detection process and enhance speed without sacrificing too much accuracy. As a result, these models became increasingly attractive for a variety of real-time applications, such as those found in nursery environments.

### Single-Stage Detectors

Single-stage detectors emerged to meet the demand for faster object detection suitable for real-time applications. The concept gained prominence with the introduction of YOLO (You Only Look Once) by Redmon et al. (2015) and SSD (Single Shot MultiBox Detector) by Liu et al. (2015). Single-stage detectors perform object detection, classification, and localization in a single forward pass through the neural networks, eliminating the separate region proposal stage, as shown in Figure 2.2. Their simpler architecture reduces computation time, making them suitable for

real-time applications requiring immediate results, such as autonomous driving and video surveillance, including nursery applications. Figure 2.3 illustrates the architecture of the YOLO model, demonstrating its streamlined approach.

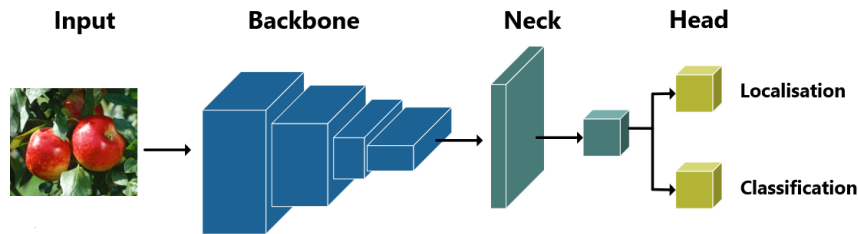


Figure 2.3: Architecture of YOLO

### Anchor-Free Detectors

While single-stage detectors streamlined the detection pipeline by eliminating the need for a separate region proposal stage, most of these methods still relied on predefined anchor boxes. These anchor-based frameworks, despite their speed, introduced additional complexities, such as extensive hyperparameter tuning for anchor sizes, aspect ratios, and scales. To further simplify the process and reduce computational overhead, researchers explored anchor-free approaches. By directly predicting object key points or center points instead of relying on anchor boxes, anchor-free detectors such as CornerNet (Law & Deng, 2018) and FCOS (Fully Convolutional One-Stage Object Detection) (Tian et al., 2019) push the boundaries of real-time detection even further. While still facing challenges with crowded or complex scenes, this anchor-free paradigm represents a natural evolution of single-

stage detection methods, aiming to balance both speed and accuracy for a wide range of practical applications.

### Transformer-Based Models

While anchor-free methods represented a significant step towards simplifying the object detection pipeline and reducing reliance on predefined bounding boxes, the next wave of innovation came from transformer-based architectures. These transformer-based detectors leverage self-attention mechanisms to capture global context and directly predict object sets without traditional anchors or post-processing heuristics. Models like the Detection Transformer (DETR) introduced by Carion et al. (2020) introduced a novel approach that reframes object detection as a direct set prediction problem, enabling the model to learn object relationships and scene structure holistically. By integrating the principles of anchor-free design with the flexibility and contextual reasoning provided by transformers, researchers have paved the way for even more streamlined and robust detection frameworks that excel in complex, real-world scenarios. As shown in Figure 2.4, DETR utilizes transformers to capture global image contexts and model relationships between objects.

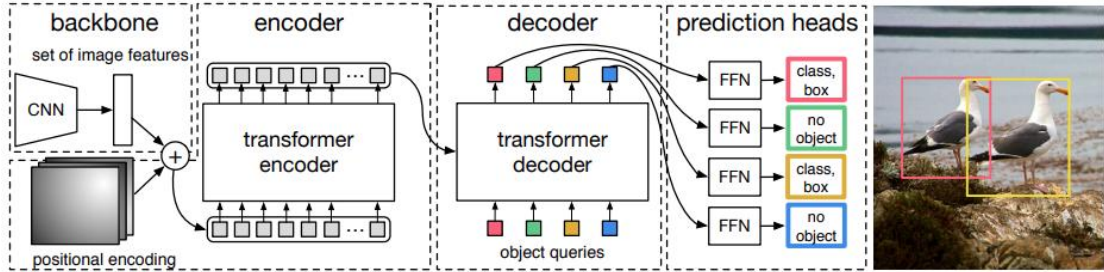


Figure 2.4: Detection Transformer [DETR] Architecture

Studies suggest that one-stage detection is faster and has relatively better accuracy compared to two-stage detection, anchor-free detectors (J. Huang et al., 2024; Karbouj et al., 2024; Li et al., 2023). Each of these models offers trade-offs between speed and accuracy. One-stage detector, YOLO, offers faster detection with a balance between speed and accuracy (Huang et al., 2017). In agricultural contexts, where processing large volumes of data quickly is essential, advanced YOLO models are often preferred (Al et al., 2024).

### ***2.2.2 Seedling Detection and Counting***

Recent advancements in computer vision have led to a wide range of agricultural applications (Dhanya et al., 2022), particularly plant counting, using techniques such as convolutional neural networks (CNNs), object detection frameworks like YOLO, and semantic segmentation models. Advancements include various tasks, such as crop monitoring (Monteiro et al., 2021), yield estimation (Ariza-Sentís et al., 2023), and disease identification (Harakannanavar et al., 2022).

Green et al. (2020) evaluated the use of unmanned aerial vehicles (UAVs) equipped with RGB cameras and computer vision techniques to count loblolly pine seedlings, highlighting the potential of UAV-based surveys despite limitations in GPS accuracy and image resolution. While the study reported acceptable accuracy under specific conditions, limitations stemming from equipment quality—such as GPS errors and lower image resolution—compromised data reliability by reducing the precision of spatial measurements of the seedlings.

Bidese-Puhl et al. (2023) developed a ground-based counting system using a multi-camera setup and optical flow for pre-harvest seedling inventory, demonstrating that deep learning-based regression models can outperform manual counting practices

in terms of accuracy and efficiency. By integrating video and optical flow data with convolutional and recurrent neural networks, their approach achieved a mean absolute percentage error (MAPE) of 7.5%, indicating a high level of accuracy that is crucial for a reliable inventory system for pre-harvest seedlings (~0.3 m) in large-scale forest nurseries.

However, an automated counting system for early-stage seedlings remains unexplored due to challenges such as their small size (<0.15 m), dense seedlings, and viewing perspective of the seedlings. Jiang et al. (2019) combined object detection and Kalman filtering techniques to achieve high-precision counting of smaller cotton seedlings ( $R^2 = 0.98$ ), showcasing the effectiveness of integrated deep learning and tracking frameworks. Similarly, Wang et al. (2021) developed an ultra-compact field robot using deep learning-based corn stand counting algorithms, achieving robust and reliable counting even under variable field conditions through high-precision control and automated phenotyping and Zhang et al. (2020) employed YOLO-based models integrated with Kalman filters for real-time corn stand counting from RGB videos with  $R = 0.96$ .

Despite these technological advances, research on early-stage pine seedling counting remains unexplored, emphasizing the need for innovative methods tailored to forestry applications. This gap highlights an opportunity to advance the automation of forest nursery management by developing a robust counting system tailored for early-stage seedlings. Leveraging high-resolution RGB imaging and deep learning models for object detection in cluttered forest nursery environments could effectively address this need.

### 2.2.3 Camera Imaging (Viewing) Angles

Camera imaging angles, or viewing angles, define the viewing perspective of the objects present in the scene. These angles determine how objects are perceived in the captured imagery (Berndt et al., 2023). As illustrated in Figure 2.5, the appearance of a car viewed from directly overhead (nadir) differs notably from its appearance at an oblique imaging angle. At more oblique angles, objects may partially occlude or overlap with each other, underscoring the influence of viewing angles on how smaller objects are captured.

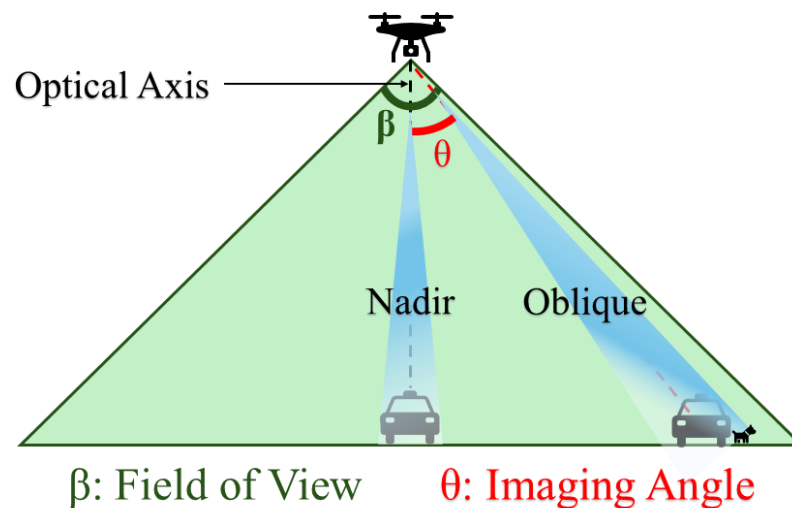


Figure 2.5: Illustration of different viewing perspectives of the object located at the nadir and at an oblique viewing angle.

In developing an automated counting system for early-stage seedlings, the selection of viewing angles is critical. For instance, a wide-angle lens can provide a more comprehensive overview of the scene, aiding in capturing objects closely at extreme angles (Paine & Kiser, 2012), while a narrow angle may highlight subtle

features of the seedlings. Similarly, different angles may reveal different views of the objects, influencing the clarity and separability of targets within a complex environment such as a forest nursery. This potentially affects the detection of objects that are located at various oblique angles, making detection more challenging (Berndt et al., 2023).

Despite the acknowledged importance of viewing angles, very few empirical studies systematically analyzed how different viewing angles affect detection performance in agricultural settings. It is possible that the viewing perspective of the seedlings may influence the performance of the detection model. Gaining a better understanding of how viewing angles shape detection outcomes is essential for designing effective camera configurations in nursery environments. This knowledge can inform best practices for an efficient automated system, ensuring that the chosen viewing angles align with the goals of accurate and reliable inventory management.

Accordingly, this study aims to address this by examining how various viewing angles affect the detection accuracy of pine seedlings. By assessing detection performance across multiple perspectives, the research seeks to provide actionable insights for automated counting systems for forest nurseries.

## **2.3 Methodology**

### **2.3.1 Data Collection**

Data collection was performed at a commercial forest nursery (IFCO BAREROOT, Verbena, AL, USA) in June 2021. Video of seven seedling beds was recorded using an action camera (VIRB ULTRA 30, GARMIN, Olathe, KS, USA) at 30 frames per second (fps) with a resolution of 3840×1920 pixels, covering eight crop

rows per bed spanning 1.2m in width, details can be found in Table 2.1. The camera was mounted at the center of a push bar attached to a tractor traveling at 0.5 m/s, and the camera was positioned at approximately 0.5 m above the bed (Bidese-Puhl et al., 2023).

Table 2.1: Details of RGB video samples recorded at IFCO BARERROOT Nursery and used in this study.

Bed ID	Number of video samples	Frames
1	4	240
2	2	120
3	4	240
4	2	120
5	2	120
6	2	120
7	2	120

### 2.3.2 Data Preprocessing

#### Data Annotation

Frames were extracted from the collected videos using the OpenCV library (Bradski & Kaehler, 2000). The extracted frames were then manually annotated using the Computer Vision Annotation Tool (CVAT) (Sekachev et al., 2020), where there were 147 seedlings per frame. Annotations were exported in the YOLO format for training purposes and in the MOT format for evaluating crop counts. The MOT labels included unique identity numbers (IDs) for each seedling, allowing for consistent tracking of individual seedlings throughout the video sequences, which is essential for accurate counting.

#### Data Augmentation

Data augmentation techniques were applied using Roboflow to enhance the robustness and generalizability of the detection models. Images were rotated between -15 and +15 degrees to simulate the natural variations in seedling orientations. Horizontal and vertical flips were also applied to augment the dataset further, ensuring that the model is not biased toward any specific orientation. Additional augmentations included adjustments to color, brightness (ranging from -17% to +17%), exposure (from -5% to +5%), and saturation (from -25% to +25%) to simulate variable lighting conditions commonly encountered in outdoor environments.

### **Dataset Composition**

The collected videos were strategically split to train robust models and evaluate them under various conditions. The training set comprised 480 images randomly selected from multiple videos, ensuring a diverse representation of scenarios. A validation set included 120 images, which were used for hyperparameter tuning and monitoring model performance to prevent overfitting. The test set consisted of 8 unseen videos, encompassing a total of 480 frames. Evaluating the model on these unseen videos allowed for an in-depth assessment of its effectiveness, providing insights into operational reliability and accuracy in real-world conditions.

### **2.3.3 Model Implementation**

#### **Model Selection Rationale**

The selection of detection models was guided by several key criteria, including detection accuracy, processing speed, the ability to detect small and densely packed objects, and computational efficiency. Precise localization was essential for accurate counting, making detection accuracy a top priority. Additionally, the models needed to handle high-resolution images at higher frame rates to ensure timely processing. The

capability to effectively detect small and densely packed seedlings was crucial for the application's success. Balancing latency against speed was also important to maintain computational efficiency for inventory purposes. Taking these factors into account, the YOLO architecture was chosen, specifically YOLOv8, YOLOv9, and YOLOv10, due to their significant advancements in speed, accuracy, and scalability (Wang et al., 2024).

### **2.3.3.1 Model Architectures**

#### **YOLOv8**

YOLOv8, developed by Ultralytics and released in January 2023, builds upon the successes of its predecessors to deliver enhanced performance in object detection tasks (Glenn Jocher et al, 2023). YOLOv8 introduces an anchor-free detection mechanism, simplifying the model architecture and improving performance on small objects—a crucial factor in detecting pine seedlings. Its real-time detection capabilities make it suitable for applications requiring immediate processing, such as automated inventory systems.

YOLOv8 incorporates advanced architectural features and training methodologies, allowing for precise object detection even in complex scenes. It features a new backbone network utilizing Cross Stage Partial (CSP) DarkNet53 connections (Wang et al., 2019) can be seen in Figure 2.6, enhancing feature extraction while reducing computational load. The model employs an improved neck for better feature aggregation and an anchor-free head for more efficient detection. Additionally, YOLOv8 integrates advanced training strategies and loss functions that contribute to faster convergence and higher accuracy. YOLOv8 is available in

multiple variants—Nano, Small, Medium, Large, and Extra-Large—to suit different performance and resource requirements.

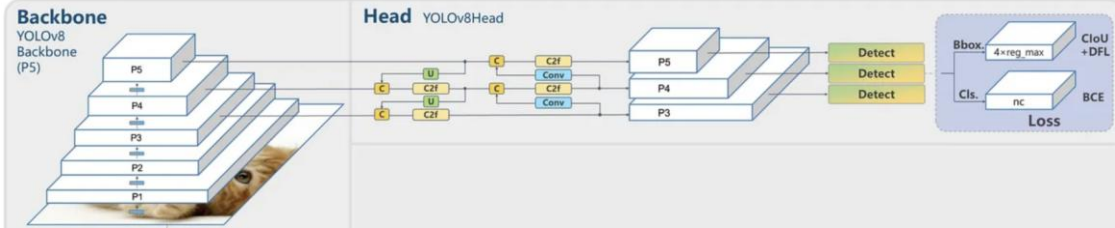


Figure 2.6: YOLOv8 Architecture.

