Computers and Electronics in Agriculture 232 (2025) 110057

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

In-field blueberry fruit phenotyping with a MARS-PhenoBot and customized BerryNet

Zhengkun Li [a], Rui Xu [a], Changying Li [a,*], Patricio Munoz [b], Fumiomi Takeda [c], Bruno Leme [b]

[a] Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL, USA
[b] Horticultural Science Department, University of Florida, Gainesville, FL, USA
[c] Berry Innovation LLC, Scottsdale, AZ, USA

ARTICLE INFO

Keywords:
Blueberry phenotyping
Yield
Maturity
Fruit compactness
Deep learning
Segment Anything Model (SAM)

ABSTRACT

Accurate blueberry fruit phenotyping, including yield, fruit maturity, and cluster compactness, is crucial for optimizing crop breeding and management practices. Recent advances in machine vision and deep learning have shown promising potential to automate phenotyping and replace manual sampling. This paper presented a robotic blueberry phenotyping system, called MARS-Phenobot, that collects data in the field and measures fruit-related phenotypic traits such as fruit number, maturity, and compactness. Our workflow comprised four components: a robotic multi-view imaging system for high-throughput data collection, a vision foundation model (Segment Anything Model, SAM) for mask-free data labeling, a customized BerryNet deep learning model for detecting blueberry clusters and segmenting fruit, as well as a post-processing module for estimating yield, maturity, and cluster compactness. A customized deep learning model, BerryNet, was designed for detecting fruit clusters and segmenting individual berries by integrating low-level pyramid features, rapid partial convolutional blocks, and BiFPN feature fusion. It outperformed other networks and achieved mean average precision (mAP50) of 54.9 % in cluster detection and 85.8 % in fruit segmentation with fewer parameters and fewer computation requirements. We evaluated the phenotypic traits derived from our methods and the ground truth on 26 individual blueberry plants across 17 genotypes. The results demonstrated that both the fruit count and cluster count extracted from images were strongly correlated with the yield. Integrating multi-view fruit counts enhanced yield estimation accuracy, achieving a Mean Absolute Percentage Error (MAPE) of 23.1 % and the highest R^2 value of 0.73, while maturity level estimations closely aligned with manual calculations, exhibiting a Mean Absolute Error (MAE) of approximately 5 %. Furthermore, two metrics related to fruit compactness were introduced, including cluster compactness and fruit distance, which could be useful for breeders to assess the machine and hand harvestability across genotypes. Finally, we evaluated the proposed robotic blueberry fruit phenotyping pipeline on eleven blueberry genotypes, proving the potential to distinguish the high-yield, early-maturity, and loose-clustering cultivars. Our methodology provides a promising solution for automated in-field blueberry fruit phenotyping, potentially replacing labor-intensive manual sampling. Furthermore, this approach could advance blueberry breeding programs, precision management, and mechanical/robotic harvesting.

1. Introduction

In 2021, the U.S. blueberry industry produced 774.1 million pounds of blueberries, valued at approximately $1.1 billion (Morgan, 2022). Phenotypic traits such as yield, maturity, and cluster compactness are crucial for blueberry breeders and producers to make informed decisions in selecting genotypes with desirable machine-harvestable traits such as uniform maturity and loose clusters, as well as for growers to estimate the yield and plan harvest schedules (Ni et al., 2020). Blueberries, which tend to grow in clusters, are often obscured by leaves and nearby berries, making visual detection of all berries challenging. The non-uniform maturity of blueberries requires strategic harvest timing based on fruit maturity, which is crucial for harvesting the fruit at the proper maturity stage and for managing and projecting the size of picking crews to maximize harvest efficiency and profitability (Yang et al., 2012).

There are two primary approaches to measuring in-field phenotypic traits: the direct method, which involves sampling of a few plants or clusters to estimate the overall characteristics of plots or the entire field;

* Corresponding author.
E-mail address: cli2@ufl.edu (C. Li).

https://doi.org/10.1016/j.compag.2025.110057
Received 4 October 2024; Received in revised form 4 November 2024; Accepted 28 January 2025
Available online 7 February 2025

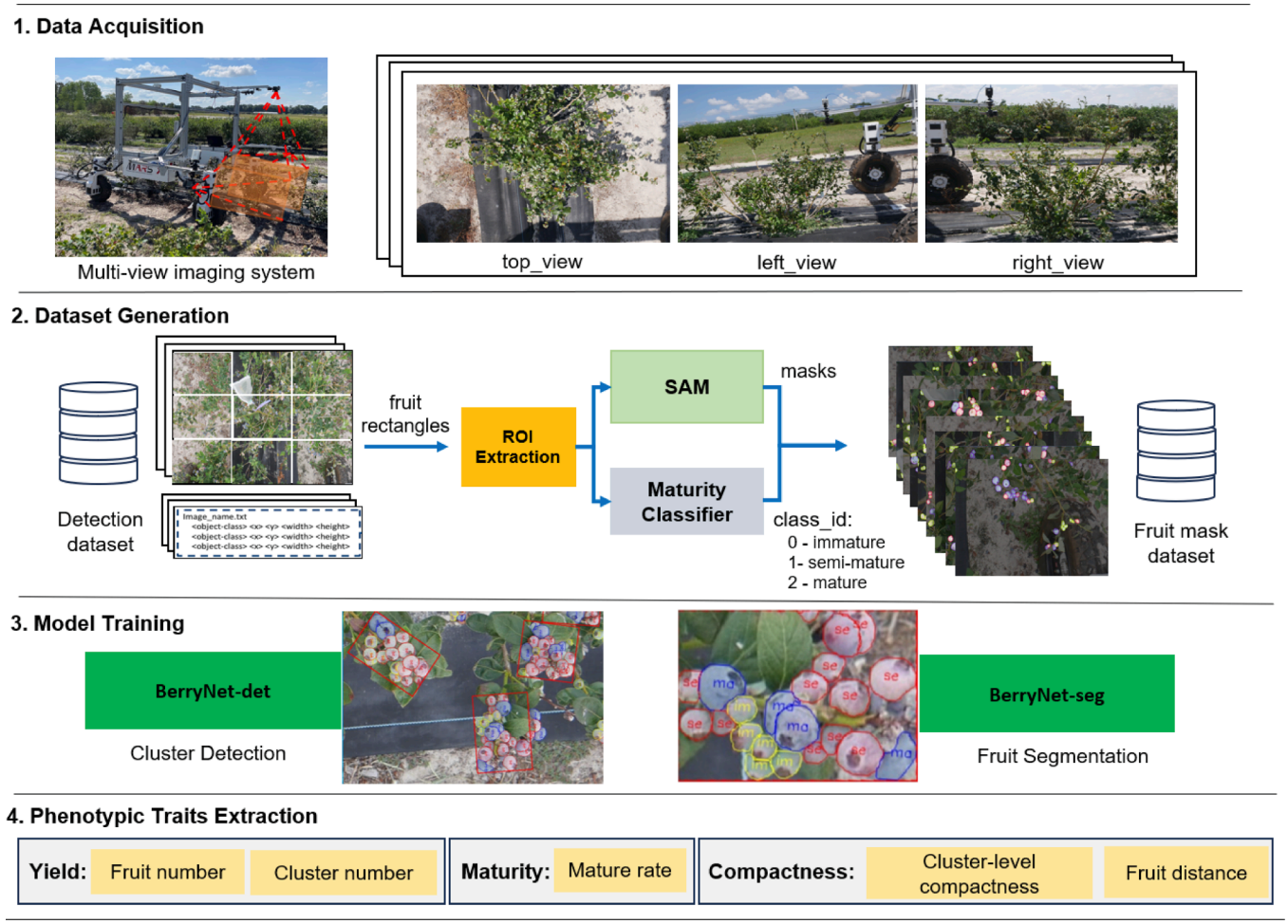


Fig. 1. Diagram of the proposed blueberry fruit phenotyping workflow which involves four stages: data collection, training dataset generation, model training, and phenotypic traits extraction.

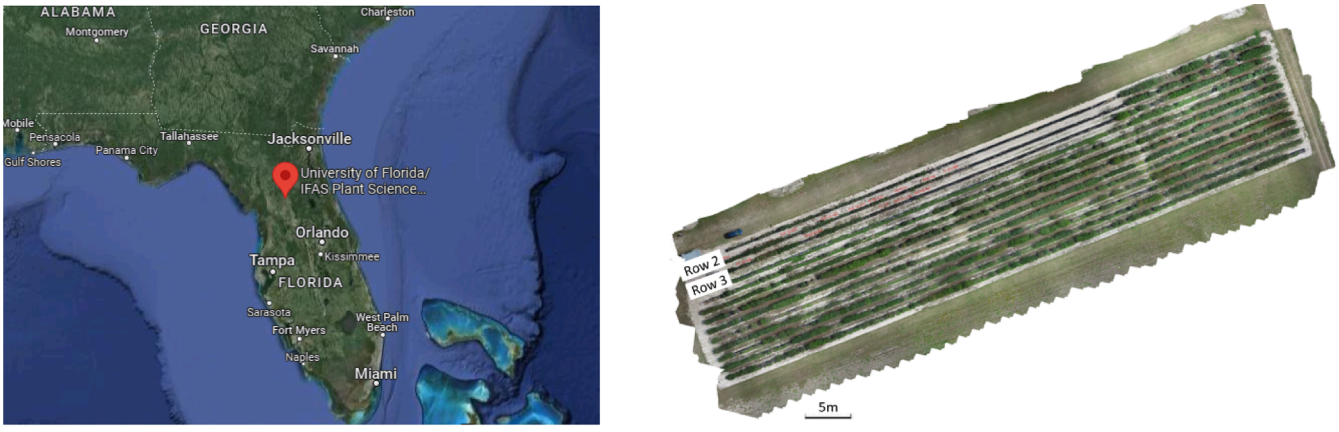


Fig. 2. Overview of the blueberry field located in Citra, Florida. (a) Geographic location on Google Maps; (b) Orthophotographs of rows 2 and 3 within the field, which were specifically targeted and scanned using our mobile platform.

and the indirect method, which uses predictive models based on related features through regression analysis (He et al., 2022; Niedbała et al., 2022). Direct methods for assessing blueberry plant traits often involve measuring attributes such as flower or bud count, fruit size, and fruit weight. Performed by humans, these methods are time-consuming, labor-intensive, and limited by sampling method quality, timing, weather, and available personnel. Indirect methods, employed across large geographic areas, utilize remote sensors such as satellites or drones to calibrate crop traits using vegetation indices, weather data, and soil information (Bai et al., 2021; Niedbała et al., 2022; Van Beek et al., 2015). Although these approaches effectively scale observations, they rely on complex models and diverse data sources, making analysis

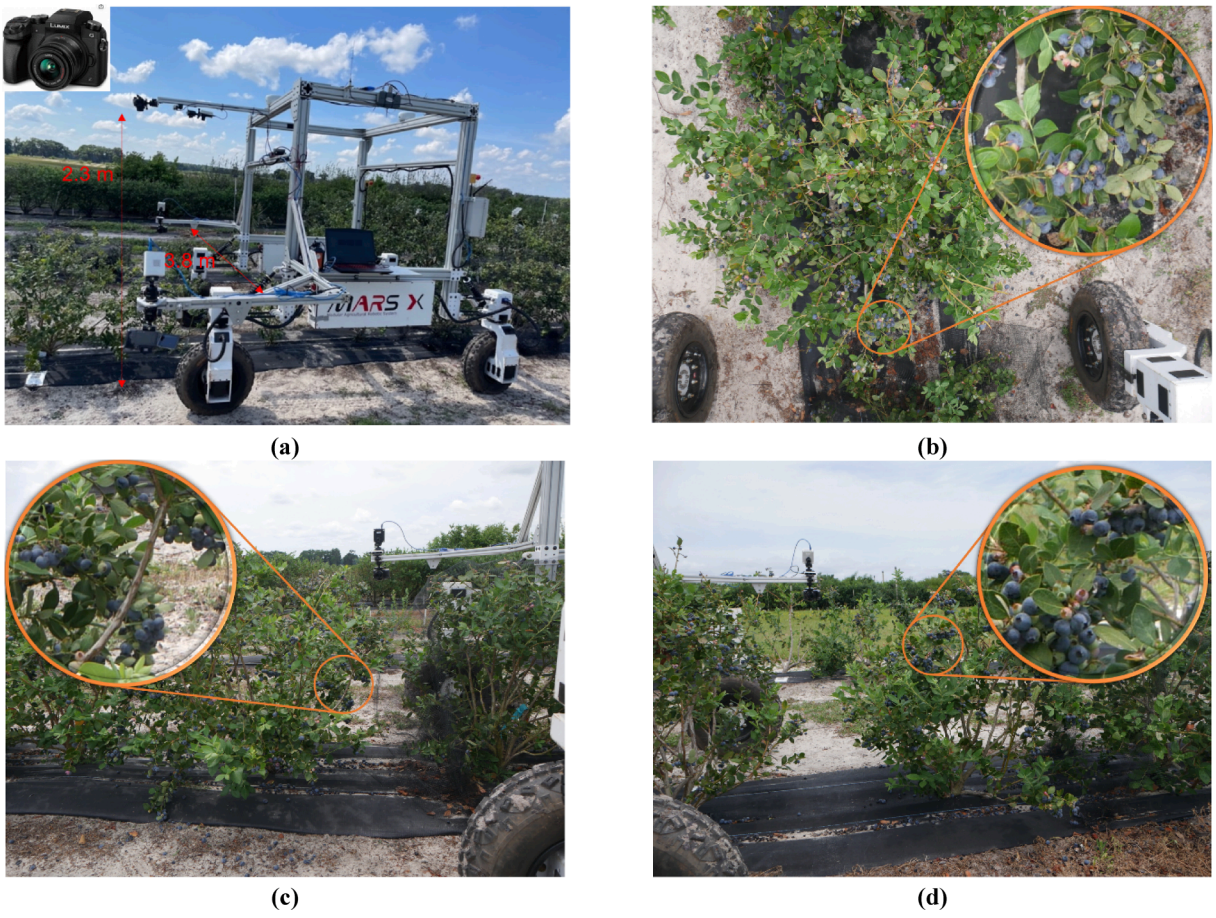


Fig. 3. In-field robotic phenotyping platform and data acquisition. (a) Ground-based phenotyping robot with a multi-view imaging system; (b)-(d) Examples of top, left, and right views (genotype of FLR 14–442) captured by the platform.