### YOLOv9

YOLOv9, introduced by Wang et al. (2024), builds upon the foundations of the YOLO series, addressing key challenges such as information loss and unreliable gradient flows in deep neural networks. The model introduces Programmable Gradient Information (PGI) and a Generalized Efficient Layer Aggregation Network (GELAN), which collectively enhance gradient flow and information retention. PGI allows for precise gradient computations necessary for effective model training, promoting better learning dynamics and feature extraction. PGI can be seen compared with other networks in Figure 2.7. These features enable YOLOv9 to achieve high accuracy while maintaining operational efficiency, particularly notable in handling deep supervision and lightweight network architectures.

YOLOv9 is available in several variants—tiny, small, medium, compact, and extensive—each optimized for specific performance and computational needs. This range allows for a balance between speed, accuracy, and computational efficiency, catering to diverse application scenarios.

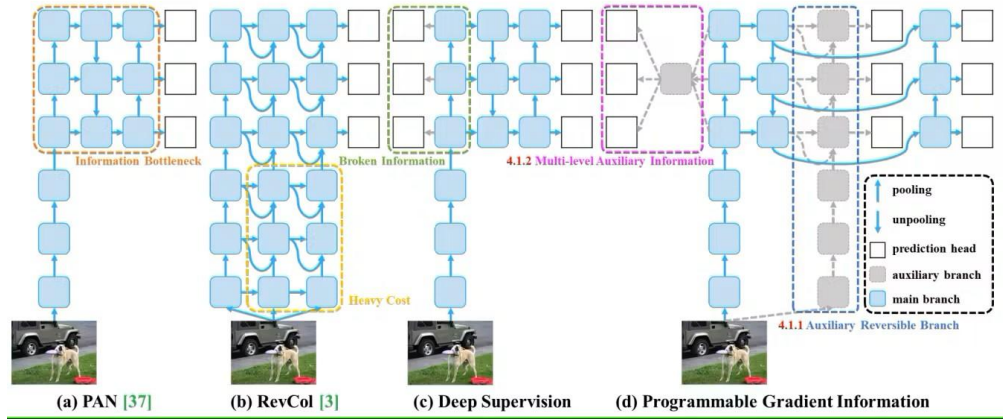


Figure 2.7: PGI and related network architectures.

## YOLOv10

YOLOv10, introduced by Wang et al. (2024), represents a significant advancement in the YOLO series. It addresses the limitations of previous versions while introducing innovative features to enhance performance and efficiency. YOLOv10 focuses on balancing efficiency and accuracy through optimizations in both architectural design and training protocols. It incorporates advanced techniques like Non-Maximum Suppression (NMS)-free training and holistic model design, making it particularly efficient for edge devices with limited computational resources. The model also includes architectural elements like large-kernel convolutions and partial self-attention to boost performance without extensive computational overhead. Figure 2.8 shows the advancements to YOLOv10, a consistent dual assignment for NMS-free training.

YOLOv10 demonstrates enhanced capability in detecting smaller objects and exhibits a lower rate of false predictions, improving overall detection reliability. It offers flexibility in deployment, with multiple variants such as nano, small, medium,

balanced, large, and extra-large to cater to different performance needs and computational capabilities.

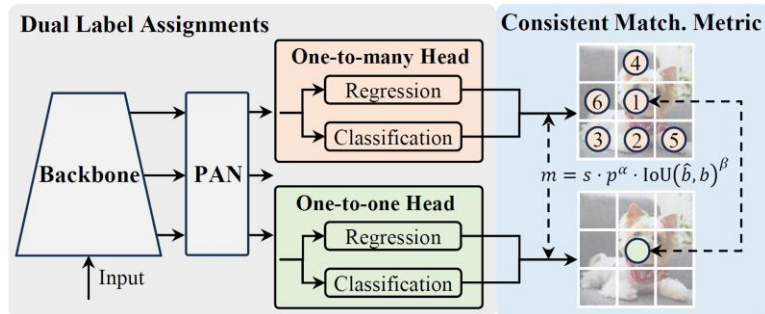


Figure 2.8: Consistent dual assignment for NMS-free training.

A variety of YOLO models were selected for benchmarking to assess how different architectures perform in the context of pine seedling detection. This included YOLOv8 (Nano, Small, Medium, Large), YOLOv9 (Tiny, Small, Medium, Compact), and YOLOv10 (Nano, Small, Medium, Balanced) models. Table 2.2 provides an overview of the selected model variants, including their parameters, floating-point operations per second (FLOPs), and mean Average Precision (mAP) values. These models were trained and validated on the MS COCO dataset (Lin et al., 2014). This comprehensive selection allows for a detailed comparison of how model parameters, layers, and architectural nuances influence the effectiveness of detecting pine seedlings.

Table 2.2: Overview of the YOLOv8, YOLOv9, and YOLOv10 model variants, which were trained on the COCO dataset, including 80 pre-trained classes.

Model	Parameters(M)	Layers	FLOPs(G)	mAP-val (50-95)
-------	---------------	--------	----------	-----------------

YOLOv8n	3.2	225	8.7	37.3
YOLOv8s	11.2	225	28.6	44.9
YOLOv8m	25.9	295	78.9	50.2
YOLOv8l	43.7	365	165.2	52.9
YOLOv9t	2.0	917	7.7	38.3
YOLOv9s	7.2	917	26.7	46.8
YOLOv9m	20.1	603	76.8	51.4
YOLOv9c	25.5	618	102.8	53.0
YOLOv10n	2.3	385	6.7	39.5
YOLOv10s	7.2	402	21.6	46.8
YOLOv10m	15.4	498	59.1	51.3
YOLOv10b	19.1	518	92.0	52.7

### 2.3.3.2 Model Training

The models were trained over 100 to 250 epochs to ensure adequate learning across all configurations. An image size of 960 pixels was used, optimized for capturing detailed features of pine seedlings, with a batch size of 16 to balance memory constraints and training efficiency. An initial learning rate of 0.002 was selected to balance training stability and convergence speed across different model architectures. Learning rate adjustments were made dynamically during training to adapt to each model's learning progress. The AdamW optimizer was employed for its advantages in handling sparse gradients and optimizing deep neural networks effectively. A momentum value of 0.9 was used to facilitate faster convergence and stability during training.

### 2.3.3.3 Training Environments

The training process was carried out using NVIDIA A100 GPUs with high RAM configurations, accessed through Google Colab Pro+. This robust hardware setup provided the essential computational resources necessary for efficiently training deep learning models. On the software side, Python 3.12.1, Ultralytics 8.1.1 (Glenn

Jocher et al., 2023), the PyTorch 2.5 (Paszke et al., 2019), and CUDA 11.8 were utilized for their flexibility and efficiency in handling deep learning models. Additionally, CUDA was employed to leverage GPU acceleration alongside other essential libraries and dependencies required for implementing the YOLO models. Together, this hardware and software configuration ensured a seamless and effective environment for training advanced detection models.

### **2.3.4 Evaluation Metrics**

Evaluation metrics are essential for assessing the performance of object detection models. The following metrics were used in this study:

#### **2.3.4.1 Intersection over Union (IoU)**

IoU, also known as the Jaccard Index, measures the overlap between the predicted bounding box and the ground truth bounding box. It is calculated as the area of overlap divided by the area of union of the two boxes. An IoU score ranges from 0 to 1, with higher scores indicating better localization accuracy. It can be seen in the following Figure 2.9.

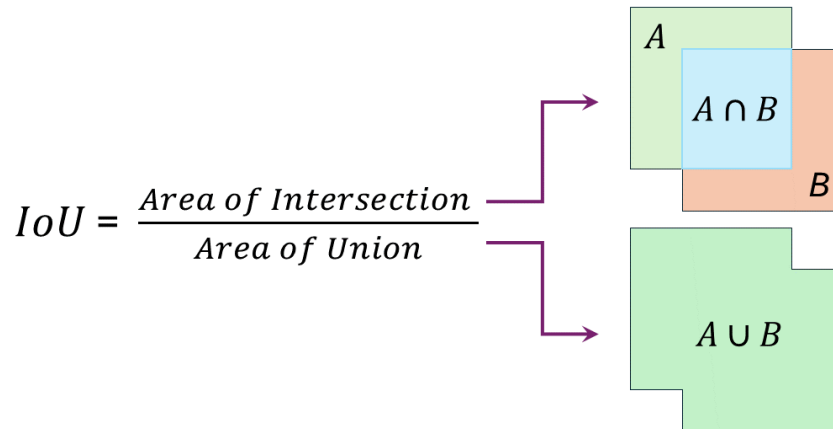


Figure 2.9: Intersection over Union.

To evaluate the object detection model's effectiveness in accurately identifying and localizing seedlings within images, we computed the IoU for each detected seedling by comparing the predicted bounding boxes against the ground truth labels.

#### 2.3.4.2 Precision, Recall and F1-Score

Precision, Recall, and F1-Score are fundamental metrics used to assess the performance of object detection models. These metrics provide valuable insights into the model's ability to identify objects of interest within images.

In object detection, True Positives (TP) occur when the model correctly identifies and localizes objects, with the IoU score between the predicted and ground truth bounding boxes meeting or exceeding a specified threshold. False Positives (FP) happen when the model incorrectly detects objects that do not exist in the ground truth or when the predicted bounding boxes have IoU scores below the threshold. False Negatives (FN) are instances where the model fails to detect objects that are present in

the ground truth. True Negatives (TN) are not applicable in object detection, as the primary goal is to identify the presence of objects rather than confirm their absence.

Precision measures the ratio of true positive detections to the total number of positive detections made by the model. A high precision indicates a low false positive rate.

$$precision (p) = \frac{TP}{TP + FP} \quad (1)$$

Recall measures the ratio of true positive detections to the total number of actual positive instances in the ground truth. A high recall indicates a low false negative rate.

$$recall (r) = \frac{TP}{TP + FN} \quad (2)$$

Precision can be seen as a measure of quality, while recall can be seen as a measure of quantity. A model with higher precision returns more relevant results than irrelevant ones, while a model with higher recall returns most of the appropriate results.

$F_1$  – score is the harmonic mean of precision and recall. It provides a balanced measure of the model’s performance, considering both false positives and false negatives.

$$F_1 = \frac{2 * (precision * recall)}{(precision + recall)} \quad (3)$$

### 2.3.4.3 Average Precision (AP)

Average Precision (AP) is the area under the Precision-Recall (PR) curve, which plots precision (1) against recall (2). It summarizes the PR Curve to one scalar value. AP is high when both precision and recall are high and low when either of them is low across a range of confidence threshold values.

$$\text{Average Precision (AP)} = \int_{r=0}^1 p(r) dr \quad (4)$$

where  $r$  denotes recall, and  $p(r)$  denotes the precision at a given recall level ( $r$ ).

### 2.3.4.4 Mean Average Precision (mAP)

Calculated by averaging the AP values across all classes. For this study, mAP@0.50 was used. The mAP quantifies the performance of the detection model; the higher the mAP score, the more accurate the model is in detecting and making correct predictions. We can calculate the mean Average Precision across all the  $k$  classes.

$$\text{mean Average Precision (mAP)} = \frac{1}{k} \sum_{i=1}^k AP_i \quad (5)$$

where  $k$  is the number of classes (starting from  $i=1$ ),  $AP_i$  is the average precision of class  $i$ .

To get mAP50, the mean average precision is calculated at an IoU threshold of 0.50. It gives a comprehensive view of the model's performance across different levels of detection difficulty.

### 2.3.5 Camera Imaging Angles

The concept of imaging angle can be explained through the understanding of horizontal and vertical viewing angles. The horizontal viewing angle (HVA) ( $\theta_h$ ) is the angle between the camera's optical axis and the object's horizontal position (projection from the center to the horizontal plane). Vertical viewing angle (VVA) ( $\theta_v$ ) is the angle between the camera's optical axis and the object's vertical position (projection from the center to the vertical plane). HVA and VVA for a pixel or for the object using its center coordinates of the bounding box can be calculated as follows:

$$HVA(\theta_h) = \tan^{-1} \left( \frac{\frac{x - \frac{w}{2}}{w}}{2 \tan \left( \frac{HFOV \times \left( \frac{\pi}{180^\circ} \right)}{2} \right)} \right) \times \left( \frac{180^\circ}{\pi} \right) \quad (6)$$

$$VVA(\theta_v) = \tan^{-1} \left( \frac{\frac{\frac{h}{2} - y}{h}}{2 \tan \left( \frac{VFOV \times \left( \frac{\pi}{180^\circ} \right)}{2} \right)} \right) \times \left( \frac{180^\circ}{\pi} \right) \quad (7)$$

where  $x$  and  $y$  are the pixel coordinates of the object's center,  $w$  and  $h$  are the image width and the image height in pixels, and HFoV and VFoV are the camera's horizontal field of view and vertical field of view in degrees, respectively. The horizontal and vertical viewing angles of an *object* from the camera can be illustrated in Figure 2.10.

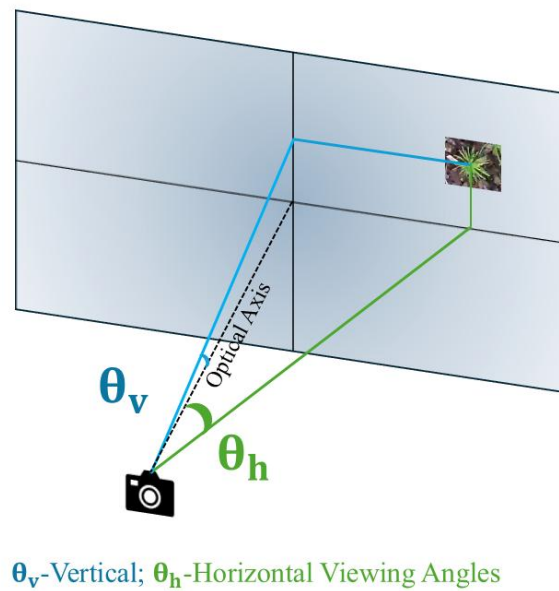


Figure 2.10: Horizontal ( $\theta_h$ ) and vertical ( $\theta_v$ ) viewing angles of the object from the camera.

To analyze the effect of viewing angles on detection performance, we computed the IoU score and horizontal and vertical viewing angles for all of the pine seedlings and analyzed the relationship between the IoU score and viewing angles.

From this, we can determine the influence of imaging (viewing) angles on detection performance.

## 2.4 Results

To evaluate our detection models, we used eight ground-truth datasets of manually annotated pine seedlings. Each image in the datasets was carefully annotated, and the number of seedlings was recorded. This process resulted in an average count of 152.5 seedlings per image ( $\pm 9.3$ ), with counts ranging from a minimum of 133 to a maximum of 174. These figures indicate a relatively consistent density of pine seedlings across the datasets.

### 2.4.1 Latency Comparisons of Different Model Architectures

Table 2.3 presents a comparison between similar architectures of different YOLO models, highlighting the better latencies with bold letters. Overall, YOLOv10 models has the fastest inference compared to YOLOv8 and YOLOv9, except for the biggest architecture, where YOLOv9 is faster than the other two. It clearly indicates that as the size of the architecture increases, so do the inference timings.

Table 2.3: Latency and similar architectural comparison of the models.

Model	Parameters (million)	Layers	GFLOPS	Latency (ms)
YOLOv8nano	3.1	225	8.2	18.625
YOLOv9tiny	2.0	917	7.8	32.725
YOLOv10nano	2.7	385	8.4	<b>16.225</b>
YOLOv8small	11.1	225	28.6	21.700
YOLOv9small	7.2	917	27.4	29.500
YOLOv10small	8.1	402	24.8	<b>16.875</b>
YOLOv8medium	25.9	295	79.3	23.970
YOLOv9medium	20.1	603	77.5	27.300

YOLOv10medium	16.5	498	64.0	<b>22.175</b>
YOLOv8large	43.7	365	165.7	33.225
YOLOv9compact	25.5	618	103.7	<b>26.800</b>
YOLOv10balanced	20.4	518	98.7	28.825

### 2.4.2 Detection Performance

Table 2.4 shows the performance of YOLOv8, YOLOv9, and YOLOv10 in the detection of pine seedlings. The YOLOv10balanced model achieved the highest mAP50 of 0.955 and an F1 Score of 0.920, indicating superior detection capabilities. As depicted in Figure 2.11, the detections confirm the model's effectiveness in accurately detecting multiple pine seedlings. Additionally, YOLOv10medium and YOLOv8large also exhibit similar accuracies. The results demonstrate that the bigger the architecture, the better the model is at reducing incorrect and missing detections and increasing accuracy.

Table 2.4: Detection metrics of all the models.

Model	TP	FP	FN	P	R	F1	mAP50
YOLOv8nano	31731	3048	4867	0.913	0.868	0.89	0.936
YOLOv9tiny	31855	2768	4743	0.92	0.871	0.89	0.943
YOLOv10nano	31107	2170	5491	0.935	0.851	0.89	0.945
YOLOv8small	31737	2636	4861	0.924	0.868	0.895	0.937
YOLOv9small	32316	2585	4282	0.928	0.884	0.905	0.949
YOLOv10small	32094	2495	4504	0.93	0.878	0.903	0.947
YOLOv8medium	32348	2678	4250	0.926	0.885	0.905	0.943
YOLOv9medium	32129	2532	4469	0.929	0.879	0.903	0.947
YOLOv10medium	32064	<b>1891</b>	4534	<b>0.945</b>	0.876	0.91	0.952
YOLOv8large	32513	2293	4085	0.935	0.889	0.911	0.95
YOLOv9compact	32224	2760	4374	0.923	0.881	0.901	0.939
YOLOv10balanced	<b>32821</b>	1951	<b>3777</b>	0.944	<b>0.897</b>	<b>0.92</b>	<b>0.955</b>

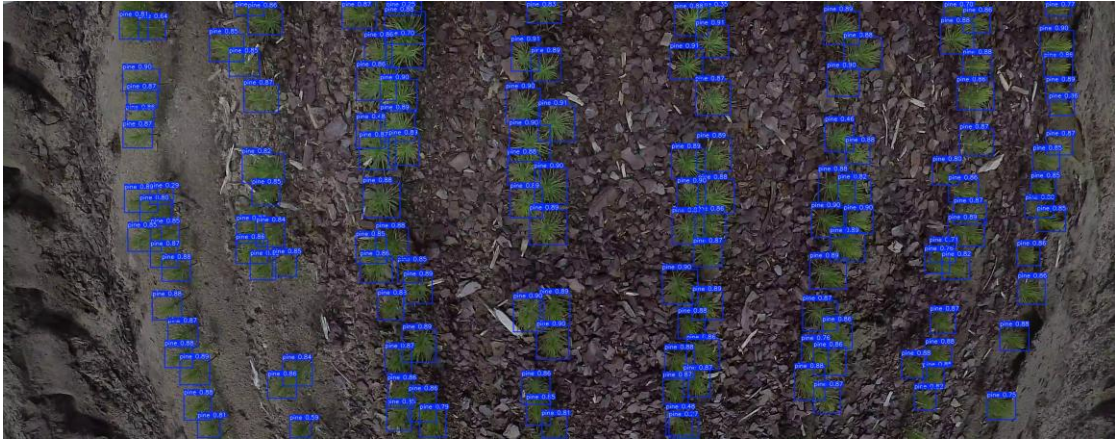


Figure 2.11: Detections by YOLOv10-balanced model (label: object, confidence).

The resulting IoU scores are visualized in Figure 2.12, where each point represents a seedling's center coordinates within the image plane, and a color gradient ranging from blue to red indicates IoU scores from 0 to 1. The spatial distribution of these seedlings reveals a prominent concentration of higher IoU scores in the central region of the seedling bed, suggesting high detection accuracy. In contrast, as the seedling positions extend towards the periphery of the image, the detection accuracy decreases.

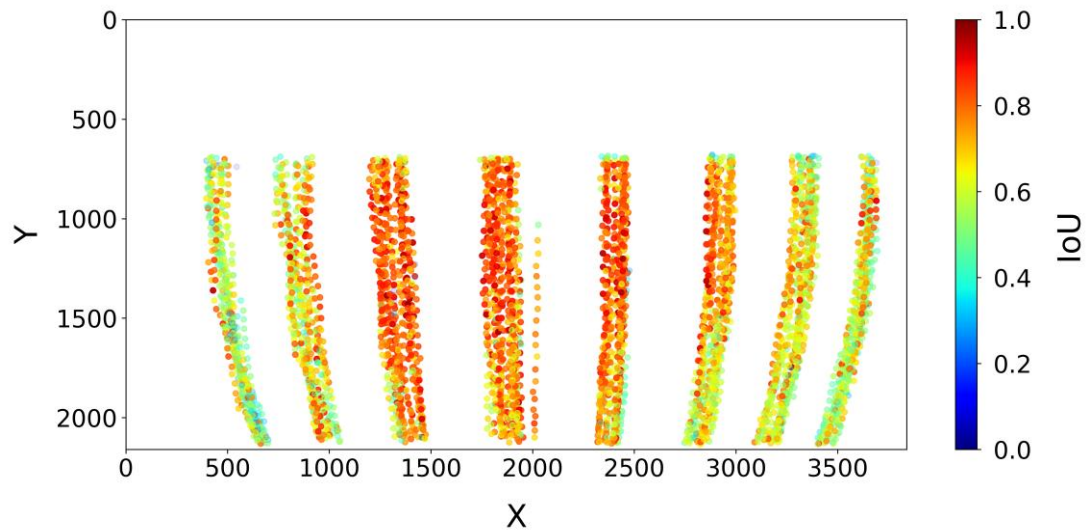


Figure 2.12: Spatial distribution of IoU scores.

To understand whether there were more crops at the center or uniformly distributed, each crop row was divided into three segments: top, central, and bottom segments. The top segment has been left out due to the presence of a push bar. The mean IoU scores were computed for each segment; Figure 2.13 indicates the higher mean in the middle rows. On the other hand, central segments possess a greater mean than the bottom segments.

Pearson correlation analysis was conducted to assess the relationship between the mean Intersection over Union (meanIoU) and the number of crops in the segments. Table 2.5 summarizes the correlation coefficients ( $r$ ) and p-values ( $p$ ). The central and bottom segments possess very low correlation, indicating no association between the number of crops and the mean IoU.

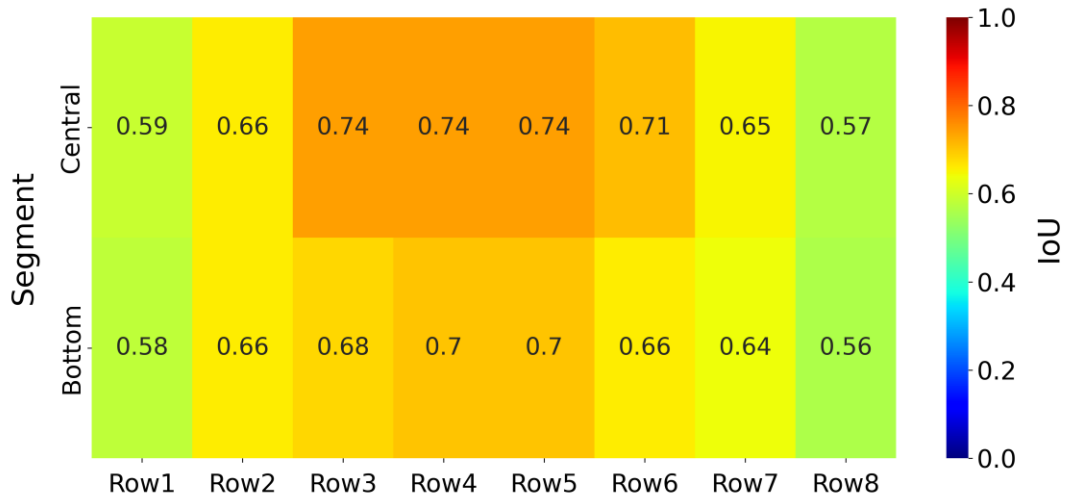


Figure 2.13: Mean IoU of all the rows in central and bottom segments.

Table 2.5: Pearson Correlation ( $r$ ) between mean IoU and the number of crops.

Model	Central Segment		Bottom Segment	
	$r$	$p$	$r$	$p$
YOLOv8nano	.098	.0010	-0.033	.18
YOLOv9tiny	.13	.0020	-0.121	.026
YOLOv10nano	.107	.001	-0.179	.000
YOLOv8small	.155	.0000	-0.096	.071
YOLOv9small	.103	.125	-0.109	.096
YOLOv10small	.09	.111	-0.125	.062
YOLOv8medium	.096	.0050	-0.037	.095
YOLOv9medium	.097	.01	-0.101	.004
YOLOv10medium	.116	.000	-0.118	.012
YOLOv8large	.122	.0000	-0.085	.052
YOLOv9compact	.101	.004	-0.124	.033
YOLOv10balanced	.132	.015	-0.071	.147

### 2.4.3 Effect of Viewing Angles

The distributions of IoU scores relative to both horizontal ( $-66.8^\circ$ ,  $66.8^\circ$ ) and vertical ( $-34.6^\circ$ ,  $34.6^\circ$ ) viewing angles were analyzed to assess the impact of camera imaging angles on the detection accuracy of pine seedlings. Figure 2.14 reveals that

the highest concentrations of seedlings, particularly those with IoU scores around 0.8, are clustered around the central viewing angles of  $-40^\circ$ ,  $0^\circ$ , and  $40^\circ$ . As the HVA extends toward the extremes of  $-60^\circ$  and  $60^\circ$ , there is a noticeable variability in the distribution of IoU scores, with some seedlings registering an IoU of zero, indicating decreased detection quality at these angles.

Similarly, the distribution in Figure 2.15 indicates a nearly uniform distribution of IoU scores as the angle deviates from  $0^\circ$  towards  $-30^\circ$  and  $20^\circ$ . Notably, a significant frequency exhibits an IoU of zero at the negative extreme angles, suggesting diminished detection capability in these locations. Overall, HVA indicates that IoU scores are maximized at central viewing angles and decrease at extreme angles, whereas VVA shows minimal influence on IoU scores.

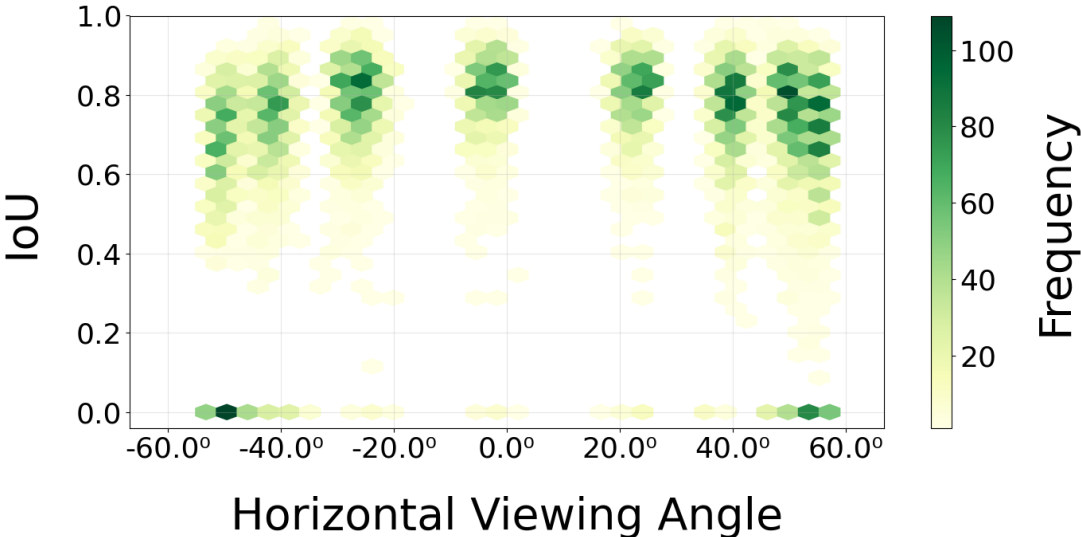


Figure 2.14: Distribution of IoU scores at different horizontal viewing angles.

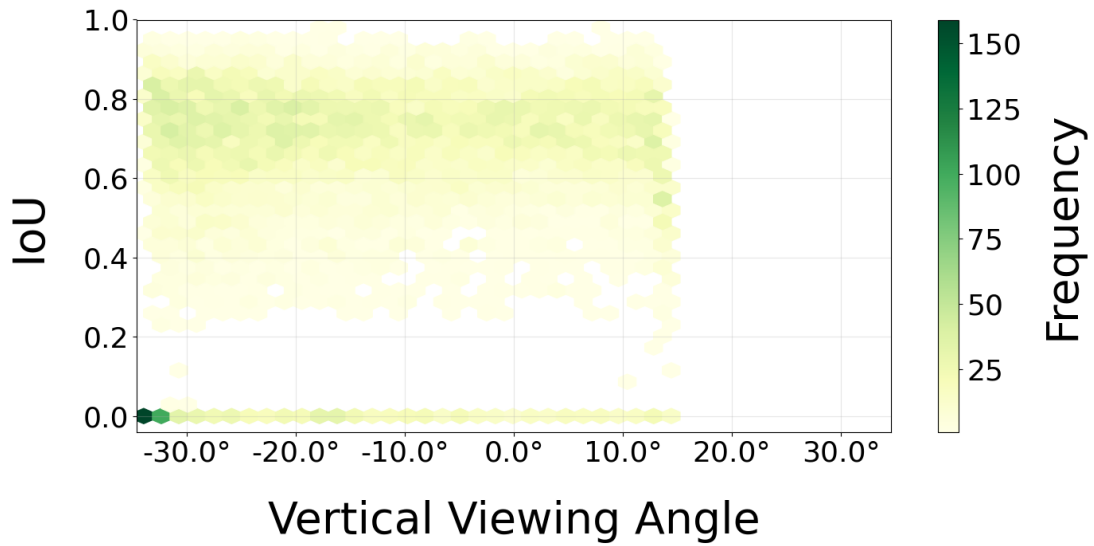


Figure 2.15: Distribution of IoU scores at different vertical viewing angles.

The table 2.6 summarizes the statistical significance of HVA (bin size=15) and VVA (bin size=10) among different angle bins for finding the change in IoU at different angles. The Kruskal-Wallis Test indicates significant differences in both HVA and VVA. Dunn’s Test reveals significance across all pairwise bins for HVA, while VVA shows significance only in central and extreme angle bins. Cliff’s Delta effect sizes range from small to large for HVA and negligible for VVA, indicating varying degrees of effect across angles.

Table 2.6: Statistical test for significance of HVA and VVA against IoU.