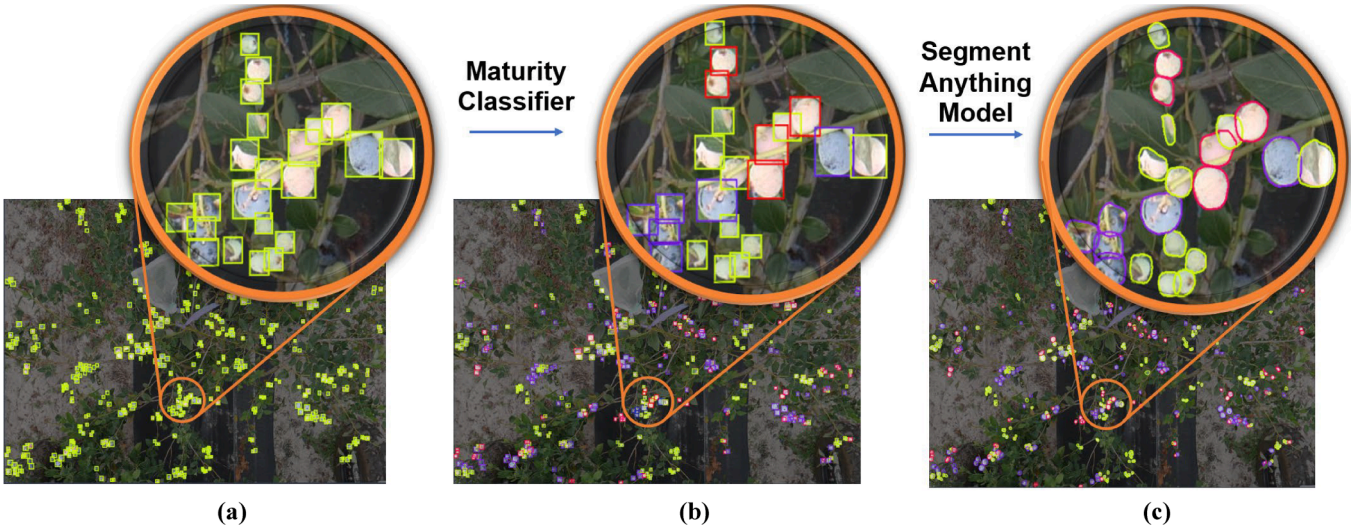


Fig. 4. Illustration of the proposed automated pixel-wise label generation process for berries at different maturity stages. (a) Bounding boxes from a previous detection dataset (Z. Li et al., 2023); (b) Bounding boxes re-classified into three categories: immature (yellow), semi-mature(red), and mature (blue), using a maturity classifier; (c) Pixel-wise mask labels generated using the Segment Anything Model.

challenging without advanced expertise. Additionally, they require calibration with historical data and are often limited by the spatial resolution of the data, which is not able to measure plant-scale or cluster-scale traits.

In blueberry breeding programs, breeders often focus on individual plants (given that some genotypes may only have a few plants) or single clusters, to precisely assess each genotype's traits. These programs mostly rely on direct manual sampling of a few berries or clusters to estimate the plant's situation. For example, yield is often estimated by counting the number of flowers or buds during the flowering stage, or by

Table 1
Comparative summary of blueberry detection dataset.

Dataset	Images	Labels	Source	Resolution	Annotation	
					Type	Class
HandSet	215	62,709	Digital cameras; phones.	720p ~ 4 k	bounding box	fruit
OpenSet	79	3,512	Roboflow	640 × 640	bounding box	fruit
RoboSet	128	42,877	Robotic imaging system	4 k	bounding box	fruit

measuring the size and weight of the fruit at the fruiting stage. Several metrics need to be measured periodically. For example, the fruit maturity rate is generally measured once or twice a week, further burdening the workload. In addition, mechanical harvesting demands genotypes with looser clusters to facilitate fruit detachment from the plant during harvest. However, there is no existing official standard for defining the level of compactness in fruit clusters, which currently relies on expert experience and is difficult to classify consistently. Some studies have adapted the concept of grape bunch compactness to assess blueberry clusters by calculating the ratio of the fruit region to the overall cluster region in the image (Gai et al., 2024; Ni et al., 2020). However, this approach is heavily influenced by the camera angle and distance due to the irregular shape of blueberry clusters, leading to inconsistent compactness values even for the same cluster. For those genotypes with small or sparsely clustered berries, directly measuring the distances between berries within the whole image could provide a solution to avoid the difficulties associated with defining clusters.

Various high throughput phenotyping platforms have been developed and shown promise in replacing humans for measuring phenotypic traits, including gantry (Vasconez et al., 2020), ground mobile platforms (Gutierrez et al., 2019; Williams et al., 2020), and aerial platforms (Chen et al., 2017; Zheng et al., 2022). These platforms are always equipped with an imaging system and adopt machine vision technologies to achieve automated plant phenotyping. In blueberry breeding, the early studies used RGB cameras to estimate the yield of wild lowbush blueberries with the top-view observation (Swain et al., 2010). A customized four-wheel Farm Motorized Vehicles (FMVs) equipped with digital cameras enabled real-time yield mapping of wild blueberries by calculating the blue pixel ratio in images. Subsequently, an automated yield monitoring system was further developed based on the mobile platform and imaging processing algorithm, creating fruit yield maps for commercial harvesters by combining the geo-referenced coordinates (Chang et al., 2012). Their results showed a strong correlation ($R^2 > 0.9$) between the percentage of blue pixels and actual fruit yield in low-bush wild blueberry fields using top-view observations. However, the blue pixel extraction algorithms relied on color thresholding, which led to incorrect identifications under varying lighting conditions, such as reflections and deep shadows. In high-bush blueberry cultivars, the spatial distribution of the berries is more complex due to the diverse plant architecture, making top-view imaging less effective in covering the fruit accurately. In addition, several recent studies applied drones to evaluate blueberry architecture (Patrick & Li, 2017), plant height (Haydar et al., 2024), and yield (Qu et al., 2024). While these methods achieved fast, high-throughput measurements, they primarily captured the overall plant shape and were not able to capture finer details that were critical to yield estimation.There is a need of developing novel methods to adequately address challenges such as fruit overlap, variable sizes, occlusions and lighting conditions, which are essential for precise and reliable measurements.

Machine learning has further advanced automated outdoor blueberry detection and phenotyping in the past decade. In the early stage, the color features were used for training a K-nearest neighbour (KNN)-based classifier to distinguish different maturity levels of blueberries captured by RGB images (Li et al., 2014), but the method cannot count the number of berries. A Histogram Oriented Gradients (HOG) based KNN classifier was developed to make the model robust in different illuminations, while it can only distinguish individual fruit cropped within regions of interest (RoIs) rather than detection in an image (Tan et al., 2018). After that, deep learning models such as the You Look Only Once (YOLO), achieved higher accuracy in outdoor blueberry fruit detection (Redmon et al., 2016; Terven & Cordova-Esparza, 2023). Its early versions, such as YOLOv3 and YOLOv4, have demonstrated real-time detection of lowbush wild blueberries at various maturity stages with over an 85 % accuracy (Haydar et al., 2023; MacEachern et al., 2023; Schumann et al., 2019). Other studies modified the different versions of YOLO model to enhance the blueberry fruit recognition within the fruit clusters of the commercial highbush plants (Liu et al., 2023; Yang et al., 2022). Furthermore, some studies have adopted instance segmentation to count and evaluate blueberry clusters for quantification (Aguilera et al., 2023; Gonzalez et al., 2019) and extract fruit traits associated with harvestability and yield, such as fruit number, maturity, and cluster compactness (Ni et al., 2020, 2021). Those studies achieved competitive segmentation performance, accurately delineating individual blueberries' boundaries. Compared to the detection model, instance segmentation models provide more detailed information about blueberries, such as boundaries, which made it possible to derive other useful traits such as cluster compactness. However, their high performance relied on the of the fruit clusters' imaging quality and the images were manually taken with specific angles and close distances to the fruit to avoid occlusion.

In the past studies, training and evaluation were performed with a small dataset and most images were taken manually with artificial backgrounds, limiting their usefulness in large breeding fields or commercial farms. For example, the image data are typically collected manually from ideal angles or close distances to minimize fruit occlusion and enhance feature visibility of plant parts of interest. These conditions fail to reflect the complexity of commercial or breeding fields, where diverse plant morphologies, varying light conditions, and different bush heights pose challenges that have not been addressed by existing methods. One potential solution is the use of multi-view imaging, which captures images from multiple angles to provide a more complete representation of the plants and reduce occlusion-related errors. This approach, combined with 3D reconstruction techniques, can help account for variations in plant structure and improve the accuracy of fruit detection (Yu et al., 2024).

Another major challenge in deep learning model development is that the training data requires extensive manual labeling that is time-consuming and prone to inconsistencies, particularly under the complex conditions typical of commercial blueberry fields. Recently, vision foundation models, such as the Segment Anything Model (SAM) (Kirillov et al., 2023), have demonstrated zero-shot capabilities, enabling them to recognize and classify images without prior training, which is particularly advantageous in rapidly adapting to diverse and complex agricultural environments (Chunhui Zhang et al., 2023). Recent studies have applied zero-shot models in agricultural applications, such as farmland regionalization (Gui et al., 2024; Tripathy et al., 2024; Chen Zhang et al., 2023), pest segmentation (Y. Li et al., 2023), crop/weed segmentation (Nguyen et al., 2023), and potato leaf segmentation (Williams et al., 2024). The vision foundation models could solve the semantic model training challenges without relying on extensive pixel-wise labeling, potentially enhancing the efficiency and effectiveness of phenotyping in blueberry cultivation.

To fill the technological gap in terms of high throughput data collection and more efficient data analysis with minimized annotation,

Table 2
Summary of maturity classifier dataset.

Dataset	Maturity level		
	Immature	Semi-mature	Mature
Training	2,458	1,384	1,429
Validation	684	372	443
Testing	337	197	210

we made two main contributions in this study. First, an in-field blueberry phenotyping robot equipped with multi-view cameras was developed to provide high throughput data collection and a comprehensive view coverage of blueberry plants to address fruit occlusion. Second, a customized deep learning model was developed to detect fruit clusters and segment individual fruit *in situ*. By leveraging the zero-shot ability of SAM, pixel-wise labels of berries were generated using the bounding box prompts provided by a berry detection dataset (Z. Li et al., 2023).

The overall goal of this study was to build a robotic phenotyping system for infield data collection and develop a deep learning-based data processing workflow to assess blueberry yield, fruit maturity, and cluster compactness of multiple genotypes. Specific objectives were to: 1) Develop a blueberry phenotyping system using a field robot equipped with multi-view cameras; 2) Leverage a vision foundation model (SAM) to train segmentation models to reduce the burden of mask annotation; 3) Customize BerryNet to enhance the cluster detection and blueberry segmentation; 4) Derive and evaluate blueberry fruit traits including yield, maturity, and cluster compactness across various genotypes.

2. Materials and methods

This study focused on developing a robotic multi-view vision system that was capable of realizing automated blueberry fruit phenotyping, assessing traits such as yield, maturity, and compactness. The overall procedure consisted of four stages: data collection, training dataset generation, model training, and phenotypic traits extraction (Fig. 1). We customized a mobile platform (Xu & Li, 2022) equipped with a multi-view imaging system (top, left, and right) to scan blueberry plants by navigating over crop rows. Using these field images, the datasets for cluster detection and fruit segmentation were generated with the aid of machine-assistance annotation. An automated pixel-wise labels generation method was developed for generating the masks of immature, semi-mature, and mature blueberries. Additionally, these mask labels served to identify preliminary RoIs for fruit clusters, which were then refined through further human modification. We developed BerryNet, an improved lightweight model based on YOLOv8, for detecting fruit clusters and segmenting different levels of maturity. Our pipeline successfully extracted phenotypic traits at both the plant and cluster scales, providing the metrics of yield, maturity, and compactness. Further details of each step were provided in the subsequent subsections.

2.1. Data acquisition and robotic imaging system

The experimental field was located at the Plant Science Research and Education Unit (PSREU) of the University of Florida in Citra, Florida, USA (Fig. 2). The field (29.408985° N, − 82.143210° W) grew blueberries with various genotypes for blueberry breeding. Image data were collected from the second and third rows, including 154 plants of 22 genotypes that were planted in 2021. The distance between rows was 3 m, and the distance between the blueberry crops was 0.5 m. Within each row, the height of blueberry plants varied from 0.4 to 2.0 m.

To navigate the blueberry field, the MARS-X mobile robotic platform, previously developed by (Xu & Li, 2022), was customized in its dimensions to traverse effectively over the plants (Fig. 3a). The mobile robot navigated across blueberry fields at a constant speed of 0.3 m/s with dual RTK-GNSS navigation, ensuring consistent imaging conditions. To cover the blueberry canopy and minimize occlusion as much as possible, the multi-view imaging system was customized to capture the blueberry image from the top, left, and right sides (Fig. 3b-d). To avoid the crushing and breaking of blueberry stems and shaking of the fruit clusters, the cameras were equipped at a relatively further distance from the plant. Specifically, the distance between the left and right cameras was 3.8 m with a height of 1 m from the ground; and the top camera was mounted with a height of 2.3 m. High-resolution images (3448 × 4592) were captured using Panasonic Lumix G7 (Panasonic, Osaka, Japanese) cameras with a 14–42 mm adjustable lens, ensuring comprehensive coverage of each plant. To maximize coverage, the camera's lens was set to its widest field of view (FOV), 63.42° × 49.73°, and it captured the image sequences with 1 FPS. The variation in plant morphology among different genotypes led to fluctuating overlap rates of 40 % to 70 % between adjacent images. These variations were due to the differing heights and widths of the blueberry plants. Data collection occurred weekly during the blueberry fruiting stage from April to May 2023, specifically on the dates of April 5, April 12, April 19, April 24, and May 2. From the sample images (Fig. 3b-d), the berries were extremely small (a typical blueberry fruit occupied only 1 %-4% of the image's width and height) and often occluded, making accurate detection challenging. The zoomed-in areas of these images revealed that fruit clusters contained berries at various maturity stages.

2.2. Training dataset generation

The dataset preparation generated two types of datasets: one for cluster detection using bounding boxes, and another for fruit segmentation with semantic masks. The initial in-field data collection included images with hundreds to thousands of small berries (4 ∼ 7 mm diameter). To accelerate the annotation process and reduce labeling errors, an automated pixel-wise annotation method was developed to generate masks for immature, semi-mature, and mature berries. This strategy involved transforming one-class bounding box annotations (section 2.2.1) into three-class pixel-wise mask labels (section 2.2.3) using a maturity classifier and a prompt-based vision foundation model SAM (Chunhui Zhang et al., 2023) (Fig. 4). Using these mask labels, a cluster extraction algorithm was utilized to generate the preliminary RoIs of fruit clusters, providing a basic reference for cluster labeling (section 2.2.4).

The maturity classifier was a compact Convolutional Neural Network (CNN) model that converted the single-class labels "fruit" into three-class labels (immature, semi-mature, and mature). In previous works, the shallow networks containing few convolution layers have been proven to validate fruit classification, especially the relatively easy maturity classification tasks (Naranjo-Torres et al., 2020). The classifier architecture involved two convolutional layers for feature extraction, two polling layers to reduce dimensionality, and two fully connected layers for predicting the maturity classes of the blueberries. The blueberry classifier was trained using a maturity classification dataset, described in section 2.2.2.

Subsequently, the SAM utilized the bounding boxes of berries as prompts to infer the images and generated the pixel-wise mask labels. SAM is a vision foundation model designed to handle diverse segmentation tasks dynamically. It integrated three main components: an image encoder, which extracts detailed visual features from images; a prompt encoder, which processes textual descriptions to specify the segmentation targets; and a mask decoder, which combines features from both encoders to generate precise segmentation masks. This architecture allows SAM to adapt flexibly to various user-defined segmentation tasks, making it highly versatile and effective across different applications. In our cases, the large pre-trained weight of SAM was selected to encode the images and prompts. After encoding with SAM's mask decoder, the outputting masks combine the maturity classification ID to generate three-class pixel-wise mask labels for immature, semi-mature, and mature berries.

Table 3
Summary of pixel-wise fruit segmentation dataset.

Dataset	Images	Labels			Source	Augmentation
		Immature	Semi-mature	Mature		
Training	2463	113,423	23,814	68,081	Automated pixel-wise label generation	Crop:0 ~ 20 % zoom Brightness: between ± 30 % Blur: up to 1 pixel
Validation	209	12,474	1,480	1,028		–
Testing	100	3,183	836	585	Manual labels	–

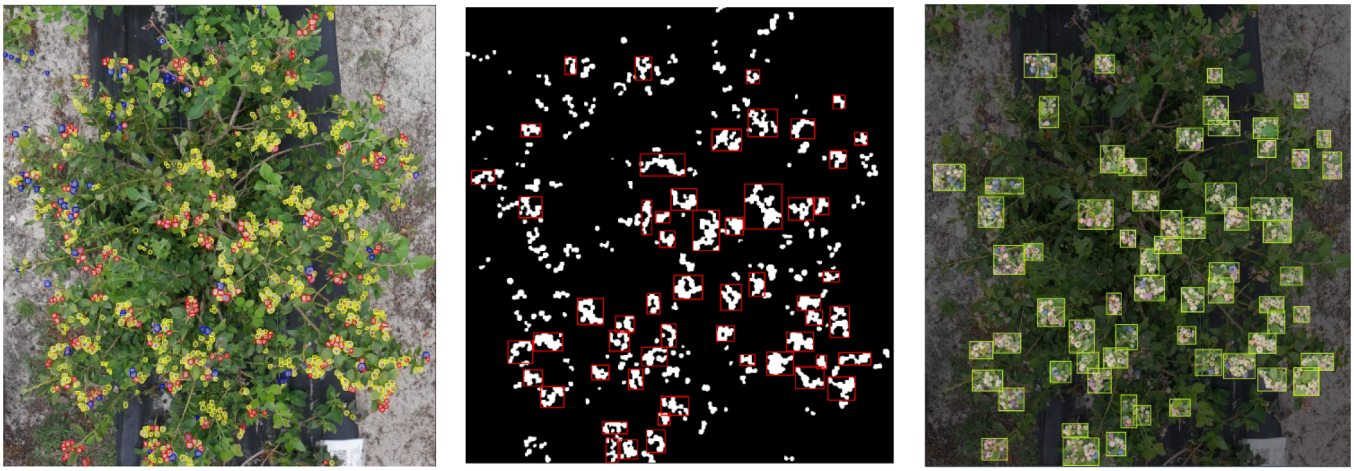


Fig. 5. Illustration of cluster detection dataset generation. (a) A visualized example from the Pixel-wise Fruit Segmentation Dataset (genotypes: FLR14-442 with yellow, red, and blue masks indicating the immature, semi-mature, and mature berries, respectively. (b) The result of the cluster extraction method is that red boxes are the extracted clusters based on the binarized berry regions; (c) Final cluster annotations after manual correction.

2.2.1. Blueberry fruit detection dataset

The blueberry fruit detection dataset was the start of the automated pixel-wise label generation process. It has two functions: one was to provide the RoIs for cropping the small patches of individual berries from the images and generate the unlabeled blueberry classification dataset; Another was to provide the bounding boxes prompts to the SAM to generate the pixel-wise labels. The dataset was from our prior work (Z. Li et al., 2023), and included three subsets: HandSet, OpenSet, and RoboSet. They featured RGB images from different locations and seasons, which annotated all visible blueberries as "fruit" with bounding boxes (Table 1). The dataset consists of a total of 405 images with over 100, 000 berries labeled.

2.2.2. Maturity classifier dataset

The dataset was prepared for training and evaluating the maturity classifier. With the existing bounding box annotations from the blueberry detection dataset, the RoIs were extracted as individual patches that contained only a single blueberry fruit. These patches then were resized to 64 × 64 × 3 and manually categorized into one of three maturity classes. Finally, the dataset was divided into training, validation, and testing sets with a ratio of 7:2:1 (Table 2).

2.2.3. Pixel-wise fruit segmentation dataset

The pixel-wise fruit dataset is used for training blueberry segmentation models. These labels are instance masks (immature, semi-mature, and mature) that were generated from the fruit detection dataset using a maturity classifier and SAM. Considering the small fruit pixel ratios occupied in the image, the images from RoboSet were divided into 3 × 3 smaller patches for the better inference result of pixel-wise masks. Finally, a total of 2672 patches with 125, 897 instances were generated using the automated pixel-wise labels generation method and they were divided into training and validation sets with a ratio of 9:1. The training set was augmented three times through the random transforms, including cropping (0 ~ 20 % zoom), brightness adjustment (between −30 ~ 30 %), and blurring (1 pixel).

Additionally, 100 images were manually annotated as the testing set, adjusting and correcting the mask classes and boundaries based on the initially generated labels. It can be used to evaluate the quality of labels generated by the pixel-wise label generation method and the performance of the fruit segmentation models. A summary of the pixel-wise fruit segmentation dataset is provided in Table 3.

2.2.4. Cluster detection dataset

The purpose of the cluster detection dataset is to annotate and detect clusters of berry fruit, which is important to evaluate cluster compactness and predict yield. This involved marking areas in images where blueberries are grouped as clusters using bounding boxes.

Visually identifying blueberry clusters in images is challenging due to the natural complexity and variability in their appearance (Fig. 5a). This complexity leads to inconsistent label annotations, as annotators might interpret and define clusters differently based on personal criteria. To improve the quality of the cluster annotations, the pixel-wise fruit segmentation dataset (section 2.2.3) was utilized to generate initial cluster bounding boxes, which were then refined with expert knowledge. The cluster extraction procedure is described as follows:

1) Convert all fruit mask annotations to a binary image.
2) Apply morphological operations to merge closely located masks.
3) Identify the contours of the connected areas.
4) Remove clusters containing fewer than six berries.
5) Determine the bounding boxes for clusters; then remove clusters with a shape ratio larger than five to avoid the mistakenly connected areas due to the multiple thin clusters connected
6) Correct incorrect or unsatisfactory labels by humans.

Table 4
Summary cluster detection dataset.

Dataset	Images	Cluster instances	Augmentation
Training	285	6231	Flipping and rotation Brightness: between ± 30 % Blur: up to 2.5px
Validation	60	1982	–
Testing	78	1148	–

An example of comparison between the cluster extraction method without (Fig. 5b) and with (Fig. 5c) human interaction reflects the specific limitations of the automated process, particularly in occlusion situations. Automated clustering sometimes fails by merging nearby clusters, being obstructed by foliage, or misinterpreting spaced berries within a cluster as separate, which hinders accurate blueberry cluster extraction.

Finally, the cluster detection dataset was compiled with a total of 423 images containing 9361 cluster instances. The training set was augmented three times through flipping, rotating, adjusting brightness, and blurring (Table 4).

2.3. Customized BerryNet

BerryNet is a customized deep-learning network, designed to enhance the blueberry cluster detection and fruit instance segmentation based on the state-of-the-art YOLOv8 (Jocher, 2024). The original YOLOv8 is designed to detect general objects and is limited in small object detection, especially for small berries. We implemented three key modifications to the YOLOv8 architecture (Fig. 6): 1) employing features from higher spatial resolution layers to enhance small feature extraction (Mudassar & Mukhopadhyay, 2019); 2) integrating Bi-directional Feature Pyramid Networks (BiFPN) for more effective feature fusion (Tan et al., 2020); and 3) introducing the FasterNet Block to improve spatial feature extraction and increase inference speed (Chen et al., 2023).

Higher spatial resolution layers in feature pyramid — P2:The initial modification to BerryNet's architecture involved leveraging higher spatial resolution layers in the backbone's feature pyramid to enhance detail extraction. In the original YOLOv8 model, the third, fourth, and fifth spatial features are typically used for feature fusion, designed to balance the accuracy and inference speed for most of the general objects and scenarios. In our adaption, the second spatial feature in the backbone was utilized for better capture of finer details, improving the detection and segmentation of small objects such as blueberries. The reason for avoiding using all of the spatial layers (P1-P5) is that incorporating the first spatial layer (P1) would significantly increase convolutional operations, leading to a substantial rise in computational demands (Floating Point Operations, FLOPs) and consequently lowering the inference speed.

More effective feature fusion — BiFPN: BiFPN, the Bi-directional Feature Pyramid network, has been proven an effective method to better understand the context around different objects, improving its ability to distinguish between closely situated or similar-looking objects (Tan et al., 2020). For blueberry detection/segmentation, BIFPN enhanced the distinction of individual berries, especially when they are clustered together or partially obscured by effectively fusing features at different scales. Unlike traditional top-down FPNs, BiFPN allows for an additional bottom-up pathway, creating a bidirectional flow of information (Fig. 7). This meant features from both higher and lower levels could be enhanced by information from all scales, leading to richer feature

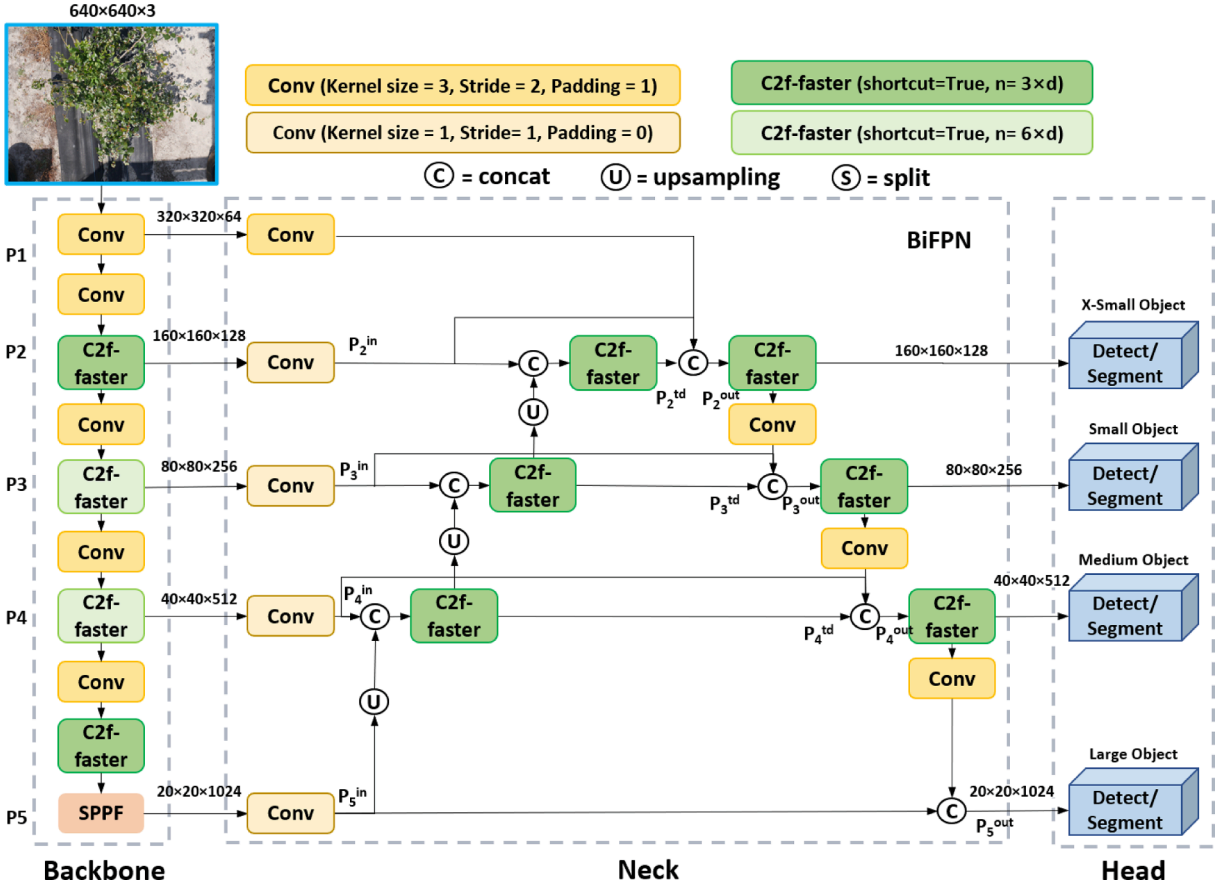


Fig. 6. Illustration of the BerryNet architecture. It incorporated three major enhancements: 1) enhancing P2 layer to better capture features of small objects; 2) implementing BiFPN for improved feature fusion, and 3) replacing C2f block with the more efficient C2f-faster block to accelerate inference.