Statistical Test	HVA	VVA
Kruskal-Wallis Test	Significant	Significant
Dunn’s Test	Significant in all pairwise bins	Only significant in central and extreme angle bins
Cliff’s Delta	Small to large	Negligible to small

## **2.5 Discussion and Conclusion**

### **2.5.1 Discussion**

The results obtained from this study provide crucial insights into the performance of different YOLO models for detecting pine seedlings. The findings contribute to addressing the existing challenges in automated seedling counting, particularly under dense planting scenarios where early-stage pine seedlings are typically smaller and harder to detect.

#### **2.5.1.1 Model Performance**

The latency comparison across YOLO models indicates a clear trade-off between model size and inference time. YOLOv10balanced demonstrated the best balance between low latency and high detection accuracy. This supports the notion that architectural optimization, as seen in YOLOv10 (Wang et al., 2024), can significantly reduce inference times while maintaining accuracy.

Our detection performance yielded a mAP50 of 0.955 and an F1 score of 0.920 for YOLOv10balanced, similar to a study on small seedlings by Jiang et al. (2019), which achieved a mAP50 score of 0.969 for cotton seedling detection using Faster R-CNN. However, the seedling density is very low (15 cotton seedlings/frame) compared to ours (147 pine seedlings/frame). This discrepancy underscores the scalability challenge addressed by our model, which successfully managed a much higher object density while maintaining competitive performance metrics.

Similarly, Wang et al. (2021) applied YOLOv3 for corn stand counting and achieved a mAP50 of 0.98, benefiting from greater intra-plant spacing and only four seedlings per frame. In comparison, our YOLOv10balanced model achieved a mAP50

of 0.955 despite operating in more visually complex and densely populated scenes, reinforcing its robustness in challenging forest nursery environments.

### **2.5.1.2 Impact of Viewing Angles**

The spatial distribution of IoU scores revealed better detection accuracy in the center than peripheral region. Analyzing the spatial variation in IoU scores revealed that the central segments of crop rows had higher mean IoU scores compared to the bottom segments. This indicates that detection accuracy is higher in the middle of the images, maybe due to a better viewing perspective of the seedlings. The Pearson correlation analysis between mean IoU and the number of crops in each segment shows that detection accuracy is driven by viewing geometry rather than plant density.

The impact of viewing angles on detection performance was evident from the IoU distribution analysis. Horizontal viewing angles significantly affected detection accuracy, with central angles between  $-40^\circ$  and  $40^\circ$  yielding the highest IoU scores, while extreme angles less than  $-60^\circ$  and greater than  $60^\circ$  exhibited considerable declines. This trend mirrors findings by Minhui et al. (2021), who demonstrated that optimal HVA ( $0^\circ$ ) significantly improved leaf parameter estimation from 3D point clouds of wheat plants, as  $30^\circ$  and  $45^\circ$  led to higher error rates because of oblique side views, causing overlapping leaves and occlusions that reduced 3D reconstruction accuracy. Vertical viewing angles had a minimal effect, suggesting that the selection of HVAs plays a more dominant role in detection performance.

### **2.5.2 Limitations**

Despite the promising results, the study faced several limitations. Due to computational constraints, high-resolution data of  $3840 \times 2160$  pixels had to be

downscaled to 960×544 pixels for training and testing, even with access to powerful GPUs like the NVIDIA A100. This downscaling could potentially lead to the loss of fine-grained details that might be crucial for accurate detection. Additionally, implementing these models in real-time applications may be computationally intensive, especially when processing high-resolution images, which could limit the practicality of the system in operational settings where quick decision-making is essential. A push bar approximately covering one-third of the image obstructed the seedlings underneath, necessitating the exclusion of that portion from the study and thereby limiting the comprehensiveness of the dataset. Furthermore, seedlings at different growth stages exhibit significant variations in size, shape, and color, challenging the model's ability to accurately detect if they were trained only on a single growth stage. The presence of shadows and inconsistent lighting conditions in the images can obscure key features of the seedlings, leading to potential misdetections or reduced accuracy. Addressing these limitations in future work could involve leveraging higher computational resources to handle high-resolution images, enhancing the dataset with images captured under varied lighting and growth stages, and implementing advanced techniques to mitigate the effects of shadows and obstructions.

### **2.5.3 Recommendations**

Based on the findings and identified limitations, future research should focus on optimizing camera placements to maintain central and consistent viewing angles, potentially by deploying multiple cameras to cover different angles for more effective detection. To address the issue of shadows and inconsistent lighting, implementing controlled lighting environments or advanced image preprocessing techniques like

shadow removal algorithms (Guo et al., 2023; Liu et al., 2021) is recommended to enhance detection accuracy. Additionally, leveraging more advanced computational resources or developing optimized algorithms will allow for the utilization of high-resolution images without significant loss of detail, thereby improving the model's precision. Expanding and diversifying the dataset to include seedlings at various growth stages will enhance the model's ability to generalize across diverse scenarios. Incorporating real-time processing capabilities through model optimization and the use of specialized hardware accelerators can make the detection system more feasible for operational settings. Finally, conducting large-scale field trials to test the models in operational nursery settings would help assess their practical applicability. Feedback from these trials could inform adjustments to the models.

#### **2.5.4 Conclusion**

This study successfully developed and implemented an automated counting system for spring pine seedlings, providing a solution where previous research has largely focused on different growth stages or other plant species. By employing ground-based, high-resolution imagery and advanced deep-learning techniques, our approach achieved reliable detection performance even at challenging early developmental stages. In doing so, we investigated the effects of imaging angle on detection that can overcome the inherent difficulties presented by smaller, highly dense seedlings. These findings are particularly valuable for nursery management, as they help develop an optimized automated counting system. By improving the precision and efficiency of the counting process, our work sets the stage for enhanced crop monitoring strategies that can be adapted to a range of seedling sizes, species, and production environments. In spite of these advancements, further refinement is

possible. Future research could explore integrating tracking algorithms for efficient counting of seedlings. Such steps would further increase the resilience and scalability of automated counting systems, ultimately contributing to more sustainable and data-driven approaches in nursery and agricultural management.

## Chapter 3

### PINE SEEDLING TRACKING AND COUNTING

#### 3.1 Introduction

Tracking each seedling allows precise counting by maintaining the identities of seedlings across sequential frames in video data. This process, known as Multi-Object Tracking (MOT), involves identifying and continuously following multiple objects over time. In nursery environments, MOT is essential for precise monitoring and effective management, as it enables the consistent tracking of seedlings, avoiding double counting.

While Chapter 2 addressed the influence of HVA on detection accuracy, this chapter focuses on investigating the effects of the camera viewing angles and FoV on tracking and counting pine seedlings. Specifically, we explore how HVA and vertical field of view (VFoV) impact the performance of the combination of SoTA object detection and MOT algorithms in counting pine seedlings.

In this study, we implement and evaluate various MOT algorithms in conjunction with detections from the YOLO models discussed previously. We analyze their performance under different VFoVs, aiming to understand how they affect tracking accuracy and counting. Additionally, we compare and assess the effectiveness of these tracking algorithms in the context of pine seedling counting, identifying the most suitable approaches for nursery applications.

The chapter is organized as follows: Section 3.2 provides a literature review on the significance of MOT in agriculture and its applications in plant counting. Section

3.3 describes the methodology, including the implementation of tracking algorithms and the experimental setup. Section 3.4 presents and discusses the results obtained from counting pine seedlings, highlighting the impact of the FoV on tracking performance. Finally, Section 3.5 offers conclusions and discusses the implications of our findings, as well as directions for future research.

## **3.2 Literature Review**

### **3.2.1 *Multi-Object Tracking***

MOT is an important task in computer vision that involves identifying and locating multiple targets in an image, assigning them unique identities, and maintaining these identities across consecutive frames (Luo et al., 2021). With the availability of robust object detectors like YOLO, MOT is predominantly approached using the tracking-by-detection paradigm, as illustrated in Figure 3.1. This paradigm entails first detecting objects in each frame and then associating detections of the same object across frames based on motion and appearance cues.

In recent years, agriculture has seen a shift towards automation and digitization, often referred to as precision agriculture or smart farming (Liakos et al., 2018). Within this domain, MOT enables real-time monitoring of plants, animals, or machinery, providing data-driven insights to improve efficiency, crop management, and yield prediction (Sishodia et al., 2020). For instance, in fruit inventory management, tracking individual fruits allows for precise yield estimation and inventory control, aiding in decision-making processes (Zhang et al., 2022). Additionally, in livestock monitoring, tracking dairy cows enables health and behavior

analysis, which is essential for optimizing animal welfare and productivity (Zheng et al., 2023).

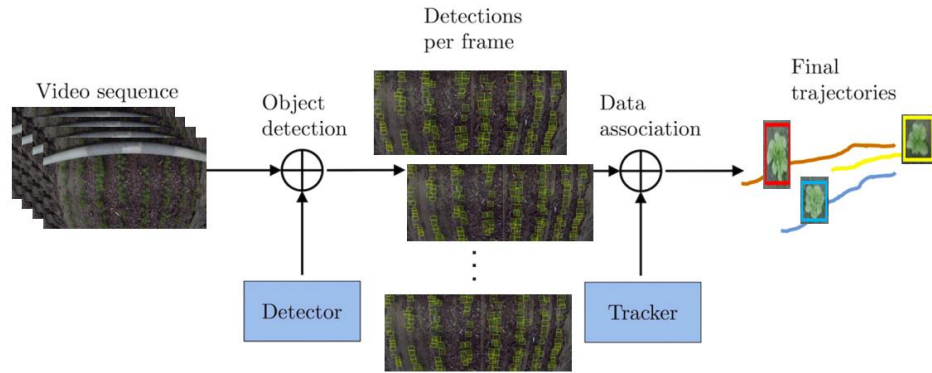


Figure 3.1: Tracking-by-detection paradigm.

Despite its extensive use in agriculture, limited research exists on its application in seedling nursery management. This gap highlights the need to explore MOT capabilities for efficient seedling counting, particularly given the challenges posed by dense planting and seedlings' similar appearance.

### 3.2.2 *Tracking Algorithms for Seedling Counting*

Several notable MOT algorithms have been employed in agricultural contexts, each with unique features and performance metrics that may influence their effectiveness in counting pine seedlings. This section discusses three such algorithms: Simple Online and Realtime Tracking (SORT), ByteTrack, and BoT-SORT.

### 3.2.2.1 Simple Online Realtime Tracking (SORT)

SORT was introduced by (Bewley et al., 2016) as a simple real-time approach to multi-object tracking. SORT uses a combination of a Kalman filter (Kalman, 1960) and the Hungarian algorithm (Kuhn, 1955) to predict the positions of objects and efficiently associate detections with these predictions, as depicted in Figure 3.2. The Kalman filter predicts the future position of each object based on its previous state, while the Hungarian algorithm solves the assignment problem of matching predicted positions to new detections. The simplicity of SORT allows it to run at high frame rates, which is beneficial for real-time applications (Bewley et al., 2016).

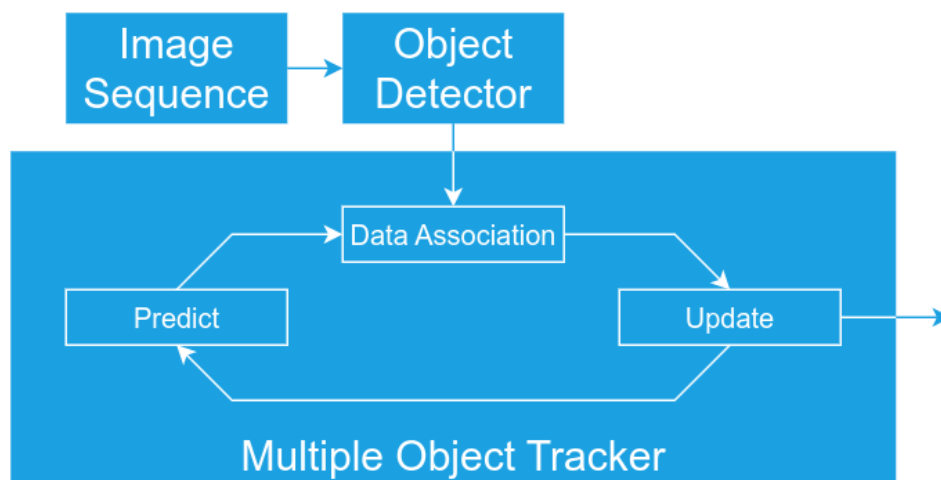


Figure 3.2: SORT: Simple Online and Realtime Tracking.

SORT has been utilized in various agricultural tracking tasks due to its computational efficiency. For instance, Chittupalli et al. (2022) applied SORT to count flowers and ripe and unripe strawberries and eventually predicted the yield. Similarly, Shen et al. (2023) used SORT to count the grape clusters to make informed decisions

before harvesting. Although these studies did not specifically focus on seedlings, they demonstrated SORT's capability to handle MOT in outdoor agricultural environments. However, SORT relies heavily on the quality of object detection and can struggle with long-term occlusions or identity switches when objects reappear after being occluded. This limitation is particularly relevant in dense nursery settings, where seedlings may occur frequently.

### **3.2.2.2 ByteTrack**

ByteTrack, introduced by Zhang et al. (2021) is an MOT algorithm designed to maintain robust tracking of objects even in complex scenarios. It refines object detections by considering both high-confidence and low-confidence detections, integrating them progressively to reduce identity switches and improve overall tracking accuracy. Figure 3.3 illustrates the ByteTrack framework, which associates every detection box, including those with low confidence scores, thereby maintaining more continuous tracks.

ByteTrack's robust tracking approach has proven useful in challenging agricultural environments. Gené-Mola et al. (2023) adopted ByteTrack to develop and evaluate a fruit counting system to predict orchard yield. By effectively handling occlusions and varying object appearances, ByteTrack improved the accuracy of fruit counting. Similarly, Cui et al. (2023) implemented ByteTrack for counting the missing seedlings in paddy fields in the early stages. These studies highlight ByteTrack's potential applicability in seedling counting tasks, where seedlings can partially occlude or overlap with each other, and maintaining consistent identities is essential for accurate counting.

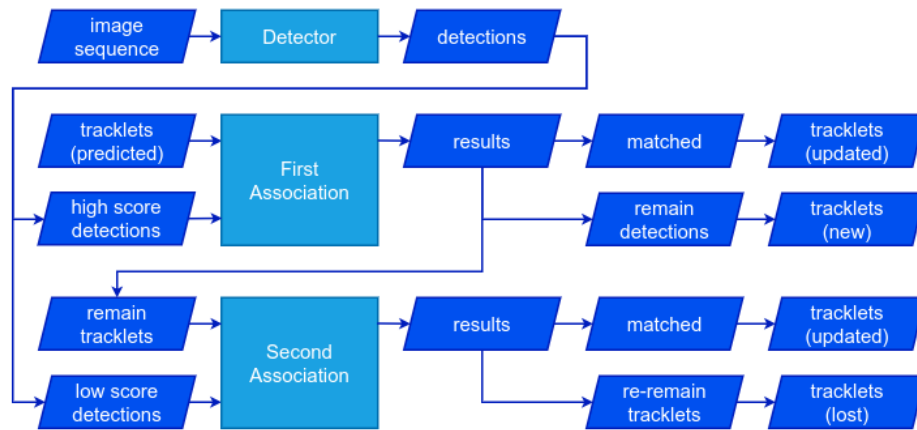


Figure 3.3: ByteTrack: Multi-Object Tracking by associating every detection box.

### 3.2.2.3 BoT-SORT (Boosted Tracking with SORT)

BoT-SORT, proposed by Aharon et al. (2022), is an enhancement of the original SORT algorithm that incorporates additional features such as Camera Motion Compensation and re-identification (ReID) models to improve object association, particularly in cases where objects are heavily occluded or temporarily leave the frame. Figure 3.4 provides an overview of the BoT-SORT pipeline, which includes these advanced components.

BoT-SORT maintains SORT's real-time capability while adding robustness to data association. The camera motion compensation module adjusts for any movement of the camera, which is particularly useful in dynamic environments or when the camera is mounted on moving platforms. The ReID module helps to re-identify objects that have been occluded or have left and re-entered the frame, reducing identity switches. While BoT-SORT is more accurate than SORT in many scenarios, it is still reliant on good initial detections and can be sensitive to long-term occlusions. Its advanced features make it a promising candidate for seedling counting in dense

nursery environments, where maintaining consistent identities of seedlings over time is crucial for accurate counting.

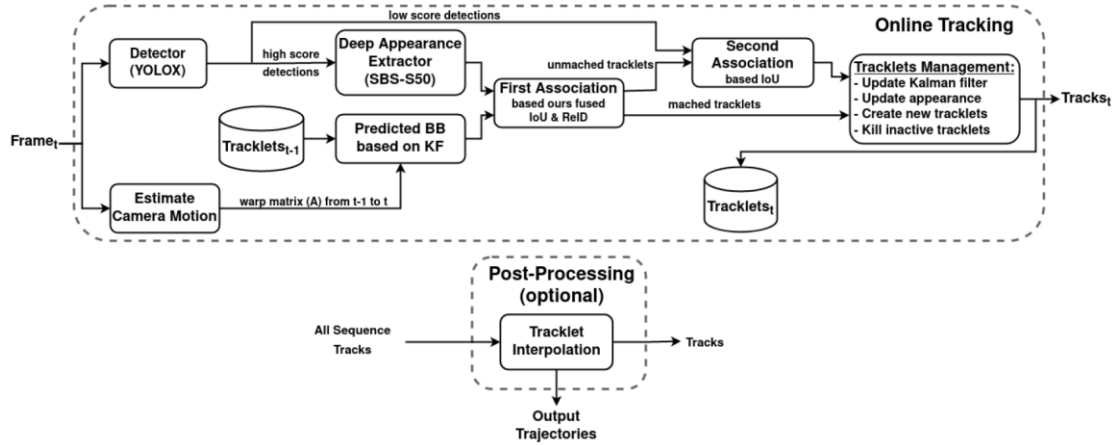


Figure 3.4: Overview of BoT-SORT pipeline (CMC and ReID).

### 3.2.3 Research Gap in Existing Research

Despite the advancements in MOT algorithms and their applications in agriculture, there is a notable gap in the literature regarding their use for nursery inventories with pine seedling tracking and counting. Most existing studies focus on fruit counting or livestock monitoring, with limited attention to seedlings, particularly pine seedlings, in nursery settings. The dense planting and uniform appearance of pine seedlings present unique challenges for tracking algorithms, such as seedling overlap, occlusion, and difficulty maintaining consistent object identities.

Furthermore, while algorithms like SORT, ByteTrack, and BoT-SORT have shown promise in other agricultural contexts, their performance in seedling counting under varying camera FOV has not been thoroughly investigated. Understanding how the camera's FoV affects tracking and counting accuracy is essential for optimizing

these algorithms for practical deployment in nurseries. This research aims to address these gaps by evaluating the performance of selected MOT algorithms—SORT, ByteTrack, and BoT-SORT—in tracking and counting pine seedlings. Additionally, the study investigates the impact of the camera's FoV on counting accuracy, providing insights that can inform the configuration of camera systems and the selection of appropriate tracking algorithms for efficient forest nursery inventory management during spring.

### **3.3 Methodology**

#### **3.3.1 *Experimental Setup***

As discussed in Chapter 2, the detection models employed were YOLOv8, YOLOv9, and YOLOv10. These models were used to detect pine seedlings in each frame of the 8 unseen videos in the test data set comprising 480. The bounding boxes provided by these detection models served as inputs for the tracking algorithms as part of the first step in the tracking-by-detection approach.

#### **3.3.2 *Implementation of Tracking Algorithms***

To track and count pine seedlings, we integrated three multi-object tracking algorithms with the detection models: SORT, ByteTrack, and BoT-SORT. Each algorithm was fine-tuned to optimize performance, with parameters determined through experimental tuning based on the dataset's characteristics, such as the uniform appearance and dense packing of seedlings.

SORT processed the bounding box outputs from the detection models. Key parameters were adjusted to ensure optimal results in the nursery environment. The *max\_age* parameter, representing the maximum number of frames a track can miss

before deletion, was set to 34 as the maximum number of frames taken by a seedling to move out of the scene, allowing consistent tracking. The *min\_hits* parameter, determining the minimum consecutive frames a detection must appear before being considered a valid track, was set to 4 to reduce false positives from sporadic detections while ensuring prompt initiation of valid tracks. The IoU threshold was set to 0.3, balancing leniency in matching detections with existing tracks and minimizing incorrect associations in the densely packed environment. These parameters were fine-tuned iteratively by evaluating the tracker's performance on validation data, focusing on metrics such as ID switches and tracking continuity.

ByteTrack employed a two-stage association strategy to incorporate both high- and low-confidence detections. Parameters were determined through iterative testing. The *track\_high\_thresh* was set to 0.5, prioritizing high-confidence detections during the first association round. The *track\_low\_thresh* was set to 0.3, allowing the inclusion of low-confidence detections in the second round to recover tracks that might otherwise be lost. The *new\_track\_thresh* was set to 0.3, enabling the initialization of new tracks for seedlings appearing at frame boundaries or reemerging after occlusions. The *track\_buffer*, determining how long inactive tracks are retained before deletion, was set to 34 to handle temporary detection gaps effectively. The *match\_thresh* was set to 0.6, optimizing the association between detections and existing tracks while minimizing ID switches. These parameters were carefully adjusted by evaluating ByteTrack on validation sequences with high object density and assessing metrics like tracking accuracy and ID consistency.

BoT-SORT enhanced tracking performance by addressing the challenges of dense and uniform seedling tracking. It utilized camera motion compensation through

the *sparseOptFlow* method, accounting for potential camera movement and stabilizing tracking in dynamic scenes. The *track\_high\_thresh* was set to 0.5 and the *track\_low\_thresh* to 0.1, balancing precision in the first association round with sensitivity to low-confidence detections in the second. The *new\_track\_thresh* was set to 0.6, requiring a higher confidence level for initiating new tracks to ensure reliability. The *track\_buffer* was set to 34, consistent with SORT and ByteTrack, allowing tracks to persist during brief occlusions or detection failures. The *match\_thresh* was set to 0.6 to refine the association between detections and existing tracks, minimizing ID switches in densely packed seedlings. The Re-Identification (ReID) module was disabled, as the uniform appearance of pine seedlings made appearance-based matching unnecessary. Instead, the tracker relied on spatial and motion-based cues, which proved more effective for this dataset. These parameters were determined through experiments, evaluating tracking metrics such as MOTA and IDF1, and tracking continuity to ensure BoT-SORT provided exceptional performance tailored to the nursery environment.

By fine-tuning these parameters, each tracking algorithm was optimized to address the specific challenges posed by the dense and uniform pine seedling dataset, enabling robust MOT in the nursery setup.

### **3.3.3 Counting Pine Seedlings**

After tracking seedlings across frames, each seedling was assigned a unique identifier (ID) for the duration it was visible. The total number of unique IDs encountered over the sequence was counted to determine the total number of seedlings. This approach helped count each seedling once in its lifetime. Additionally,

this method allowed us to evaluate the effects of HVA and VFoV, as counting was performed for each crop row of the bed with a change in VFoV.

### 3.3.4 *Impact of Horizontal Viewing Angles with Change in VFoV on Tracking and Counting*

The field of view defines the extent of the observable scene captured by the camera and is determined by the camera's focal length and sensor size. FoV is measured in degrees horizontally (Horizontal Field of View, HFoV) and vertically (Vertical Field of View, VFoV). While the imaging angles influence how the scene is captured from different viewpoints, FoV affects how much of the scene is captured.

As discussed in Chapter 2, HVA significantly affected the detection accuracy. In this chapter, we further investigated how horizontal viewing angles and VFoV influence tracking and counting accuracy. Figure 3.5 shows the horizontal and vertical field of view.

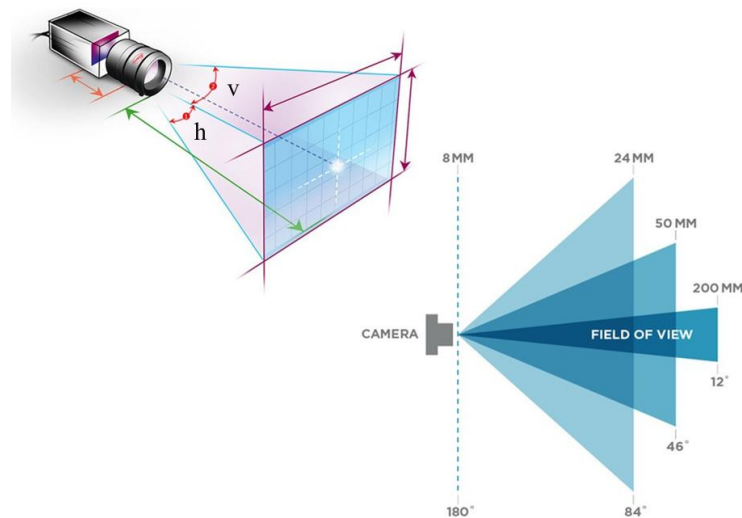


Figure 3.5: Horizontal (h) and vertical (v) field of view.

**Range of Horizontal Viewing Angles for crop rows**

Seedlings in different crop rows of the bed are located at varying horizontal viewing angles relative to the camera's optical center. We examined how changes in viewing angles affected tracking performance by analyzing seedlings located within specific horizontal angle ranges. The horizontal viewing angles corresponding to each crop row as shown in Figure 3.6 as follows: row1 (-52.77°, -44.96°), row2 (-45.63°, -35.64°), row3 (-32.83°, -19.8°), row4 (-7.63°, 2.3°), row5 (17.77°, 27.22°), row6 (35.13°, 44.21°), row7 (45.06°, 52.78°), and row8 (51.24°, 57.34°).

By analyzing tracking performance at these specific angles, we aimed to understand the impact of HVA on the counting accuracy of different crop rows.

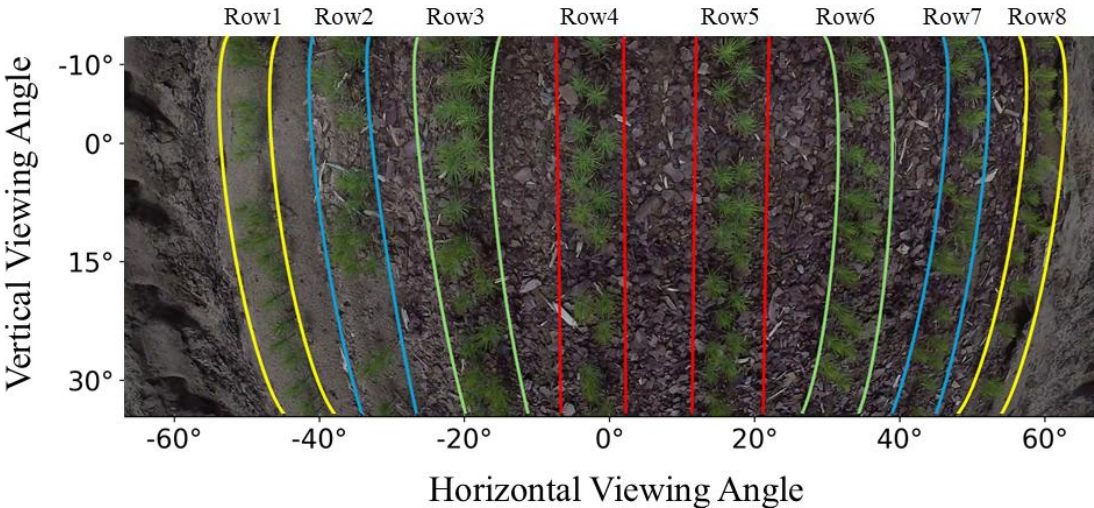


Figure 3.6: Rows (1-8) division based on HVA.

### Tracking and Counting in different VFoVs

We varied 31 VFoV starting from  $(-1.12^\circ, 1.12^\circ)$ ,  $(-2.24^\circ, 2.24^\circ)$ ,  $(-3.36, 3.36)$  up to  $(34.6^\circ, 34.6^\circ)$  to analyze how changes in the camera's vertical orientation affected tracking and counting. This analysis was conducted for seedlings in each crop row, considering their different HVA, to understand the combined effect of both variables. Figure 3.7 illustrates how the different VFoVs of a crop row (the crop row is located at HFoV of  $(-7.63^\circ, 2.3^\circ)$ ) were used to track and count pine seedlings.

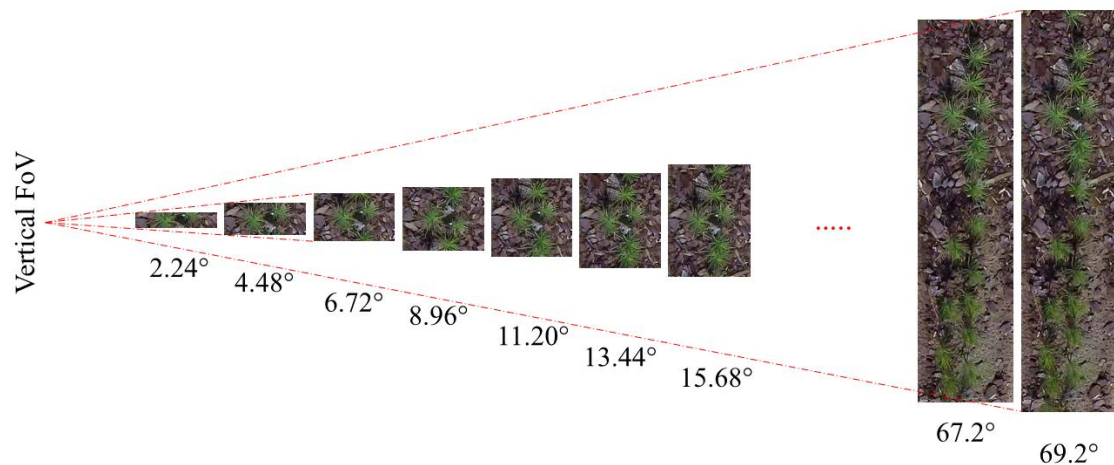


Figure 3.7: Increasing VFoV of a crop row.

By evaluating seedling tracking and counting at various VFoV configurations, we aimed to provide insights into the effects of the camera's VFoV on counting accuracy in nursery environments. This understanding is crucial for optimizing VFoV and HVA to achieve the most accurate and reliable counts, further providing insights into the optimal camera configurations.

### 3.3.5 Evaluation Metrics

To assess the performance of the tracking algorithms in counting pine seedlings, we used standardized MOT metrics (Milan et al., 2016; Stiefelhagen et al., 2007) and a counting accuracy measure.