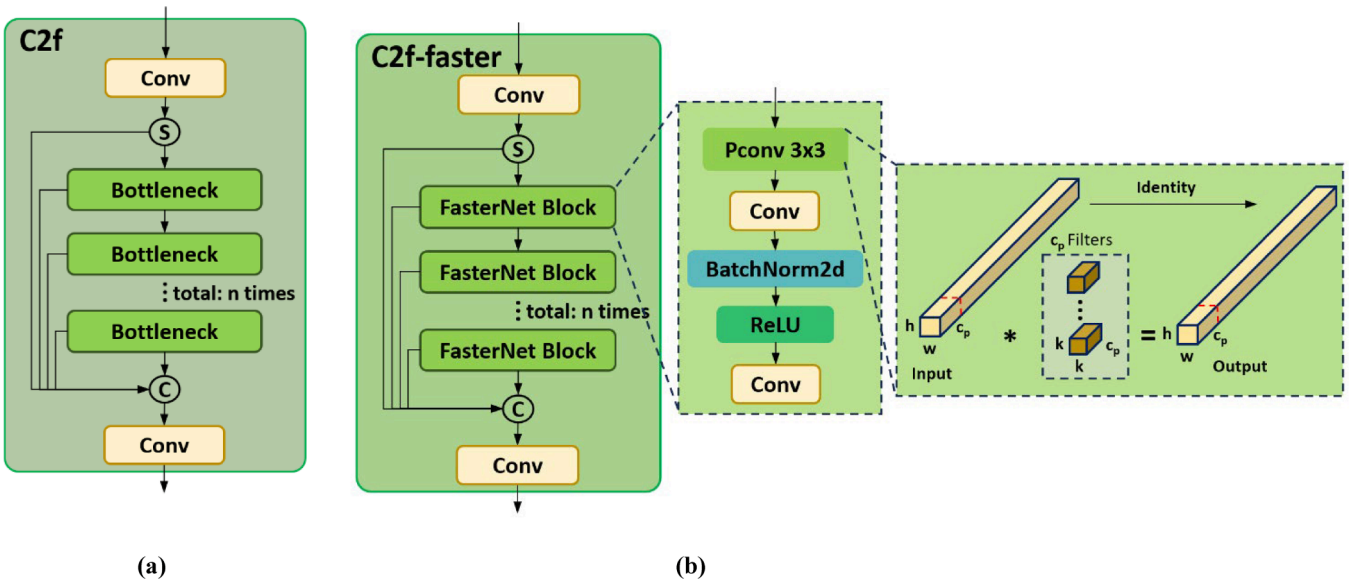


Fig. 7. Comparison between (a) the original C2f module and (b) the improved C2f-faster module. In the C2f-faster module, a FasterNet Block replaces the Bottleneck block. The Pconv in the FasterNet Block selectively processes only a subset of input channels, c_p, using regular convolutions for spatial feature extraction while leaving the remaining channels unchanged.

representations. This structure helped in refining the feature maps at all levels, leading to more accurate localization and detection/segmentation performance.

Given a multi-scale feature list $\vec{P}^{in} = (P_{l1}^{in}, P_{l2}^{in}, \ldots)$, which represented the features of the l_i layer. BiFPN effectively aggregated these diverse features to generate a new feature list:

$$\vec{P}^{out} f(\vec{P}^{in}) \tag{1}$$

As input features vary in resolution, they contribute differently to the output features. BiFPN adopted a nuanced approach, deviating from the conventional practice of treating all input features equally. It incorporated additional learnable weights w_i for each input, enabling the network to assign appropriate importance to each feature. In this study, we utilized the fast normalized fusion method as the weighted fusion technique (Pham et al., 2006).

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \tag{2}$$

To ensure each w_i is non-negative ($w_i \geq 0$), a ReLU activation function was applied after each w_i. Additionally, a small epsilon value of $\varepsilon = 0.0001$ was used to mitigate potential numerical instability. Ultimately, BiFPN integrated bidirectional cross-scale connections and utilizes fast normalized fusion for effective feature integration. The fusion of two features within BiFPN at layer i was described as Eq. (3) – (4).

$$P_i^{td} = Block\left(\frac{w_1 \cdot P_i^{in} + w_2 \cdot Resize(P_{i+1}^{in})}{w_1 + w_1 + \varepsilon}\right) \tag{3}$$

$$P_i^{out} = Block\left(\frac{w_1' \cdot P_i^{in} + w_2' \cdot P_i^{td} + w_3' \cdot Resize(P_{i-1}^{out})}{w_1' + w_2' + w_3' + \varepsilon}\right) \tag{4}$$

Here, P_i^{td} represents the intermediate feature of layer i on the top-down path, while P_i^{out} denoted the output feature on the bottom-up path. "Resize" was either an upsampling or downsampling operation used for resolution matching. Meanwhile, the "Block" typically refers to a convolutional operation used for feature processing. In this study, we employed the C2f-faster for this purpose.

Faster partial convolutional block — C2f-faster: To accelerate the neural network, we developed the C2f-Faster module, which effectively reduces the number of floating-point operations and parameters by integrating FasterNet block (Chen et al., 2023). Specifically, the C2f-faster module evolved from the original C2f module by reducing its Bottleneck component with the FasterNet Block (Fig. 7). A fast and efficient partial convolution (Pcov) was used in the FasterNet Block by applying filters on only a few input channels while leaving the remaining ones untouched. Pconv obtained lower FLOPs than the regular convolution and higher FLOPS than the depthwise/group convolution. The implementation enabled a decrease in computational complexity and computational resource requirement, benefiting higher inference speed, especially in resource-constrained environments, such as embedded systems and mobile devices.

As shown in Fig. 7b, the FasterNet Block comprised a Pconv followed by two 1 × 1 Conv layers, forming a reverse residual block, and incorporates shortcut connections to reuse important features. Normalization and activation layers were applied only after the middle 1 × 1 Conv layer, which played the role of retaining feature diversity and minimizing processing delay. In this setup, Pconv selectively utilized a subset of input channels (select c_p from c) for spatial feature extraction via regular $k \times k$ convolutional kernels (here, $k = 3$), leaving the remaining channels unaltered. This approach streamlines the feature extraction process and contributes to the overall efficiency of the neural network. Assuming that input and output feature maps had an equal number of channels, denoted as c, the FLOPs for Pconv amounted to $h \times w \times k^2 \times c_p^2$. When the partial ratio $r = c_p/c = 1/4$, the FLOPs for Pconv were only $1/16$ of those for ordinary convolution. Meanwhile, the memory access for Pconv, calculated as $h \times w \times k^2 \times c_p^2 \approx h \times w \times 2c_p^2$, was about $1/4$ of the original.

2.4. Phenotypic traits extraction

Phenotypic traits, including fruit yield, maturity, and compactness were formulated for quantitative analysis and comparison based on the BerryNet inference result of cluster detection and fruit segmentation. We conducted a manual harvesting process on 26 blueberry plants to collect the ground truth, each representing a different genotype. The process was split over two days, with 13 plants harvested on each day, specifically on April 12, and May 9, 2023. All berries, including immature, semi-mature, and mature, were harvested and counted to establish the manual ground truth. Subsequently, for each harvested plant, we selected the corresponding image frames of three views and manually

cropped these to focus solely on the individual plant. These individual plant images were annotated with bounding boxes to establish the image's ground truth.

As for the proposed phenotypic trait extraction algorithm, BerryNet was utilized to predict the cluster regions and individual fruit masks within single-view images. Subsequently, results from three distinct views were integrated to derive a comprehensive estimation of key phenotypic parameters, including the total fruit count, cluster count, the ratio of mature berries, as well as the spatial distribution of the berries, represented by a 2D mask map. This information enabled the extraction of detailed plant-level phenotypic traits.

For cluster-level phenotyping, given the initialized cluster region by BerryNet, the cluster extraction algorithm (described as step 1 to step 5 in section 2.2.4) was used to refine the bounding box with a rotation angle, that can decrease pixels of the background and other clusters within the bounding box. By combining fruit mask information within the clusters' RoIs, a detailed analysis of each cluster could be conducted further.

2.4.1. Yield estimation

Fruit Count: Yield was directly determined by the number of blueberries per plant. We referenced the count of predicted blueberries and used a multiple linear regression model to correlate these counts for individual plants with segmentation-based predictions from three different views. The relationship was formulated as:

$$Y_{multi} = a_0 + a_1 Y_{top} + a_2 Y_{left} + a_3 Y_{right} \tag{5}$$

Here, Y_{multi} represented the predicted blueberry count per plant, while $Y_{top}, Y_{left}, Y_{right}$ denoted the predicted counts from the respective views. The coefficients a_1, a_2, a_3 were the weights assigned to the top, left, and right views, with a_0 being the intercept.

Cluster Count: The number of fruit clusters per plant was also considered an important metric of yield. Using the cluster detection result from BerryNet, we extracted the cluster counts from three views. Considering the lack of the manual cluster count as the ground truth, the multi-view cluster count was formulated using the mean value of the single views (Equation (6). These cluster counts of the single/multi-views were separately used to regress the yield (manual fruit counts).

$$C_{multi} = Mean(C_{top}, C_{left}, C_{right}) \tag{6}$$

Where $C_{top}, C_{left}, C_{right}$ were the cluster count predictions from the BerryNet, while the C_{multi} was the estimated multi-view cluster count.

Visibility: Considering errors caused by occlusion, especially when comparing image-based counts to manual ground truth, visibility was used to evaluate the ratio of visible berries in an image to the actual berry count. The metric of visibility was measured by the ratio of annotated berries in each single-view image (observable berries) to the number of manually harvested berries (ground truth number):

$$Visibility = \frac{annotatedfruitnumberintheimage}{harvestedfruitnumberbyhuman} \tag{7}$$

2.4.2. Maturity level

The maturity level was assessed by calculating the ratio of the number of mature berries to total berries in an image. Using the predicted masks for immature, semi-mature, and mature berries. We averaged the maturity ratings derived from three views to determine the final maturity level.

$$M = \frac{\text{number } of \text{ mature fruit}}{\text{number } of \text{ total fruit}} \tag{8}$$

$$M_{multi} = Mean(M_{top}, M_{left}, M_{right}) \tag{9}$$

Where $M_{top}, M_{left}, M_{right}$ were the maturity rate predictions from the BerryNet, while the M_{multi} was the final maturity rate of the whole plant.

2.4.3. Compactness

Cluster-level Compactness: To calculate cluster-level compactness, we utilized the cluster RoI, identified by the rotational rectangles of a cluster (by cluster detector and cluster extraction algorithm) and fruit masks (by segmentation model) within the RoI (Fig. 8). Compactness for each cluster was calculated from the area of the fruit masks and the area of the minimum enclosing rectangle (determined by the masks) (Ni et al., 2020):

$$Compactness_{cluster} = \frac{Areaoffruits'mask}{\text{Area } of \text{ minimum } cluster's\text{rectangle}} \tag{10}$$

However, the compactness of blueberry clusters is significantly influenced by their irregular shape. Even for the same cluster, the compactness value can vary due to differences in imaging angles. In real in-field scanning, occlusion further complicates this variability. For a single plant, calculating the compactness of all clusters can provide insight into the overall distribution; however, the wide range of values makes it challenging to distinguish between different plants or genotypes. Since the harvesting process is most affected by clusters with tight fruit, the top 50 % of clusters were selected to minimize this variability and focus on the most relevant measurements. The plant's overall cluster compactness was then determined by averaging the calculated value of these clusters.

Fruit distance: Utilizing the 2D mask map produced by the segmentation model the normalized fruit distance was calculated by

Fig. 8. Illustration of cluster-level compactness calculation. From left to right: a rough region of fruit cluster predicted with BerryNet; minimum enclosing rectangle defined by the distribution of fruit masks within the cluster region; and calculation of cluster-level compactness.

examining the distribution of fruit's central points using the Voronoi-based nearest neighbors algorithm (Kolahdouzan & Shahabi, 2004). This algorithm partitioned the space around each fruit into regions, ensuring that every point within a given region was closer to its associated fruit center than to any other. The process was described as follows (Fig. 9):

1) Convert the prediction of blueberry masks Mask_i into a binary image.
2) Determine the central points of each fruit P_i
3) Identify the nearest neighbors P_j, for each central point P_i using the Voronoi-based search algorithm.
4) Calculate the distance L_i, as $\|P_i - P_j\|_2$, where $\|.\|_2$ denotes the Euclidean distance between the central points of two adjacent berries.
5) To adjust for variations in fruit size and prevent bias in compactness measure due to fruit scale, the fruit distance was normalized by calculating the ratio of distance L_i and the estimated diameter of the corresponding fruit's mask $2\|area(mask_i)/\pi\|_{\frac{1}{2}}$. Then compute the mean of these normalized distances that are only smaller than three to filter the non-clustering berries: $L_{Norm} = mean\left(\frac{L_i}{2\|area(mask_i)/\pi\|_{\frac{1}{2}}}\right)$.

A larger L_{Norm} value indicated a less compact plant with more widely spaced fruit, while a smaller value suggests a more compact arrangement with closely spaced fruit.

For each plant, three views were analyzed, and multiple clusters from these views were combined to evaluate the overall distribution of compactness. To statistically discern and group plants with similar cluster compactness profiles, an ANOVA test ($P = 0.05$) followed by a Tukey HSD test was employed, ensuring that observed differences in cluster compactness between plant groups are statistically significant. This provides a reliable basis for comparing the phenotypic traits of different blueberry genotypes.

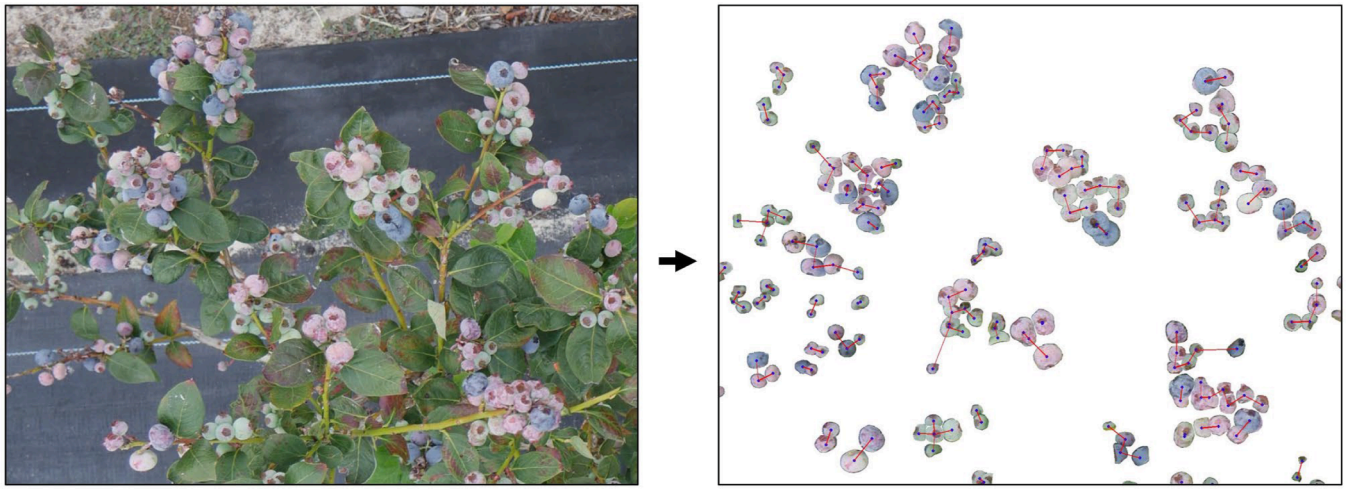


Fig. 9. Illustration of fruit distance calculation. The left image was a patch of raw image collected in the field, and the right image was the visualization of nearest neighbors for each berry. The blue dots were the central points, and the red lines connected their corresponding nearest neighbor.