#### 3.3.5.1 MOTA (Multi-Object Tracking Accuracy)

MOTA is the most widely used metric that measures the overall accuracy of both the tracking system, accounting for false negatives, false positives, and identity switches. It is defined as:

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDsw_t)}{\sum_t GT_t} \quad (8)$$

where  $t$  is the frame index,  $FN$  is False Negatives,  $FP$  is False Positives,  $IDsw$  is the number of Identity Switches, and  $GT$  is the number of ground truth objects.

Figure 3.8 illustrates four cases of tracker-to-target assignments, highlighting situations such as identity switches, fragmentation, and suboptimal tracking.

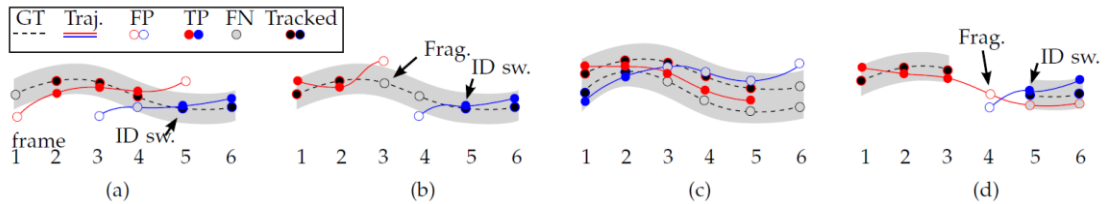


Figure 3.8: Four cases illustrating tracker-to-target assignments.

ID Switch (IDsw) (a) occurs when the assigned target changes from one track (red) to another (blue). Track Fragmentation (b) happens when tracking stops (frames 1–2), resumes later, and introduces a new track, causing both fragmentation and an ID switch. Suboptimal Tracking (c) propagated single-frame assignment leads to missed targets (FN) and false positives (FP). No fragmentations occur when interrupted tracking is not resumed. Degenerate Case (d) target re-identification fails, causing fragmentation and an unintuitive ID switch when a new assignment (blue) is made to avoid conflicts from earlier frames.

### 3.3.5.2 IDF1

IDF1 addresses the limitations of MOTA by focusing on how long the tracker correctly identifies an object. It focuses on the correctness of object identities over time:

$$IDF1 = \frac{2 * IDTP}{2 * IDTP + IDFP + IDFN} \quad (9)$$

where IDTP is the ID True Positive (correctly identified detections), IDFP is the ID False Positives (tracker predictions not matching any ground truth), and IDFN ID False Negatives (ground truth identities not tracked).

### 3.3.5.3 MOTP (Multi-Object Tracking Precision)

MOTP measures the average localization precision of the tracker by calculating the overlap between the predicted bounding boxes and the ground truth:

$$MOTP = \frac{\sum_{i,t} d_{t,i}}{\sum_t c_t} \quad (10)$$

where  $c_t$  denotes the total number of matches in frame  $t$  and  $d_{t,i}$  is the bounding box overlap of target  $i$  with its assigned ground truth object. MOTP quantifies the tracker's ability to localize objects accurately but provides limited insight into tracking consistency over time.

#### 3.3.5.4 Additional MOT Metrics

The evaluation of Multiple Object Tracking (MOT) performance was further enhanced by incorporating additional metrics to provide a comprehensive understanding of the tracking system's effectiveness. Ground Truth (GT) represents the total number of ground truth detections across all frames, serving as a baseline for assessing detection accuracy. Ground Truth Unique Identities (GTID) refers to the total number of unique object IDs present in the sequence, which is crucial for evaluating the system's ability to track individual objects over time consistently. Mostly Tracked (MT) counts the number of objects that were successfully tracked for more than 80% of their lifespan, indicating a high level of tracking reliability. Partially Tracked (PT) denotes the number of objects tracked between 20% and 80% of their lifespan, reflecting moderate tracking performance where some continuity is maintained. Lastly, Mostly Lost (ML) measures the number of objects that were tracked for less than 20% of their lifespan, highlighting instances where the tracking system struggled to maintain consistent object identification. Together, these metrics provide a detailed breakdown of the tracker's performance concerning object longevity and tracking accuracy.

### 3.3.6 Counting Accuracy

Beyond the standard MOT metrics, counting accuracy is critical for the specific task of counting seedlings. We defined counting accuracy using the error rate between the predicted count and the ground truth count:

$$error_{rate} = \left| \frac{(pred - gt)}{gt} \right| \quad (11)$$

$$accuracy = (1 - error_{rate}) * 100 \quad (12)$$

where *pred* is the predicted count of seedlings, and *gt* is the ground truth count. This metric measures the deviation of the predicted count from the actual count, providing insight into the accuracy of the tracking algorithm in performing counting tasks. An accuracy of 100% indicates perfect counting.

## 3.4 Results

The table 3.1 lists the manually counted ground truth of seedlings in each video sample, serving as a reference for evaluating MOT algorithms and counting accuracy.

Table 3.1: Ground truth count of seedlings in each video sample used for evaluation.

Video sample ID	Seedling count (Ground Truth)
1	244
2	250
3	249
4	223
5	234
6	218

### 3.4.1 *Multi-Object Tracking Results*

An effective tracking algorithm is essential to accurately count pine seedlings. To this end, we evaluated the performance of various tracking algorithms employed in this study by comparing the MOTA scores across different combinations of detection models, as illustrated in Figure 3.9. The analysis reveals that BoT-SORT consistently outperforms the other two tracking algorithms across all models, maintaining a MOTA score predominantly above 80. In stark contrast, ByteTrack and SORT exhibit significantly lower performance, with MOTA scores generally ranging between 20 to 30 for SORT and 40 to 60 for ByteTrack. These findings highlight the superior reliability of BoT-SORT in tracking pine seedlings across diverse conditions and model architectures.

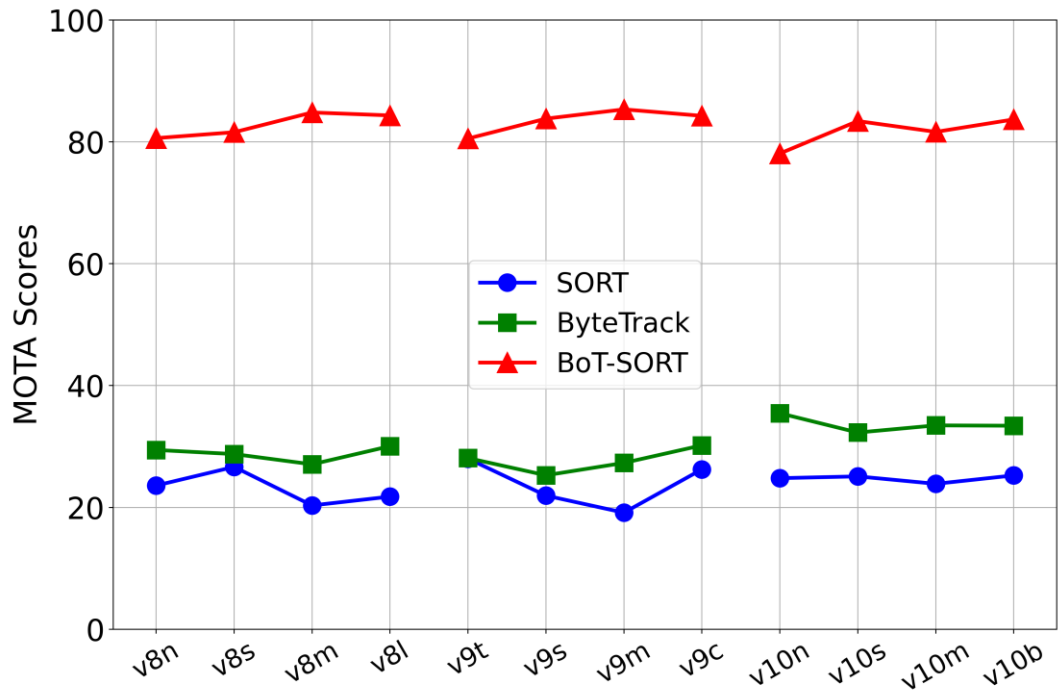


Figure 3.9: MOTA scores across trackers with varying detection models.

The comparative results of the multi-object tracking algorithms on the test set are summarized in Table 3.2. The BoT-SORT algorithm demonstrates superior performance to other algorithms across nearly all evaluated metrics, with the exception of the MOTP, which primarily measures the precision of object localization rather than tracking accuracy. BoT-SORT performed well in maintaining the least number of mostly lost (ML) objects and object IDswitch and maintained an acceptable level of tracking accuracy as measured by the MOTA. On the other hand, ByteTrack accounted for mostly lost objects and higher ID switches compared to SORT, which also suffered the same problem. This performance indicates robustness across different architectural detection models, emphasizing BoT-SORT's effectiveness in consistent object tracking.

Table 3.2: Comparison of optimal detection models by architecture based on MOTA with three tracking algorithms.

Model	Tracker	IDF1	MT	PT	ML	IDsw	MOTP	MOTA
v8nano	SORT	0.437	109	128	215	374	75.71	23.59
	ByteTrack	0.39	159	190	103	242	73.10	29.43
	BoT-SORT	0.886	358	68	26	47	75.94	80.61
v9small	SORT	0.384	73	159	220	427	76.60	21.94
	ByteTrack	0.324	137	221	94	241	73.39	25.23
	BoT-SORT	0.902	366	66	20	<b>31</b>	76.73	83.81
v9med	SORT	0.35	83	117	252	426	<b>77.08</b>	19.14
	ByteTrack	0.343	154	213	85	291	73.5	27.31
	BoT-SORT	<b>0.909</b>	<b>375</b>	<b>58</b>	19	42	76.92	<b>85.33</b>
v8large	SORT	0.379	75	144	233	463	76.09	21.77
	ByteTrack	0.371	162	223	67	288	73.12	30.05
	BoT-SORT	0.906	367	68	<b>17</b>	49	76.15	84.34

BoT-SORT consistently maintained track continuity, as Figure 3.10 illustrates its performance in one of the middle crop rows, whereas it faced challenges with trajectory swaps in peripheral rows, as shown in Figure 3.11. The significance of colors is as follows: red color (switching or assigning a new track (ID)), green (maintaining track), and black (missing track). This occasional inconsistency could compromise overall counting accuracy in farther rows. In contrast, ByteTrack and SORT exhibited decent performances. Figure 3.12 shows how ByteTrack faced significant issues with identity switches in the middle row. SORT performance in Figure 3.13 shows how SORT struggled in complex interactions and rapid movements, leading to frequent identity switches and, overall, less reliable tracking in dense environments by ByteTrack and SORT.

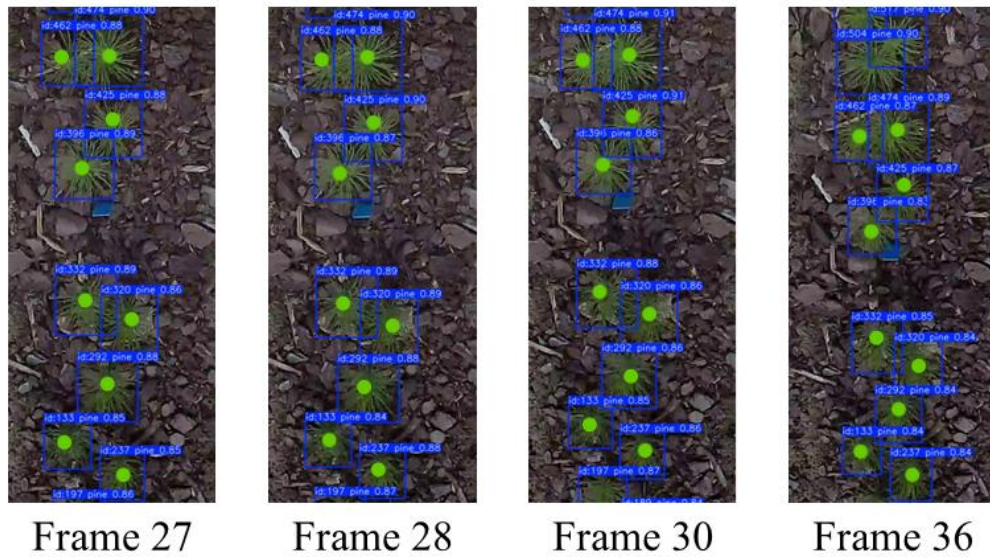


Figure 3.10: BoT-SORT’s performance in the central row across multiple frames. The green points indicate the continuity of the track ID in subsequent frames.

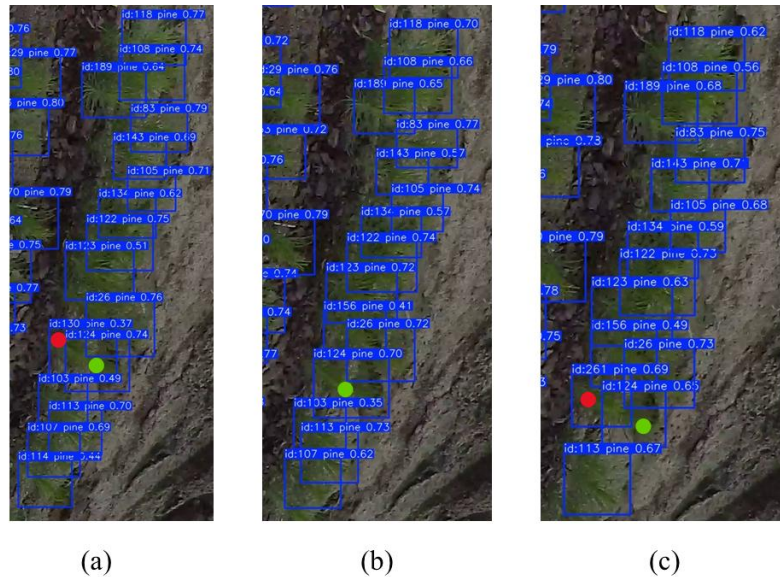


Figure 3.11: BoT-SORT is switching ID from (a) to (c) after being lost in (b) in consecutive frames, where red color denotes the switching or re-assigning of a track ID, green denotes maintaining continuity, and black refers to a missing track.

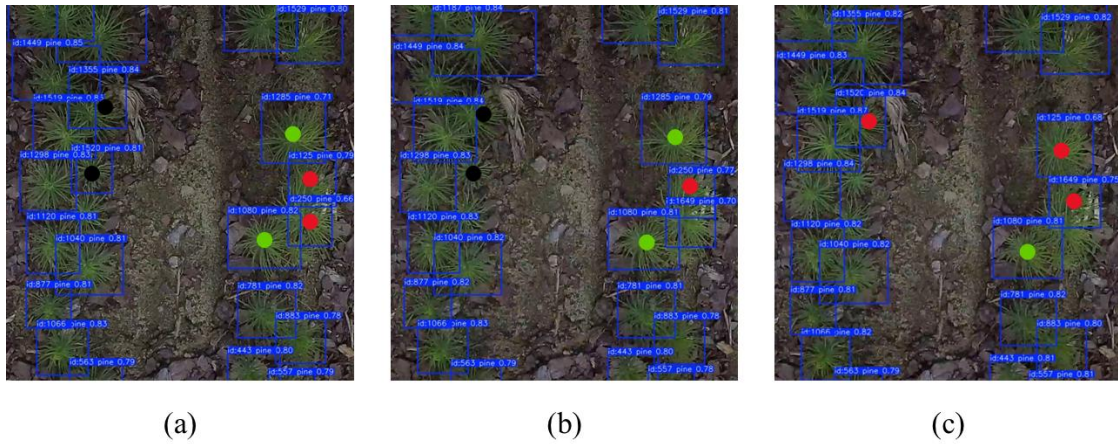


Figure 3.12: ByteTrack loses tracks in (a) and (b) and assigns an existing ID to a new track in (c) in consecutive frames, where red color denotes the switching or re-assigning of a track ID, green denotes maintaining continuity, and black refers to a missing track.