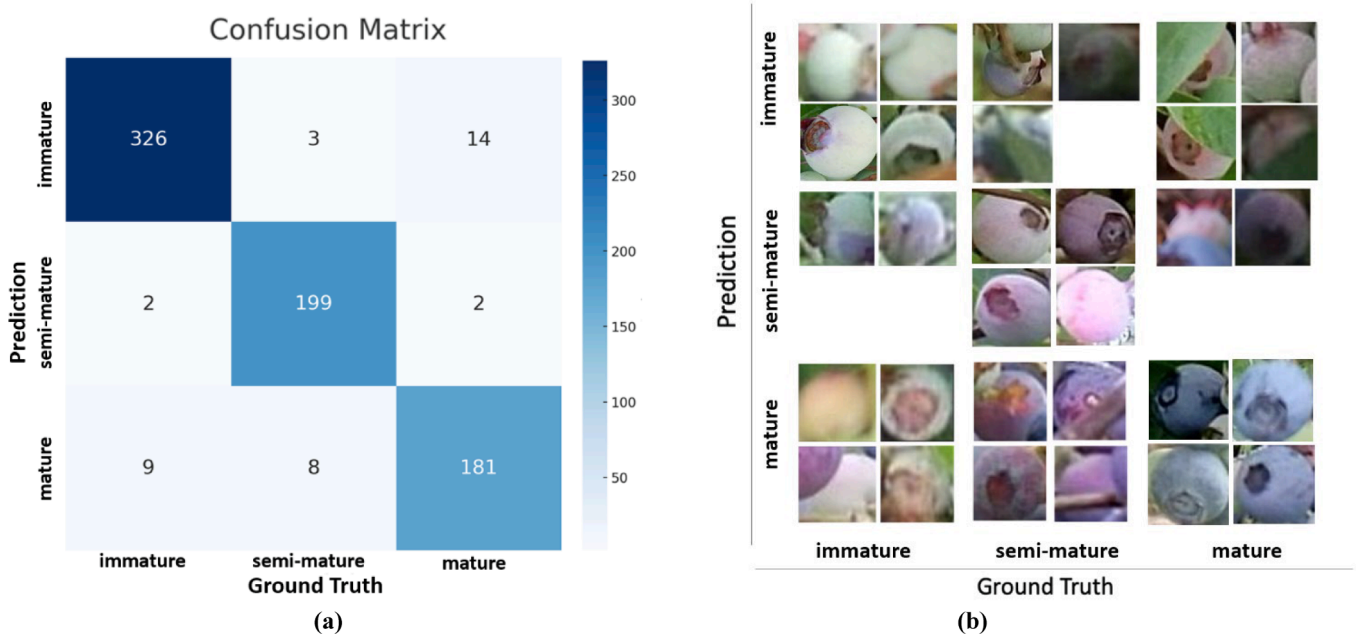


Fig. 10. Evaluation of the fruit maturity classifier. (a) Confusion Matrix; (b) Examples of prediction cases. The primary failures stem from occlusion caused by other fruit or leaves, poor image quality characterized by low resolution or low contrast, and inaccuracies in manual annotation.

2.5. Time-series analysis for genotypes comparison

With the proposed blueberry fruit phenotyping extraction methods, data collected during the fruiting period were utilized for genotype comparison. Our mobile robot scanned each blueberry plot at a rate of one frame per second, capturing images based on spatial geographic information. For each genotype, seven plants were grouped, yielding approximately 21 to 27 frames per scan from the three views for further analysis. These images were processed with BerryNet and extracted the phenotypic traits including fruit count, cluster count, maturity level, cluster-level compactness, and normalized fruit distance. Finally. the data collected on April 5, April 12, April 24, and May 2 were processed to generate the time-series boxplots of eleven blueberry genotypes in section 3.4 The ANOVA test ($P = 0.05$) and Tukey HSD test were conducted to statistically analyze the significance and group genotypes based on these metrics. These metrics provided insight into the developmental characteristics and potential yield of the genotypes across various dates, reflecting genetic expressions and environmental influences.

2.6. Evaluation metrics

To highlight the benefits of BerryNet for cluster detection and fruit segmentation tasks, a comparison was made with several classic detection and segmentation models. These models were trained and tested on the same datasets. The detection models included Faster-RCNN (Ren et al., 2015), RTMDet (Lyu et al., 2022), CenterNet (Zhou et al., 2019), RT-DETR (Zhao et al., 2024), YOLOv5 and YOLOv8 (Jocher et al., 2022). For segmentation tasks, the models used were Mask-RCNN (He et al., 2017), U-Net (Ronneberger et al., 2015), SOLOv2 (Wang et al., 2020), RTMDet-ins (Lyu et al., 2022), and Queryinst (Fang et al., 2021), YOLOv8-seg (Jocher et al., 2022). For the model comparisons, the lightweight version of these models was selected for faster training and inference. In addition, ablation studies were performed to analyze the functions of the three improved modules, including the P2 feature layer, BiFPN, and the c2f-faster model. By comparing the performance of cluster detection and fruit segmentation, the three modules were assessed in section 3.2.2. Besides, the different model configurations of BerryNet were also discussed in section 3.2.3.

The training and inference of cluster detectors and fruit

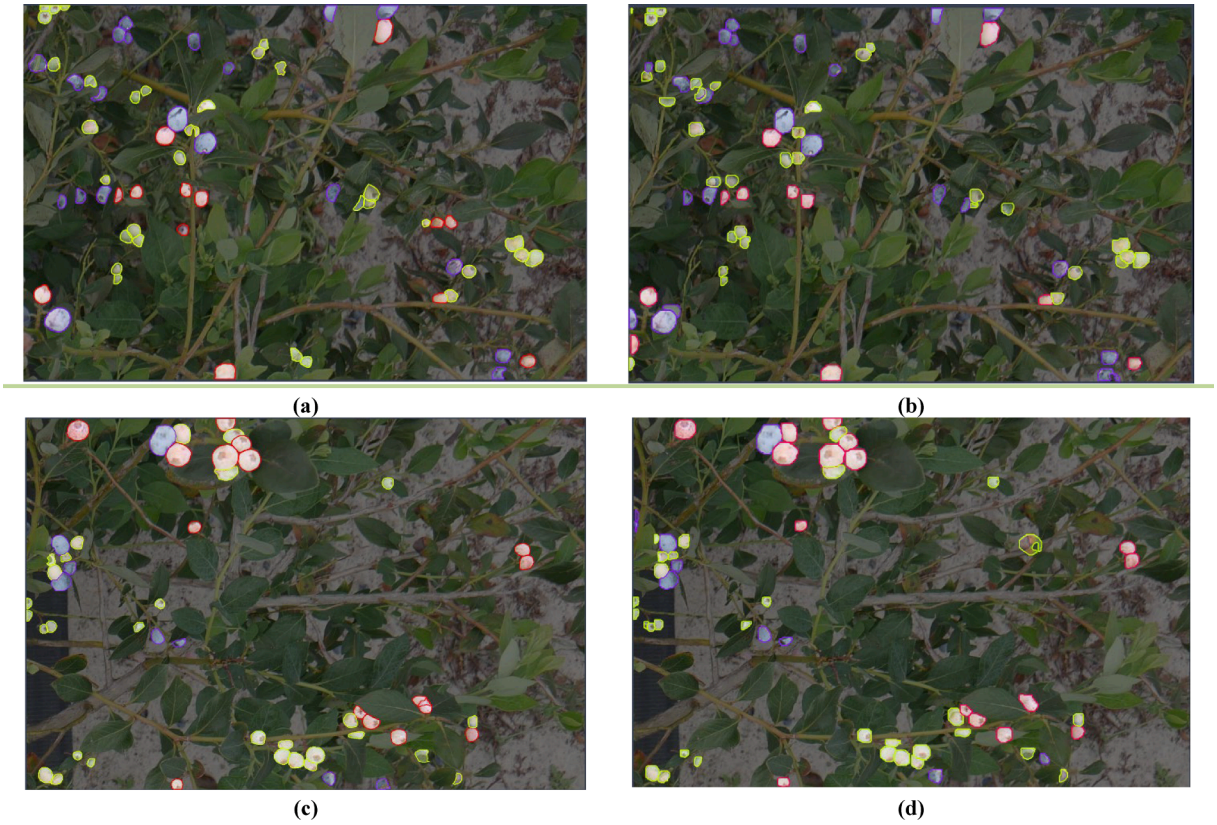


Fig. 11. Comparison of the masks from manual annotations and predictions with Segment Anything Model: (a) and (c) were blueberry masks annotated manually (ground truth); (b) and (d) were blueberry masks from automated pixel-wise labels generation method. The yellow, red, and blue masks represented the immature, semi-mature, and mature blueberries, respectively.

Table 5
Performance comparison of recent lightweight object detectors and BerryNet.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)	Params (M)	GFLOPs
Faster-RCNN- r50	44.1	43.2	40.3	25.2	41.34	90.8
RTMDet-tiny	38.8	29.0	38.8	17.0	4.8	8.1
CenterNet-r18	47.4	41.3	47.4	16.9	14.2	16.2
RT-DETR	55.3	50.0	48.3	24.4	19.9	56.9
YOLOv5n	50.4	48.9	43.9	20.6	1.9	4.5
YOLOv8n	51.2	49.7	45.7	22.2	3.01	8.1
BerryNet (Ours)	**55.8**	**54.2**	**52.8**	**27.3**	**1.77**	15.7

Table 6
Performance comparison of recent lightweight segmentation models and BerryNet.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)	mIoU (%)	Params (M)	GFLOPs
Mask-RCNN-r50	68.7	58.3	68.4	52.5	50.4	44.4	134.4
Unet-r50	55.4	53.2	51.1	49.2	29.7	31.06	16.59
SOLOv2-r18	44.4	31.6	44.0	31.6	23.6	18.1	42.5
RTMDet-ins-tiny	65.6	54.8	65.6	46.9	48.6	5.6	11.8
Queryinst-r50	39.9	41.7	39.8	28.9	21.4	46.7	156.7
YOLOv8n-seg	75.2	72.6	77.7	43.9	49.7	3.2	12
BerryNet (Ours)	**75.4**	**71.3**	**78.7**	**56.0**	**54.3**	**2.0**	34.1

segmentation models were implemented on a High-Performance Computing platform HiperGator at the University of Florida with a NVIDIA DGX A100 GPU (80G). The operating system was RedHat with software tools including CUDA 12.2, cuDNN 8.9, and Python 3.10 with OpenCV 4.7.0 library. YOLO and RT-DETR were trained and evaluated based on the official implementation of Ultralytics repository (Jocher et al., 2022) with default parameters, while other benchmarks were implemented using the MMDetection framework (Chen et al., 2019) developed by OpenMMLab.

Mean intersection over Union (mIoU), precision, recall, mAP50, and mAP95 were used to evaluate detection and segmentation accuracy. The following parameters were used in the formulae for some of the above evaluation metrics: TP (predicted as a positive sample and as a positive sample as well), FP (predicted as a positive sample, though it was a negative sample), and FN (predicted as a negative sample, though it was a positive sample). Intersection over Union (IoU) represented the ratio of the intersection area to the union area between the predicted segment and the true segment. The mean Intersection over Union (mIoU) was the average of the IoU values across all classes,

$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoU_i \tag{11}$$

where IoU_i denoted the IoU value for the i-th class, and N denoted the number of classes in the dataset.

Precision was the ratio of the number of positive samples predicted by the model to the number of all detected samples,

$$Precision = \frac{\mathrm{TP}}{TP + FP} \tag{12}$$

Recall was the ratio of the number of positive samples correctly predicted by the model to the number of positive samples that appeared,

$$Recall = \frac{\mathrm{TP}}{TP + FN} \tag{13}$$

The average precision (AP) was equal to the area under the precision–recall curve,

$$AP = \int_0^1 Precision(Recall)d(recall) \tag{14}$$

Mean average precision (mAP) was the result obtained by the weighted average of AP values of all sample categories, which is used to measure the detection performance of the model in all categories,

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{15}$$

Where, AP_i denoted the AP value of category index value i, and N denoted the number of categories of the samples in the training dataset (in this paper, N was 3). mAP50 denotes the average accuracy when the IoU of the detection model is set to 0.5, and mAP95 denotes the average accuracy when the IoU of the detection model is set from 0.5 to 0.95 (with values taken at intervals of 0.5).

Furthermore, parameter numbers (Params) of model weights and FLOPs Floating Point Operations per Second (FLOPs) were also used for evaluating a model's complexity and inference performance. Params metric refers to the total number of learnable parameters of the weight in the model, representing its capacity to capture and learn features from the data. While the FLOPs metric measures the computational workload required for a single forward pass through the model, reflecting its computational complexity and inference speed, higher FLOPs indicate more processing power and potentially slower inference. Balancing Params and FLOPs is essential for optimizing both the efficiency and accuracy of deep learning models.

3. Results

3.1. Automated pixel-wise label generation

Compared to manual labeling without any algorithm assistance, the data annotation process combining our automated pixel-wise label generation method was generally 10–20 times faster for generating labels, especially for the semantic labels. The success of the method depended on the quality of the original blueberry detection dataset and the accuracy of the maturity classifier and SAM. The evaluation of the maturity classifier and SAM are presented as follows.

Maturity classifier: The maturity classifier was trained using the dataset from section 2.2.2, achieving an average accuracy of 94.89 % on the testing set (Fig. 10). The high classification accuracy was efficient in distinguishing the immature, semi-mature, and mature fruit. Among those misclassified cases, one major source of error was the occlusion caused by leaves and other berries, making fruit partially visible. The green leaves tend to make the classifier mistakenly recognize the semi-mature/mature fruit as immature due to the similar appearance of leaves and immature fruit. Another source of error was from the image quality, especially the low-resolution and low-contrast situations. For example, the red semi-mature berries were sometimes classified into mature fruit due to the similar color under the weak-lighting condition. One more source of error was from incorrect annotations by human annotators. The annotation error accounted for about 20 % of the failed cases. With the classifier, these cases with incorrect labels could be classified into the correct class (Fig. 10 b). The maturity classification could be further enhanced if these sources of errors can be mitigated.

Segment Anything Model: By comparing the labels annotated by humans, the SAM method demonstrated the satisfied pixel-wise mask generation ability to segment berries in the images. The result showed mean Intersection over Union (mIoU) scores of 91.64 %, 90.67 %, and 93.41 % for immature, semi-mature and mature blueberry masks, respectively. As shown in Fig. 11, the automated pixel-wise mask generation pipeline effectively handled maturity classification and mask boundary generation in most situations. Most failures occurred when over 50 % of the blueberries were occluded, leading to inaccuracies in mask boundary generation. Occlusion was a common and inevitable problem especially for these plants with dense leaves. Accurately identifying and segmenting partially obscured objects is a common challenge in the agricultural field. Overall, the accuracy of SAM was efficient enough for mask generation.

3.2. BerryNet evaluation

3.2.1. Performance comparison with other models

Cluster detector: The BerryNet was compared with the state-of-the-

Table 7
Ablation studies of BerryNet for cluster detection.

P2	C2f-faster	BiFPN	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)	Params (M)	GFLOPs
×	×	×	51.2	49.7	45.7	22.2	3.01	8.1
✓	×	×	53.7	48.1	47.5	23.5	2.92	12.2
×	✓	×	**57.0**	49.9	51.2	25.0	2.3	**6.3**
×	×	✓	50.0	50.8	46.9	22.4	1.99	7.1
✓	✓	✓	55.8	**54.2**	**52.8**	**27.3**	**1.77**	15.7

art object detection models to detect the fruit cluster on the cluster detection dataset and the results showed that the BerryNet achieved the highest mAP using the least parameters (Table. 5). Overall, BerryNet outperformed other AI models with the increases of 4.5 % to 14.0 % in mAP50. Compared to the YOLOv8 baseline, BerryNet achieved a 7.1 % higher mAP50 while using 40 % fewer parameters. Although the GFLOPs required by BerryNet were almost twice that of the YOLOv8 baseline, BerryNet was still competitive among lightweight models.

Nevertheless, even the top-performing BerryNet model only achieved a mAP50 of 52.8 %, which cannot be considered highly accurate. Typically, a mAP50 greater than 80 % is expected for robust performance. The inconsistent performance can be attributed to the inherent variability in the shapes of blueberry clusters. Unlike standardized vertical rectangles, blueberry clusters come in various shapes, sizes, and orientations, often blending with background elements or adjacent clusters. This complexity posed significant challenges for detection algorithms which struggle to delineate and accurately identify individual clusters amidst such variability.

Fruit instance segmentation: In terms of segmentation performance, BerryNet demonstrated the highest instance segmentation accuracy for berries among the latest lightweight instance segmentation networks (Table 6) on the pixel-wise fruit segmentation dataset (section 2.2.3). The results indicated that BerryNet achieved an average improvement of 13.8 % in mAP50 over the off-the-shelf models. Compared to the baseline YOLOv8-seg, BerryNet outperformed in segmenting more accurate and detailed masks, with a 12.1 % higher mAP95 and a 4.6 % increase in mIoU. Mask-RCNN and the RTMDet model showed significantly lower accuracy, underperforming by more than 10 % of mAP50. The remaining models exhibited even poorer results, with mAP50 scores below 50 %, showing the challenge in blueberry segmentation. Additionally, BerryNet has the advantages of the smallest model size and competitive GFLOPs for faster inference.

3.2.2. Ablation studies

Ablation studies on cluster detection (Table 7) and fruit segmentation (Table 8) revealed that each module contributed specific advantages: P2 enhanced accuracy in detecting small blueberries, C2f-faster improved precision while reducing computational load, and BiFPN increased recall and delineation of boundaries.

The P2 module led to a noticeable 3 % increase in both precision and mIoU, underscoring its importance in accurately predicting the presence and contours of objects. Conversely, the C2f-faster module had a pronounced effect on model precision, particularly in cluster detection. The model with C2f-faster demonstrated higher precision compared to those without this module. However, its impact on recall was less pronounced, suggesting that C2f-faster enhanced the model selectivity. It did not necessarily improve the identification of all present objects. The BiFPN module, designed to enhance feature integration across scales, significantly strengthened recall and mAP95, particularly in the complex task of fruit segmentation. The improvement in mAP95 reflected average precision at a stricter IoU threshold, demonstrating that BiFPN excels in capturing detailed object boundaries. An increase in mIoU when incorporating BiFPN further underscored its role in precise boundary delineation, which was crucial for fine-grained segmentation tasks.

3.2.3. Model configuration

Considering different application scenarios, particularly online versus offline inference, various configurations of input image size and model size were compared to achieve a better tradeoff between accuracy and computational complexity tailored to specific computational requirements. Overall, the comparison revealed that increasing both the model size and input image resolution generally enhances BerryNet's accuracy both in the task of cluster detection and fruit segmentation. However, this improvement came with a substantial increase in computational demands (Table 9).

Specifically, for cluster detection, the large-version BerryNet with a 1280 × 1280 image input achieved the highest mAP50 of 54.9 %, reflecting an 8.6 % improvement over the smallest configuration. Increasing the image input size contributed more to accuracy than merely enlarging the model while requiring only one-fifth to one-fourth of the FLOPs needed for a larger model. In contrast, for fruit segmentation, increasing the image input size had a lesser effect on mask segmentation accuracy. Both the Nano and Large configurations with larger image inputs performed similarly to those with smaller image inputs. However, increasing the model size had a more significant impact, particularly with the 640 × 640 input size, resulting in a 6.8 % improvement in mAP50 and an 8.6 % boost in mAP95. Fruit segmentation focuses more on accurately capturing detailed boundaries, which required deeper layers to learn complex features and patterns, compared to cluster detection. The large BerryNet model, with a 1280 × 1280 image input, achieved the highest mAP50 of 85.8 % in fruit segmentation. However, compared to the nano version, it came at the cost of a significant increase in computational demand, with the number of parameters increasing nearly 20-fold and GFLOPs rising 10-fold.

Qualitative results demonstrated that BerryNet has robust fruit segmentation capabilities (Fig. 12). In real in-field images, the model successfully detected most blueberries, including immature fruit that closely resemble green leaves and berries partially obscured by other fruit or leaves. To enhance focus on the clusters and minimize background interference, rotational minimum bounding rectangles were applied, effectively isolating the clusters (Fig. 13). Automated clustering, however, was occasionally prone to inaccuracies, such as merging adjacent clusters, being obscured by leaves, or mistakenly categorizing loosely spaced berries within a cluster as separate clusters.

3.3. Phenotypic traits extraction

3.3.1. Yield estimation

Fruit count: Overall, both the single-view and multi-view approaches were efficient methods to predict the yield based on the predicted fruit count from images. According to the linear regression analysis (Fig. 14a), the multi-view fusion approach achieved the highest R^2 value of 0.73 and the lowest MAPE of 23.1 %. The top view alone achieved an R^2 value of 0.71, slightly less effective than the multi-view approach. As for the multi-view regression function (Equation (5), the results yielded coefficients $a_0 = 78.49, a_1 = 1.00, a_2 = 0.99, a_3 = -0.14$. Considering the high correlation between the different views, these coefficients estimated by the least square method may not reflect the contribution of each view.

Cluster count: Compared to directly predicting yield using the predicted fruit count, the method based on predicted cluster count slightly

Table 8
Ablation studies of BerryNet for fruit segmentation.

P2	C2f-faster	BiFPN	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)	mIoU	Params (M)	GFLOPs
×	×	×	75.2	72.6	77.7	43.9	49.7	3.2	12
✓	×	×	74.3	72.9	78.6	55.8	52.7	3.2	26
×	✓	×	75.4	71.0	77.8	54.0	49.2	2.5	10.2
×	×	✓	74.4	73.0	78.8	53.6	51.2	2.2	10.9
✓	✓	✓	75.4	71.3	78.7	56.0	54.3	2.0	34.1

decreased the correlation. (Fig. 14b). For each view, the R^2 value between the yield and cluster count was lower by approximately 0 ~ 0.06 than that of fruit count. The yield regression using multi-view cluster counts achieved the highest R^2 value of 0.69 and all of the single-view methods were also higher than 0.65. While the cluster count method slightly underperformed compared to fruit count in terms of correlation with yield, it still demonstrated strong predictive power. This suggests that although fruit count may provide finer granularity, cluster count remains a viable and efficient alternative for estimating yield, particularly in scenarios where individual fruit detection is challenging due to occlusions or image quality issues.

Visibility: By calculating the visibility of the three views across 26 plants, it was found that most views captured only 35 % to 55 % of the berries (Fig. 14c). The top view provided the highest average visibility ratio of 53.87 %, while the left and right views were 47.3 % and 48.1 %, respectively. The primary challenge is that a single perspective can only capture a portion of the blueberry plant's surface. Dense leaves and branches often block the camera's view, leaving many berries hidden from sight. Additionally, visibility varies across different plants and genotypes due to differences in plant structure, leaf density, and berry distribution, which further complicates yield estimation based solely on visible fruit. This variability increases the uncertainty in yield predictions based solely on visible berries, as occlusions and obscured clusters significantly affect the regression models, reducing the accuracy of yield estimates. This issue particularly existed in plants with either significantly dense or sparse leaves, where the visibility of berries can vary widely.

3.3.2. *Maturity*

The maturity rate predicted by BerryNet demonstrated a strong correlation with manual assessments of harvested fruit, with both the single-view and multi-view approaches achieving R^2 value of above 0.8 and MAE of less than 8 % (Fig. 15). Even in the situation of partial coverage of the plants, the local maturity rates were efficient in reflecting the maturity at the whole plant level. The multi-view method (mean the maturity rate of single views) slightly outperformed the single-view approach in terms of R^2 and showed a smaller RMSE. It suggested that the multi-view method was more reliable for maturity estimation. This level of precision should be sufficient for breeders to reliably assess the maturity of different genotypes.

3.3.3. *Compactness*

Cluster-level compactness: Among the 26 plants analyzed, the compactness scores ranged from about 0.5 to 0.7 and presented a similar distribution even though they were from different genotypes (Fig. 16a). Statistical test suggested that the plants of the genotypes Springhigh and Keecrisp exhibited relatively higher cluster-level compactness, indicating that the berries within these clusters are more tightly packed together. While those with the genotypes FLR16-669, FLR14-442, and Wayne showed lower compactness, indicating looser clusters that are easier to harvest. However, most of the plants presented a similar distribution without a significant difference from each other. The reason was that cluster-level compactness solely reflected the compactness level of detected clusters in the image, which was significantly influenced by the shape of the fruit cluster as it appears in the image (Fig. 16b). Additionally, cluster compactness was highly dependent on cluster detection and the specific viewpoint from which the image was captured. Although only the top 50 % of clusters with the highest compactness were selected for analysis to minimize deviations in cluster-level compactness, the distribution for individual plants remained broad. The variation in cluster characteristics within the same plants made it challenging to distinguish between different genotypes, with the range of variation being broader than the average differences observed across genotypes (from 0.53 to 0.71). This variability was particularly prone to being skewed by a few clusters, especially when a plant had only a small number of fruit clusters.

Fruit distance: Compared to cluster-level compactness, the metric of fruit distance demonstrated a higher degree of differentiation to the compactness according to the group number of significant differences (Fig. 17a). Unlike cluster-level compactness, fruit distance was calculated based on the nearest neighbors between fruit across the entire image, providing a more comprehensive assessment of the plant's overall compactness rather than focusing on individual clusters. This approach bypasses the need for cluster detection, reducing the impact of imaging viewpoints and occlusions. As a result, it minimized the variability caused by the structure of single clusters and offered a more robust representation of the spatial distribution of fruit across the plant.

Among the 26 plants studied, those measured earlier in the season (marked with an asterisk) exhibited shorter fruit distances compared to those measured later. This trend reflects a real-world scenario where mature berries drop or are harvested, resulting in increased fruit distance and looser clusters. The plants with genotypes FLR13-218, Wayne, and Endura present the highest fruit distances, indicating the loosest fruit cluster; while FLR14-442*, FLR13-218*, and FL11-137* had the tightest clusters. Visualization of the fruit distance for four plants effectively demonstrated its correlation with cluster compactness (Fig. 17b). It is worth noting that the two plants with the same genotype, FLR13-218, exhibited significant differences in fruit distance, likely due to variations in growth and the timing of harvest. These factors contributed to the noticeable disparity in cluster compactness between the plants. Overall, fruit distance serves as a useful metric that aligns with human visual assessment of cluster tightness.

There is a certain consistency in estimating compactness metrics between cluster-level compactness and fruit distance, but this consistency does not always exist. For example, the plants of Springhigh*, FLR13-218*, and FLR14-372 displayed high cluster-level compactness values and correspondingly low fruit distance values, both indicating tighter clusters. In contrast, Wayne, Endura, and FLR16-669* exhibited low cluster-level compactness and high fruit distance values, reflecting looser fruit clusters. However, there are some inconsistent cases, such as Keecrisp, where both metrics were relatively low. As shown in Fig. 17, the fruit distribution for Keecrisp was relatively even, but the detected clusters reflected high local compactness, suggesting a potential discrepancy between the two metrics. Cluster-level compactness can easily be influenced by a few detected clusters, which may lead to misleading judgments about the overall compactness. Therefore, combining both cluster-level compactness and fruit distance metrics provides a more comprehensive assessment and reduces the risk of inaccurate conclusions.

Table 9
Comparative performance of different BerryNet configurations for cluster detection and fruit segmentation. Image sizes 640 and 1280 represent: 640 × 640 and 1280 × 1280, respectively.