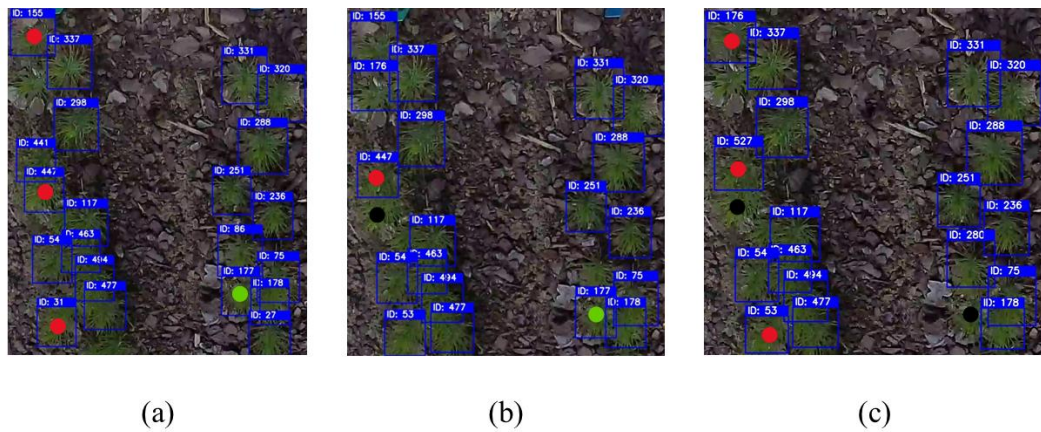


Figure 3.13: SORT switching and re-assigning IDs in consecutive frames, where red color denotes the switching or re-assigning of a track ID, green denotes maintaining continuity, and black refers to a missing track.

### 3.4.2 Effect of Imaging Angles and Field of View on Counting Accuracy

To evaluate how VFoV affects counting accuracy, we analyzed how changing the VFoV impacted counting accuracy for each model-tracker combination. Figure

3.14 and 3.15 show counting accuracy at various horizontal viewing angles with varying VFoV for the combination of the detection models having higher and lower MOTA and the three tracking algorithms (SORT, ByteTrack, BoT-SORT).

With increasing VFoV, tracking performance improved up to a certain point before plateauing with BoT-SORT, whereas SORT and ByteTrack continued to increase till the end. The counting accuracy is very low at smaller VFoV due to the minimal movement of the seedling within that view. BoT-SORT showed the most robust performance, maintaining high and consistent accuracy. SORT and ByteTrack struggled to maintain consistent accuracy at extreme VFoVs but managed to reach close to the red-colored threshold by the end in a few rows. From this data, we can conclude that the change in the horizontal viewing angle with VFoV affects tracking and counting accuracy.

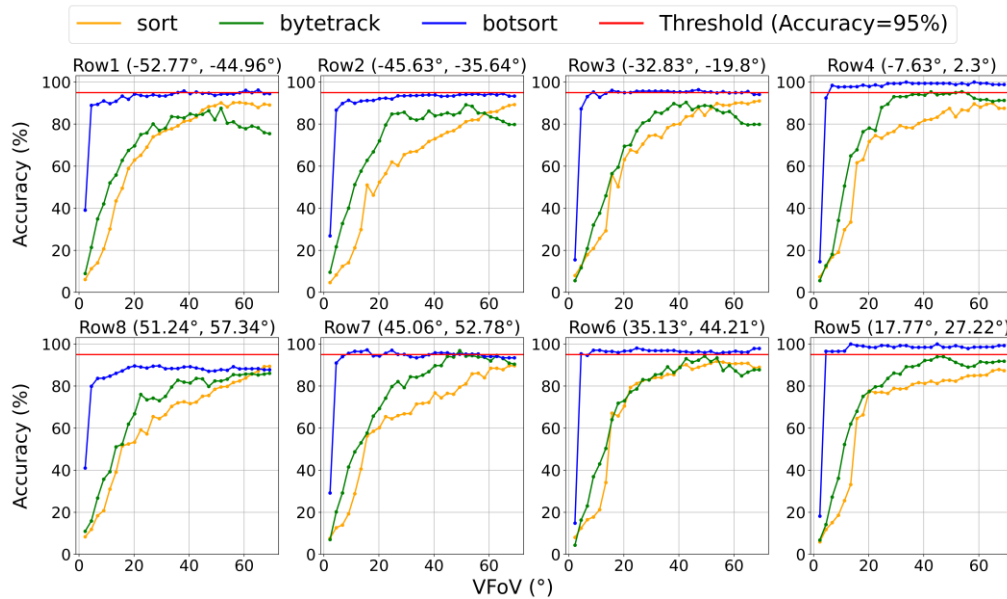


Figure 3.14: Counting accuracy at various VFoV across 8 crop rows using YOLOv9medium and the three trackers.

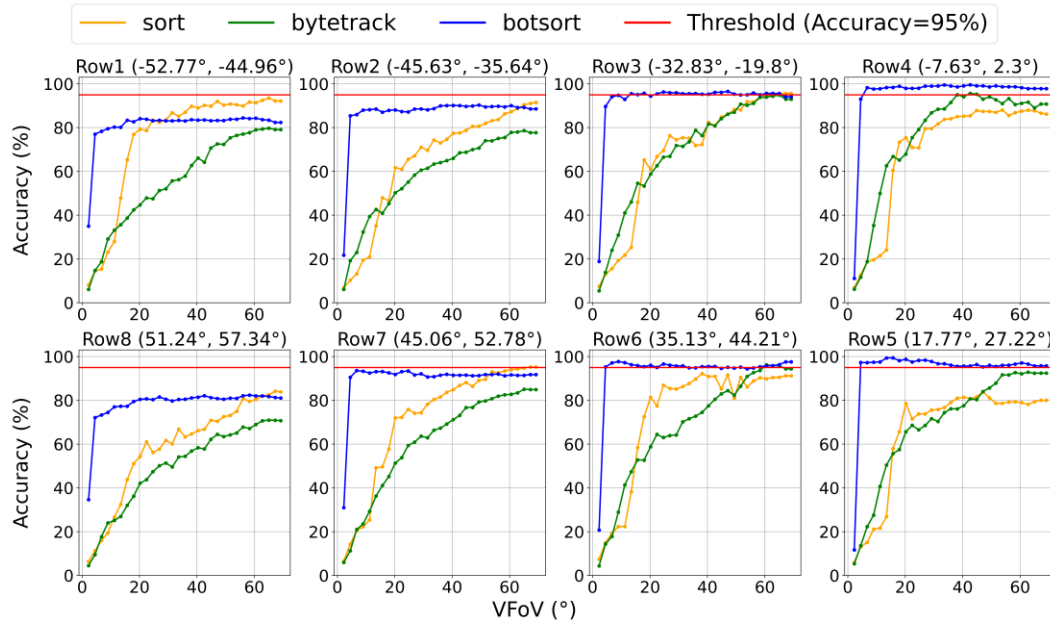


Figure 3.15: Counting accuracy at various VFoV across 8 crop rows using YOLOv8nano and the three trackers.

The average counting accuracies of the crop rows after the first VFoV with best-performing combination of YOLOv9medium and BoT-SORT are row1 (93.6%), row2 (92.77%), row3 (94.88%), row4 (98.7%), row5 (98.5%), row6 (96.5%), row7 (94.8%), and row8 (87.6%). The highest accuracy above 98% was obtained in the center rows 4 and 5. The HVA of these 2 rows is -8 to 28, accounting for approximately 30 degrees horizontally.

### 3.5 Discussion and Conclusion

#### 3.5.1 Discussion

The results presented in this study demonstrate the effectiveness of different multi-object tracking algorithms and their interaction with various detection models

for counting pine seedlings in a nursery environment. The evaluation highlights BoT-SORT's superiority over ByteTrack and SORT, particularly in challenging conditions such as dense and similar seedlings.

### **3.5.1.1 Performance of MOT Algorithms**

The evaluation of MOT algorithms based on MOTA scores highlights BoT-SORT's superior tracking capabilities. As evidenced in Table 3.1 and Figure 3.9, BoT-SORT consistently outperformed other trackers, achieving high MOTA scores across all detection models. For example, paired with YOLOv9medium, BoT-SORT attained a MOTA of 85.33 with only 19 missed tracks and 42 IDsw, demonstrating robust track continuity even in cluttered environments.

In contrast, SORT and ByteTrack underperformed, with MOTA scores just above 25 for SORT and below 40 for ByteTrack, indicating poor reliability in complex interactions and rapid seedling movements. The failure of SORT and ByteTrack aligns with the study by Gené-Mola et al. (2023) who reported MOTA scores of 64 for SORT and 68 for ByteTrack while counting apples in videos. Their relatively better performance likely stems from apples being more spatially separated compared to the pine seedlings in our study, where frequent identity switches presented a greater challenge.

BoT-SORT's superior performance can be attributed to its integration of advanced camera motion compensation techniques, enabling more stable and consistent tracking by mitigating the effects of dense object environments and any potential overlaps. These findings underscore the importance of selecting specialized trackers for forest nurseries where the movement of seedlings in the videos affects the tracking in high-density areas.

### **3.5.1.2 Effect of Field of View on Tracking and Counting Accuracy**

The relationship between VFoV and counting accuracy reveals that larger VFoVs enhance counting accuracy, particularly when paired with YOLOv9medium. This improvement is likely due to the increased visibility of seedlings, enabling better differentiation of seedling movement. The decline in counting accuracy at smaller VFoVs can be attributed to limited object movement, which hinders the tracker's ability to distinguish individual seedlings effectively. BoT-SORT's performance plateaued after a few VFoVs, while SORT and ByteTrack exhibited gradual improvements, reaching competitive accuracies only at extreme VFoVs. This pattern can be linked to frequent ID switching, as reflected in their low MOTA scores of SORT and ByteTrack.

The observed average counting accuracies with YOLOv9medium and BoT-SORT after a VFoVs, with central rows (rows 4 and 5) achieving above 98% accuracy, align with Wang et al. (2021), who reported similar accuracy for counting corn seedlings (98%). This similarity suggests that BoT-SORT effectively handles densely packed pine seedlings located near 0° HVA. Furthermore, our results surpass those of Jiang et al. (2019), who achieved 93% accuracy for cotton seedlings, likely due to the lower seedling density in their study, which presented fewer tracking challenges and experienced false positives due to the use of a simple Kalman filter for tracking. These findings underscore the necessity of employing advanced trackers like BoT-SORT for handling early-stage pine seedlings, with 2 crop rows near 0° HVA and a VFoV of 69.2° as used in this study.

These findings demonstrate the robustness of modern detection models like YOLOv9 and YOLOv10 and MOT algorithms such as BoT-SORT. By understanding the effects of HVA and VFoV in managing inventory in nursery environments, this

study highlights the potential of automating early-stage pine seedling counting, streamlining nursery management practices with optimal camera placement (e.g., four cameras), and reducing manual labor.

### **3.5.2 *Limitations and Future Research***

While this study offers valuable insights into the application of advanced tracking algorithms like BoT-SORT for nursery management, several limitations must be acknowledged. The computational demands of BoT-SORT may present challenges for deployment on edge devices with limited processing capabilities, potentially restricting its accessibility for some users.

Future research should explore optimizing detection models and tracking algorithms for lower computational environments, particularly focusing on edge devices such as the Jetson AGX Orin to enable real-time applications. This involves adapting algorithms to leverage the parallel processing capabilities and efficient power usage of these platforms, potentially through techniques like model compression, quantization, and the use of optimized libraries. Additionally, investigating the integration of cost-effective camera technologies remains essential. Expanding the study to include diverse nursery settings could enhance the generalizability of the findings. By addressing these areas, future studies can build upon the current work to develop more accessible and efficient solutions for automated seedling management.

### **3.5.3 *Conclusion***

This chapter integrated MOT tracking algorithms with a detection-by-tracking paradigm. The findings demonstrate that both the camera's field of view and the choice of tracking algorithm significantly impact the accuracy of tracking and

counting pine seedlings in nursery environments. By systematically evaluating the effects of HVA and VFoV and comparing the performance of different tracking algorithms, we have identified BoT-SORT as the most effective algorithm for this application. However, the tracking method demonstrates a small average overestimation compared to the ground truth seedling counts, with a slight variability as reflected in the limits of agreement. While the method shows promise for automated counting, its current variability may affect precision in practical settings. Optimizing camera placement and configuration, along with employing robust tracking algorithms, can substantially enhance the accuracy and reliability of automated seedling counting systems. Future efforts should focus on reducing variability in tracking performance and aligning tracked counts more closely with ground truth values. These advancements have the potential to improve nursery management practices, supporting more efficient reforestation efforts.

## Chapter 4

### CONCLUSION AND FUTURE WORK

#### 4.1 Conclusion

The accurate counting of bareroot pine seedlings in forest nurseries is a critical task for effective nursery management and successful reforestation efforts. In this thesis, we developed an automated counting system tailored to spring inventory conditions and demonstrated that HVA significantly influences detection, tracking, and counting performance. By systematically examining imaging angles and fields of view, we showed that a single-camera solution, while potentially more cost-effective, faces inherent viewing angle and perspective limitations. In contrast, implementing multiple cameras—such as four per target area—improves accuracy by covering diverse viewing angles, although this approach increases both cost and computational and hardware demands.

Our best-performing detection model, YOLOv10balanced, achieved a mAP50 of 0.955, and the BoT-SORT algorithm emerged as the most reliable multi-object tracker, delivering precise seedling counts. However, as viewing angles shifted from the nadir to more oblique perspectives, detection accuracy declined, underscoring the critical role of camera placement and orientation. Ultimately, the findings highlight that to achieve optimal counting accuracy as high as 98.5% for nursery inventory, four cameras are required, each focusing on only two crop rows.

By establishing these relationships between camera angles, the number of cameras, and detection accuracy, this work provides a framework for developing

optimized, automated inventory solutions for spring that can be readily adapted to varying forest nursery conditions. Such insights will guide future implementations and refinements, enabling forest nurseries to adopt more efficient, precise, and scalable inventory counting systems to support sustainable forestry initiatives.

## **4.2 Implications for Nursery Management**

The successful implementation of an automated counting system offers several critical advantages for forest nursery management. Enhanced accuracy provides nursery managers with more reliable data, directly improving decision-making and enabling more effective resource allocation. Additionally, timely and precise counts facilitate early detection of seedling losses, helping identify planter issues or damage caused by wildlife and diseases before they escalate. By reducing manual counting, automation also improves operational efficiency, lowering labor costs and enhancing profitability. These combined benefits establish a robust foundation for sustainable growth and responsive management practices within nursery operations.

## **4.3 Limitations and Challenges**

Several limitations of this study should be noted. The consistent vehicle speed (0.5 m/s) used during data collection may not accurately represent the variable operational speeds encountered in different nursery environments. Moreover, relying on data from a single nursery restricts the applicability of the findings across diverse soil types, lighting conditions, and seedling varieties. The focus on a single growth stage of loblolly pine seedlings further constrains the scope, as different species and growth stages may present unique detection and tracking challenges. Finally, downscaling due to computational constraints, imagery may lose the fine-grained

features, potentially affecting performance when detecting smaller seedlings. These constraints underscore the need for broader testing and methodological adaptations before implementing automated counting systems in a wider range of operational scenarios.