BerryNet	Model size	Size (pixels)	Precision	Recall	mAP50	mAP95	Params (M)	GFLOPs
Cluster detection	Nano	640	52.3	46.8	46.3	28.1	1.7	15.7
		1280	55.8	54.2	52.8	27.3		62.9
	Large	640	52.1	47.7	44.2	25.3	33.4	246.5
		1280	**58.5**	**51.6**	**54.9**	**32.3**		1363.8
Fruit segmentation	Nano	640	75.4	71.3	78.7	56.0	2.0	34.1
		1280	75.7	75.2	82.1	60.7		136.3
	Large	640	**80.2**	77.2	85.5	**64.6**	37.1	340.9
		1280	78.6	**79.0**	**85.8**	61.7		2853.8

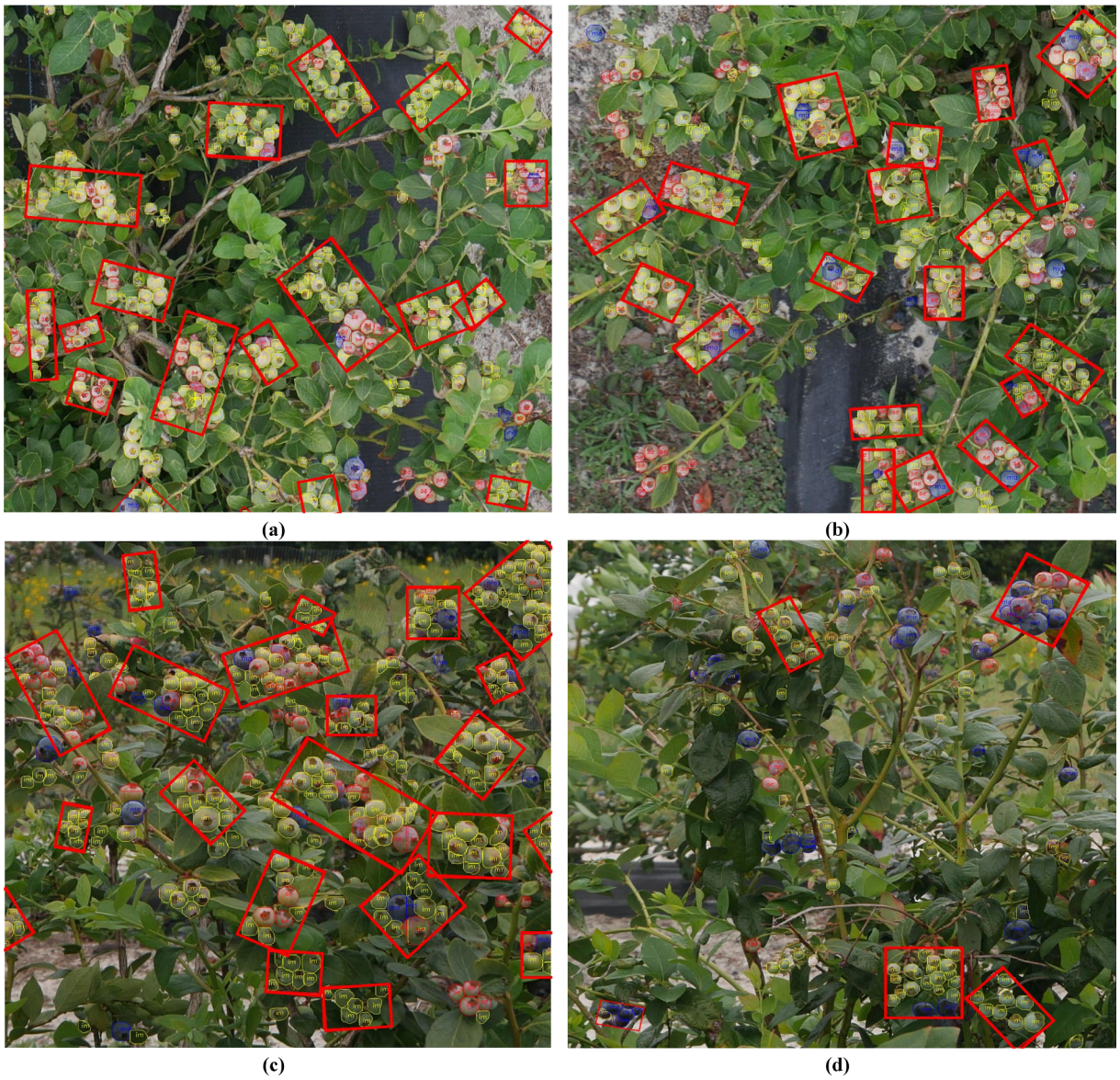


Fig. 12. Visualization of fruit cluster detection and fruit segmentation using BerryNet. (a)-(d) were the visualized examples of different views from the same plant: (a) and (b) are from the top view, while (c)and (d) are from the side view. Red rectangles indicated the predicted fruit clusters; the yellow, red, and blue masks represented the predictions for immature, semi-mature, and mature berries, respectively, labeled as 'im', 'se', and 'ma'.

3.4. Time-series analysis for genotypes comparison

With the existing feature extraction methods, the plant-level and cluster-level phenotypic traits were generated for breeders and growers (Fig. 18). This example highlighted the successful implementation of BerryNet which was capable of accurately identifying and segmenting individual fruit within a cluster. This high-resolution data from each genotype can be provided to breeders with critical information to assess genotypic differences based on yield, maturity, and compactness.

Analyzing the four-week field data of 11 genotypes revealed that different genotypes presented various phenotypic traits related to yield, maturity, and compactness using the parameters of fruit count, cluster

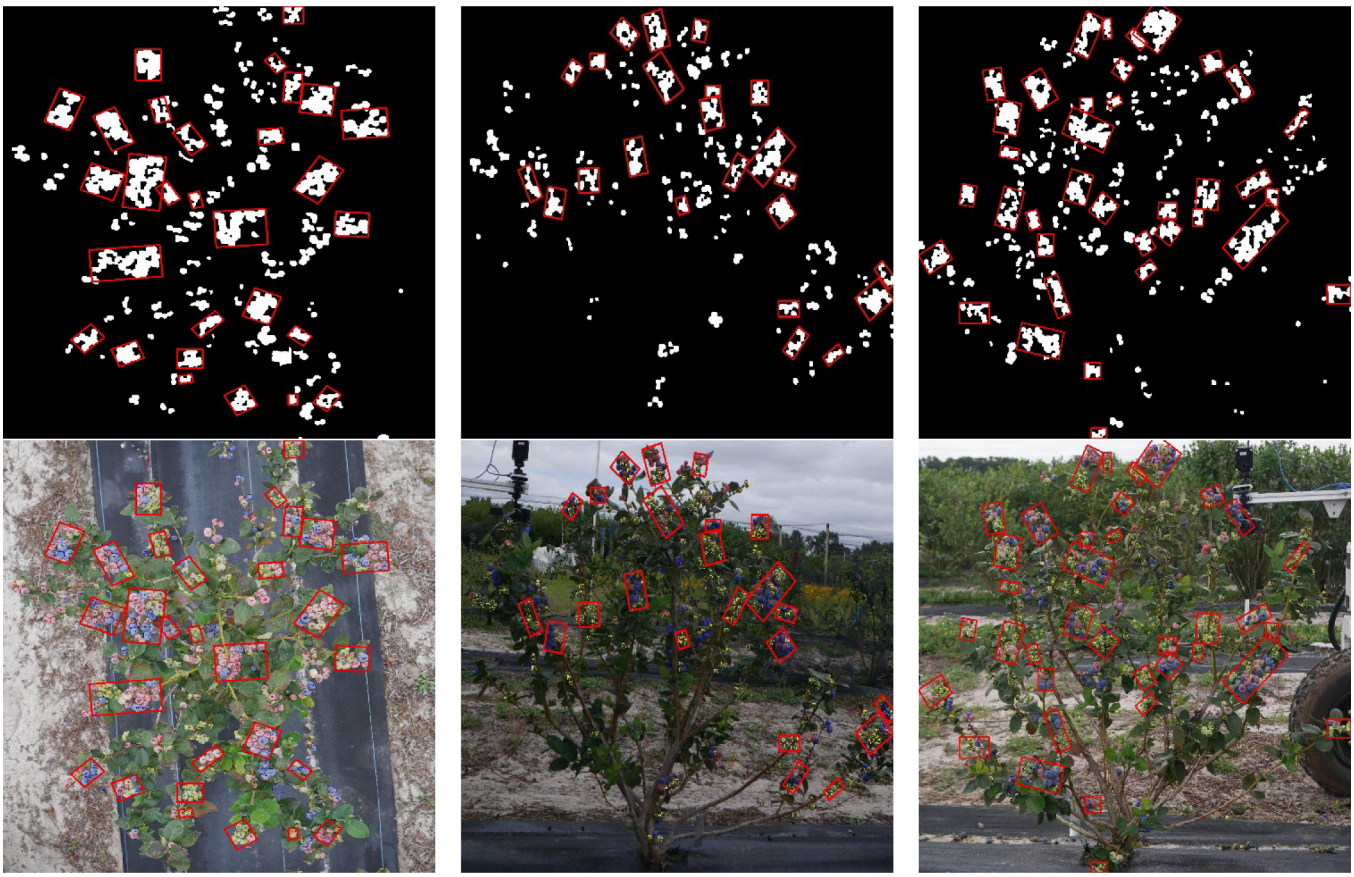


Fig. 13. Illustration of blueberry cluster extraction result (genotype: FLR-13–218). From left and right were the extraction results of the top, left, and right views. The top images were the binary mask of berries and the clusters, while the bottoms were the visualization of the raw images.

count, maturity rate, cluster-level compactness, and fruit distance (Fig. 19).

In general, fruit count and cluster count displayed similar rankings and significant differences across genotypes. Both parameters exhibited a trend of first increasing, then decreasing over time. This can be attributed to the developmental stage of most genotypes: early in the study, many plants were transitioning from flowering to fruiting, which caused an initial increase in both fruit and cluster counts. As the fruit matured, they began to drop, resulting in a gradual decline in these counts. The maturity rate across most genotypes steadily increased over time before stabilizing, reflecting the progression of fruit development. However, for most genotypes, the maturity rate remained below 60 % throughout the study. This was likely due to asynchronous fruit ripening, where not all fruit matured simultaneously. As the ripe fruit dropped, the overall maturity rate stabilized at a relatively constant level. Regarding the compactness metrics, the cluster-level compactness of most genotypes remained within a stable range, with a slight increase over time. This gradual rise can be attributed to the continuous increase in fruit size, which caused the clusters to become tighter. The fruit distance remained relatively stable for each genotype during the fruiting stage, as it was normalized to account for variations in fruit size. A smaller fruit distance indicates a tighter fruit cluster, reflecting the compactness and arrangement of the berries within each cluster. This stability suggests that fruit distance is more consistent within genotypes and primarily reflects the spatial distribution of fruit rather than their size.

Specifically, genotypes such as Windsor and FLR03-228 consistently displayed higher fruit count and cluster count, as well as the shortest average fruit distance, suggesting a strong yield potential. However, their maturity level was relatively lower than other genotypes at the same period, indicating they were late-mature genotypes. After May 9, their maturity level continued to rise, but the fruit maturity level remained below 40 %. This suggested that these genotypes may require extended growth periods or delayed harvests to reach optimal maturity, thereby affecting market availability. Conversely, genotypes such as Optimus reached the highest maturity level and maintained competitive yield from April to May, positioning itself as a preferable choice for the early-season market.

Cluster-level compactness and fruit distance revealed preferences in physical berry arrangement that could influence machine harvestability. Overall, the cluster-level compactness among different genotypes was relatively uniform due to the defined clustering parameter, with values ranging from 0.45 ~ 0.7. Genotypes such as 'Acadia' and 'Windsor' exhibited tighter clustering, which may be less favorable for both machine and hand harvesting. The mature fruit in these dense clusters requires greater force to detach, whether by shaking, air flow or manual selection. Additionally, tighter clusters often contain fruit at varying stages of maturity, making it difficult to selectively harvest only the mature berries. In contrast, FLR14-442 and Sentinel showed lower cluster compactness with lower cluster-level compactness and higher fruit distance, which may be more suitable for machine and hand harvesting.

4. Discussion

4.1. Multi-view imaging robotic system

This study successfully developed a multi-view imaging robotic system for automated in-field blueberry fruit phenotyping. The multi-view imaging robotic system improved upon methods that rely on top or side views alone and provided more comprehensive coverage of plants by capturing the top, left, and right perspectives. The visibility

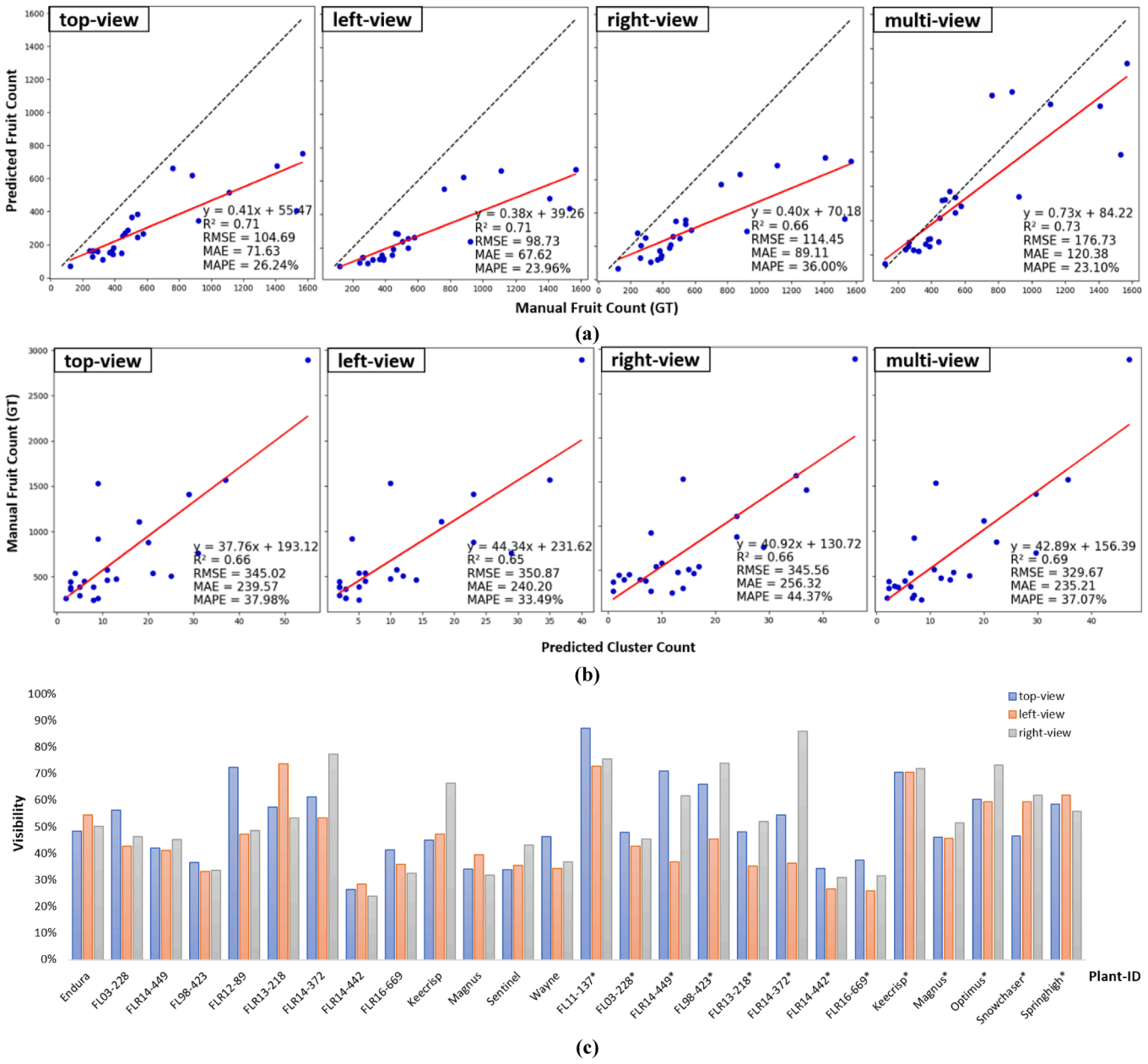


Fig. 14. Evaluation of yield estimation with BerryNet across single-view or multi-view approach. (a) Fruit counts regression using predicted fruit counts from single or multi-view; (b) Fruit counts regression using predicted cluster counts from single or multi-view; (c) comparison of blueberries visibility of three views. The plants with an asterisk (*) were harvested on April 12, and the others were on May 9, 2023.