#### **4.4 Recommendations For Future Research**

Future research should focus on expanding the dataset to include multiple nurseries and more diverse environmental conditions, thereby strengthening model robustness and enhancing generalizability. Implementing stable, controlled lighting conditions would help minimize shadows and improve detection performance across varying field scenarios. To support practical deployment, researchers should employ high-performance GPUs (e.g., Jetson AGX Orin) and optimize models for efficiency and speed. Additionally, investigating affordable camera and hardware solutions can make advanced automated systems more accessible to nursery operations, reducing costs while maintaining performance. Taken together, these recommendations guide ongoing efforts to refine and scale automated counting systems, ensuring broader applicability and operational viability in real-world nursery environments.

## REFERENCES

- Agarwal, S., Terrail, J. O. Du, & Jurie, F. (2018). *Recent Advances in Object Detection in the Age of Deep Convolutional Neural Networks*.  
<https://arxiv.org/abs/1809.03193v2>
- Aharon, N., Orfaig, R., & Bobrovsky, B. (2022). BoT-SORT: Robust Associations Multi-Pedestrian Tracking. *ArXiv.Org*.  
<https://doi.org/10.48550/ARXIV.2206.14651>
- Al, M., Alif, R., & Hussain, M. (2024). *YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain*.  
<http://arxiv.org/abs/2406.10139>
- Ariza-Sentís, M., Valente, J., Kooistra, L., Kramer, H., & Múcher, S. (2023). Estimation of spinach (*Spinacia oleracea*) seed yield with 2D UAV data and deep learning. *Smart Agricultural Technology*, 3, 100129.  
<https://doi.org/10.1016/J.ATECH.2022.100129>
- Berndt, J., Meißner, H., & Kraft, T. (2023). ON THE ACCURACY OF YOLOV8-CNN REGARDING DETECTION OF HUMANS IN NADIR AERIAL IMAGES FOR SEARCH AND RESCUE APPLICATIONS. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 48(1/W2-2023), 139–146.  
<https://doi.org/10.5194/ISPRS-ARCHIVES-XLVIII-1-W2-2023-139-2023>
- Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple Online and Realtime Tracking. *Proceedings - International Conference on Image Processing, ICIP, 2016-August*, 3464–3468.  
<https://doi.org/10.1109/ICIP.2016.7533003>
- Bidese-Puhl, R., Bao, Y., Payne, N. D., Stokes, T. A., Nadel, R. L., & Enebak, S. A. (2023). In-Field Pine Seedling Counting Using End-to-End Deep Learning for Inventory Management. *Journal of the ASABE*, 66(2), 469–477.  
<https://doi.org/10.13031/JA.15383>
- Bradski, G., & Kaehler, A. (2000). OpenCV. *Dr. Dobb's Journal of Software Tools.*, 2, 122–125.

- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12346 LNCS, 213–229. [https://doi.org/10.1007/978-3-030-58452-8\\_13](https://doi.org/10.1007/978-3-030-58452-8_13)
- Carolyn C. Pike, Diane L. Haase, Scott Enebak, Annakay Abrahams, Elizabeth Bowersock, Lori Mackey, Zhao Ma, & Jim Warren. (2023). Forest Nursery Seedling Production in the United States—Fiscal Year 2022. *Tree Planters' Notes*, 66(2 (Fall)), 73–80. <https://rngr.net/publications/tpn/66-2/forest-nursery-seedling-production-in-the-united-states2014fiscal-year-2022>
- Chittupalli, S., & Chang, Y. K. (2022). *Counting of Strawberries and Flowers in Fields using YOLOv4 and SORT*. 1-. <https://doi.org/10.13031/AIM.202200229>
- Cui, J., Zheng, H., Zeng, Z., Yang, Y., Ma, R., Tian, Y., Tan, J., Feng, X., & Qi, L. (2023). Real-time missing seedling counting in paddy fields based on lightweight network and tracking-by-detection algorithm. *Computers and Electronics in Agriculture*, 212, 108045. <https://doi.org/10.1016/J.COMPAG.2023.108045>
- Dai, J., Li, Y., He, K., & Sun, J. (2016). *R-FCN: Object Detection via Region-based Fully Convolutional Networks*. 120, 122–125. <https://github.com/daijifeng001/r-fcn>.
- Dhanya, V. G., Subeesh, A., Kushwaha, N. L., Vishwakarma, D. K., Nagesh Kumar, T., Ritika, G., & Singh, A. N. (2022). Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture*, 6, 211–229. <https://doi.org/10.1016/j.aiaa.2022.09.007>
- Fargione, J., Haase, D. L., Burney, O. T., Kildisheva, O. A., Edge, G., Cook-Patton, S. C., Chapman, T., Rempel, A., Hurteau, M. D., Davis, K. T., Dobrowski, S., Enebak, S., De La Torre, R., Bhuta, A. A. R., Cabbage, F., Kittler, B., Zhang, D., & Guldin, R. W. (2021). Challenges to the Reforestation Pipeline in the United States. *Frontiers in Forests and Global Change*, 4. <https://doi.org/10.3389/FFGC.2021.629198/BIBTEX>
- Farjon, G., Huijun, L., & Edan, Y. (2023). Deep-learning-based counting methods, datasets, and applications in agriculture: a review. *Precision Agriculture*, 24(5), 1683–1711. <https://doi.org/10.1007/S11119-023-10034-8/FIGURES/4>
- Gené-Mola, J., Felip-Pomés, M., Net-Barnés, F., Morros, J. R., Miranda, J. C., Arnó, J., Asín, L., Lordan, J., Ruiz-Hidalgo, J., & Gregorio, E. (2023). Video-Based Fruit Detection and Tracking for Apple Counting and Mapping. *2023 IEEE International Workshop on Metrology for Agriculture and Forestry*,

- MetroAgriFor 2023 - Proceedings*, 301–306.  
<https://doi.org/10.1109/METROAGRIFOR58484.2023.10424135>
- Girshick, R. (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–587. <https://doi.org/10.1109/CVPR.2014.81>
- Glenn Jocher et al. (2023). *YOLOv8*. <https://github.com/ultralytics/ultralytics>.
- Glenn Jocher, Jing Qiu, & Ayush Chaurasia. (2023). *Ultralytics YOLO*. <https://github.com/ultralytics/ultralytics>
- Green, P. C., & Burkhart, H. E. (2020). Plantation Loblolly Pine Seedling Counts with Unmanned Aerial Vehicle Imagery: A Case Study. *Journal of Forestry*, 118(5), 487–500. <https://doi.org/10.1093/jofore/fvaa020>
- Guo, L., Wang, C., Yang, W., Wang, Y., & Wen, B. (2023). Boundary-Aware Divide and Conquer: A Diffusion-based Solution for Unsupervised Shadow Removal. *Proceedings of the IEEE International Conference on Computer Vision*, 12999–13008. <https://doi.org/10.1109/ICCV51070.2023.01199>
- Harakannavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., & Pramodhini, R. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings*, 3(1), 305–310. <https://doi.org/10.1016/J.GLTP.2022.03.016>
- Hu, N., Su, D., Wang, S., Nyamsuren, P., Qiao, Y., Jiang, Y., & Cai, Y. (2022). LettuceTrack: Detection and tracking of lettuce for robotic precision spray in agriculture. *Frontiers in Plant Science*, 13, 1003243. <https://doi.org/10.3389/FPLS.2022.1003243/BIBTEX>
- Huang, J., & Li, T. (2024). Small Object Detection by DETR via Information Augmentation and Adaptive Feature Fusion. *ACM International Conference Proceeding Series*, 39–44. <https://doi.org/10.1145/3664524.3675362>
- Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., & Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January*, 3296–3305. <https://doi.org/10.1109/CVPR.2017.351>

- Jiang, Y., Li, C., Paterson, A. H., & Robertson, J. S. (2019). DeepSeedling: Deep convolutional network and Kalman filter for plant seedling detection and counting in the field. *Plant Methods*, *15*(1), 1–19. <https://doi.org/10.1186/S13007-019-0528-3/FIGURES/9>
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35–45. <https://doi.org/10.1115/1.3662552>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. <https://doi.org/10.1016/J.COMPAG.2018.02.016>
- Karbouj, B., Topalian-Rivas, G. A., & Krüger, J. (2024). Comparative Performance Evaluation of One-Stage and Two-Stage Object Detectors for Screw Head Detection and Classification in Disassembly Processes. *Procedia CIRP*, *122*, 527–532. <https://doi.org/10.1016/J.PROCIR.2024.01.077>
- Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*(1–2), 83–97. <https://doi.org/10.1002/NAV.3800020109>
- Law, H., & Deng, J. (2018). CornerNet: Detecting Objects as Paired Keypoints. *International Journal of Computer Vision*, *128*(3), 642–656. <https://doi.org/10.1007/s11263-019-01204-1>
- Li, P. (2023). A Comparative Analysis of Single-Stage Detectors from the Perspectives of Anchor-Free and Anchor-Based Approaches. *Highlights in Science, Engineering and Technology*, *72*, 774–782. <https://doi.org/10.54097/S37SGQ58>
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors 2018, Vol. 18, Page 2674*, *18*(8), 2674. <https://doi.org/10.3390/S18082674>
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS(PART 5)*, 740–755. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2015). SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*

*Bioinformatics*), 9905 LNCS, 21–37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)

- Liu, Z., Yin, H., Wu, X., Wu, Z., Mi, Y., & Wang, S. (2021). From Shadow Generation to Shadow Removal. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4925–4934. <https://doi.org/10.1109/CVPR46437.2021.00489>
- Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T. K. (2021). Multiple object tracking: A literature review. *Artificial Intelligence*, 293, 103448. <https://doi.org/10.1016/J.ARTINT.2020.103448>
- Milan, A., Leal-Taixe, L., Reid, I., Roth, S., & Schindler, K. (2016). *MOT16: A Benchmark for Multi-Object Tracking*. <https://arxiv.org/abs/1603.00831v2>
- Minhui, L., Shamshiri, R. R., Schirrmann, M., & Weltzien, C. (2021). Impact of Camera Viewing Angle for Estimating Leaf Parameters of Wheat Plants from 3D Point Clouds. *Agriculture 2021, Vol. 11, Page 563, 11(6)*, 563. <https://doi.org/10.3390/AGRICULTURE11060563>
- Monteiro, A., Santos, S., & Gonçalves, P. (2021). Precision Agriculture for Crop and Livestock Farming—Brief Review. *Animals : An Open Access Journal from MDPI*, 11(8), 2345. <https://doi.org/10.3390/ANI11082345>
- Oswalt, S. N., Smith, W. B., Miles, P. D., & Pugh, S. A. (2019). *Forest Resources of the United States, 2017 A Technical Document Supporting the Forest Service 2020 RPA Assessment*. <https://doi.org/10.2737/WO-GTR-97>
- Paine, D. P., & Kiser, J. D. (2012). Small Format Aerial Imagery. *Aerial Photography and Image Interpretation*, 243–255. <https://doi.org/10.1002/9781118110997.CH13>
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A large and persistent carbon sink in the world's forests. *Science*, 333(6045), 988–993. [https://doi.org/10.1126/SCIENCE.1201609/SUPPL\\_FILE/PAPV2.PDF](https://doi.org/10.1126/SCIENCE.1201609/SUPPL_FILE/PAPV2.PDF)
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning

Library. *Advances in Neural Information Processing Systems*, 32.  
<https://arxiv.org/abs/1912.01703v1>

- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.  
<https://doi.org/10.1109/TPAMI.2016.2577031>
- Rousseau, R. J., Roberts, S. D., & Herrin, B. L. (2012). Comparison of advanced genetic loblolly pine planting stock. In: *Butnor, John R., Ed. 2012. Proceedings of the 16th Biennial Southern Silvicultural Research Conference. e-Gen. Tech. Rep. SRS-156. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 269-272., 156, 269–272.*  
<https://research.fs.usda.gov/treearch/41462>
- Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOsmanov, Kruchinin, D., Zankevich, A., DmitriySidnev, Markelov, M., Johannes222, Chenuet, M., a-andre, telenachos, Melnikov, A., Kim, J., Ilouz, L., Glazov, N., ... Truong, T. (2020). *opencv/cvat: v1.1.0*. Zenodo.  
<https://doi.org/10.5281/zenodo.4009388>
- Shen, L., Su, J., He, R., Song, L., Huang, R., Fang, Y., Song, Y., & Su, B. (2023). Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLOv5s. *Computers and Electronics in Agriculture*, 206, 107662.  
<https://doi.org/10.1016/J.COMPAG.2023.107662>
- Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing 2020, Vol. 12, Page 3136, 12(19)*, 3136. <https://doi.org/10.3390/RS12193136>
- Stiefelhagen, R., Bernardin, K., Bowers, R., Garofolo, J., Mostefa, D., & Soundararajan, P. (2007). The CLEAR 2006 Evaluation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4122 LNCS, 1–44.  
[https://doi.org/10.1007/978-3-540-69568-4\\_1](https://doi.org/10.1007/978-3-540-69568-4_1)
- Tian, Z., Shen, C., Chen, H., & He, T. (2019). FCOS: Fully Convolutional One-Stage Object Detection. *Proceedings of the IEEE International Conference on*

- Computer Vision, 2019-October*, 9626–9635.  
<https://doi.org/10.1109/ICCV.2019.00972>
- Wakeley, P. C. (1954). Planting the southern pines. *Agriculture Monograph No. 18, 18*. <https://research.fs.usda.gov/treesearch/57741>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-Time End-to-End Object Detection*. <http://arxiv.org/abs/2405.14458>
- Wang, C. Y., Mark Liao, H. Y., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2019). CSPNet: A New Backbone that can Enhance Learning Capability of CNN. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2020-June*, 1571–1580.  
<https://doi.org/10.1109/CVPRW50498.2020.00203>
- Wang, C.-Y., & Liao, H.-Y. M. (2024). *YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems*.  
<https://doi.org/10.1561/116.20240058>
- Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*.  
<http://arxiv.org/abs/2402.13616>
- Wang, L., Xiang, L., Tang, L., & Jiang, H. (2021). A Convolutional Neural Network-Based Method for Corn Stand Counting in the Field. *Sensors 2021, Vol. 21, Page 507, 21(2)*, 507. <https://doi.org/10.3390/S21020507>
- Zhang, W., Wang, J., Liu, Y., Chen, K., Li, H., Duan, Y., Wu, W., Shi, Y., & Guo, W. (2022). Deep-learning-based in-field citrus fruit detection and tracking. *Horticulture Research, 9*. <https://doi.org/10.1093/HR/UHAC003>
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2021). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13682 LNCS*, 1–21.  
[https://doi.org/10.1007/978-3-031-20047-2\\_1](https://doi.org/10.1007/978-3-031-20047-2_1)
- Zhang, Z., Kayacan, E., Thompson, B., & Chowdhary, G. (2020). High precision control and deep learning-based corn stand counting algorithms for agricultural robot. *Autonomous Robots, 44(7)*, 1289–1302. <https://doi.org/10.1007/S10514-020-09915-Y/FIGURES/9>

Zheng, Z., Li, J., & Qin, L. (2023). YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. *Computers and Electronics in Agriculture*, 209, 107857. <https://doi.org/10.1016/J.COMPAG.2023.107857>