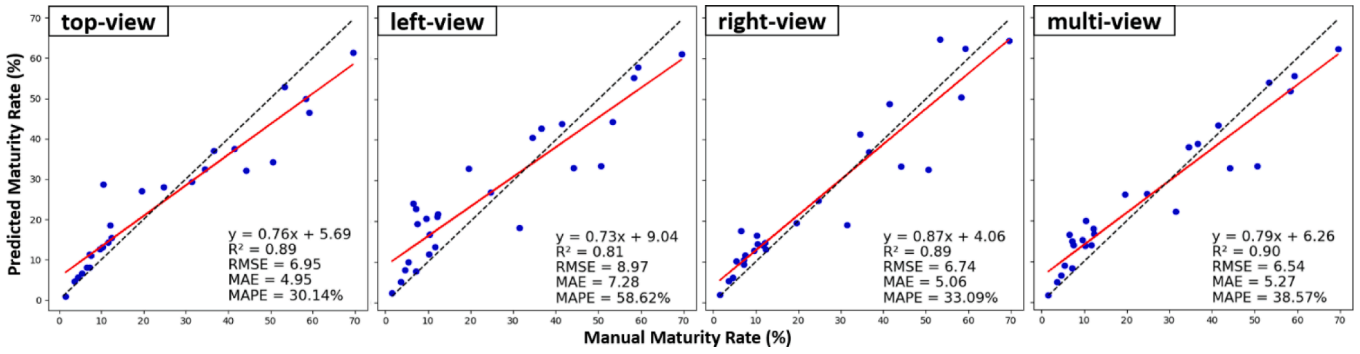
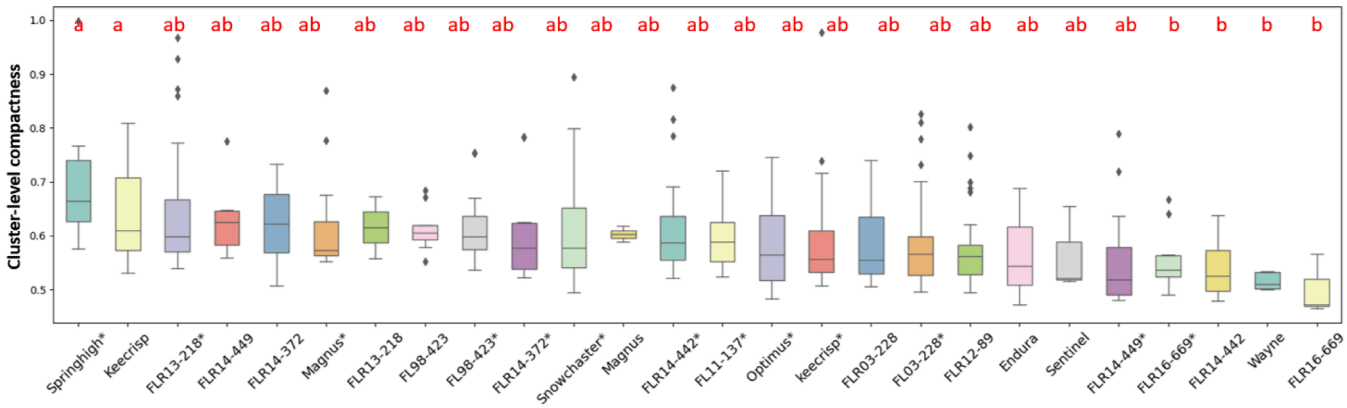


Fig. 15. Evaluation of maturity estimation with BerryNet. Linear regression from single/multi-view for maturity rate estimation.



(a)

(b)

Fig. 16. Comparison of cluster-level compactness among different genotypes. (a) Boxplot of Cluster-level compactness among 26 plants; from left to right, the cluster-level compactness becomes smaller, which means the lower compactness. The red labels above the boxes are significant letters. (b) Examples of cluster-level compactness from different plants.

statistics showed that images from a single view captured only about half or sometimes even fewer, of the actual berries present. The multi-view approach could reduce the likelihood of missing berries that were occluded or not visible from a single perspective. Additionally, the fusion of data from different angles could help overcome challenges related to plant morphology and berry occlusion, providing a more reliable assessment of fruit traits.

The automated phenotyping process also greatly accelerated data collection efficiency compared to traditional manual methods. The system's ability to operate autonomously in the field reduced labor requirements and enabled more frequent data collection over the growing season, leading to more detailed tracking of plant development. This increase in data resolution offers greater insights into genotype performance under different environmental conditions, potentially aiding in more informed breeding decisions. Data integration from multiple views remains a significant challenge. The current three-view configuration lacks sufficient overlap between camera perspectives, making it difficult to register the cameras to identify or eliminate duplicates of the same

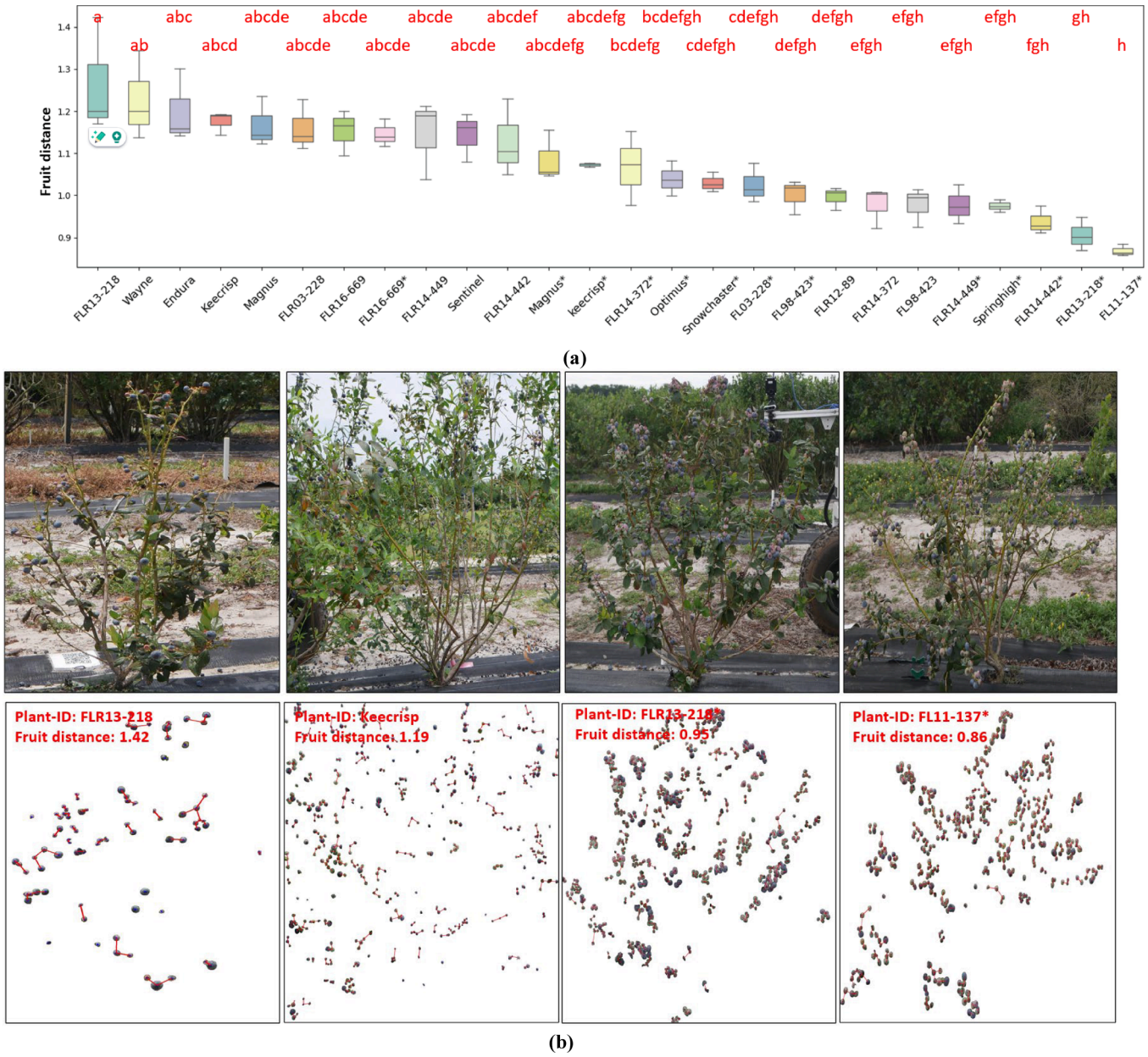


Fig. 17. Comparison of fruit distance among different genotypes. (a) Fruit distance: from left to right, the fruit distance becomes smaller, which means the compactness is higher. The red labels above the boxes are significant letters. (b) Visualization of fruit distance: from the left to the right, the fruit distance becomes smaller and means higher compactness.

clusters or berries across different cameras. This issue should be further investigated in future studies. A potential solution to improve multi-view alignment is to adopt a circular camera array with additional cameras. This configuration would increase the overlap between views, allowing for more precise alignment, better detection of duplicate clusters or berries, and improved overall accuracy in the registration process. However, this approach would also increase equipment cost and computational demand. Integrating a depth sensor is also a potential solution for improving multi-view alignment by transforming data into a unified 3D coordinate system. However, the resolution limitations of current commercial depth sensors could pose challenges in detecting small or partially occluded berries, especially when mounted on a robot and used in outdoor environments. Variations in lighting, movement, and complex plant structures may further exacerbate these challenges, impacting the accuracy of the sensor's data in real-world applications.

There are several challenges to implementing this system in large blueberry fields. For example, navigating the platform over the plants in fields with various plant sizes requires careful dimension adjustments to prevent interference with plants that have varying architectures and dimensions. In addition, the proximity of larger plants to the camera can also prevent complete data capture, reducing the overall coverage and making it difficult to accurately assess the entire plant structure. One potential solution is to navigate the robot between the rows to minimize interaction with the plants. However, this approach would require scanning both sides separately and then aligning the left and right data, which introduces additional challenges and may increase processing time.

4.2. *Automated pixel-wise labeling method*

The automated pixel-wise labeling method developed in this study utilizes vision foundation models, notably the Segment Anything Model

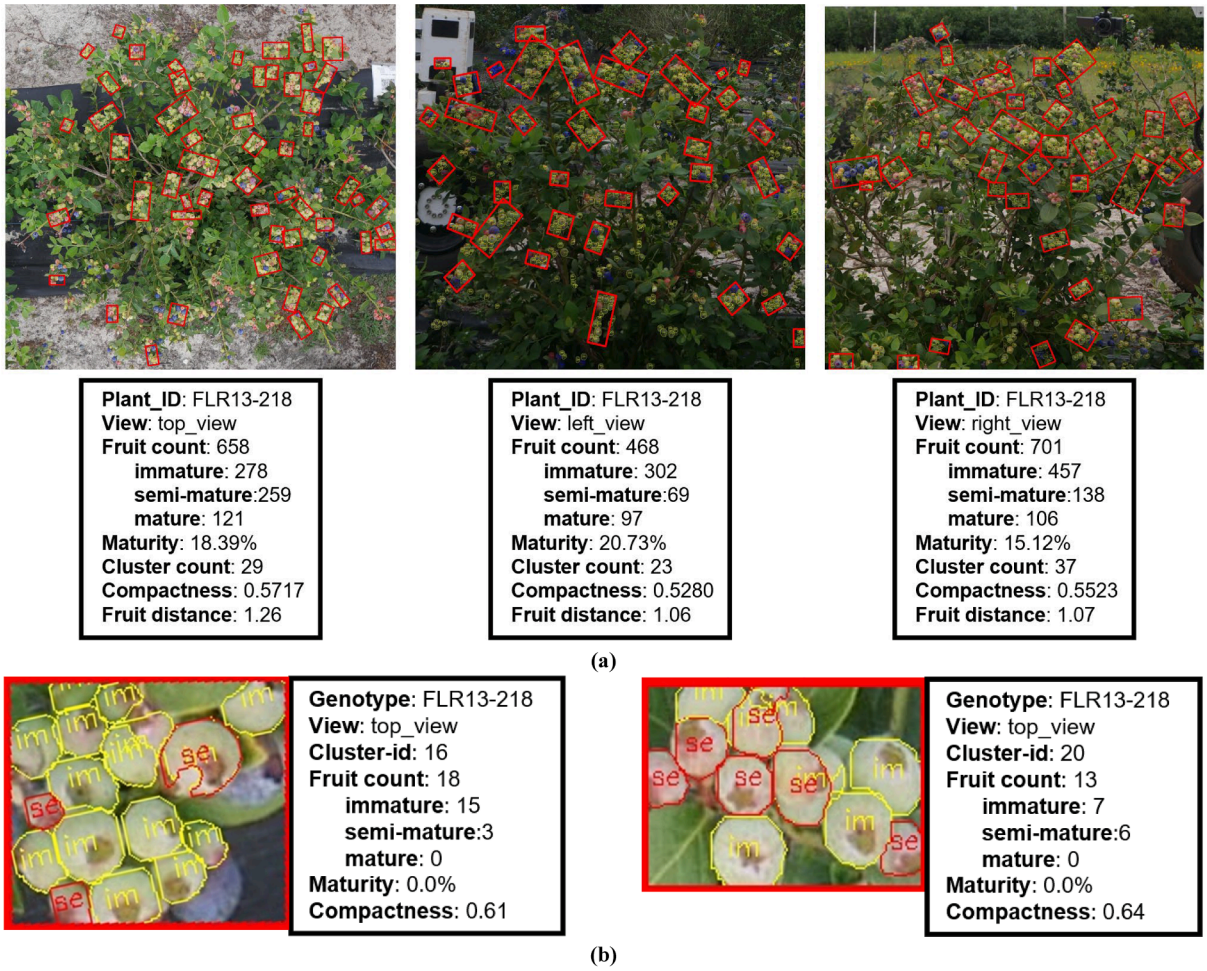


Fig. 18. Demonstrations of plant-level and cluster-level phenotypic traits extraction. (a) Plant-level phenotyping result from different views of the plant FLR13-218. (b) Examples of cluster-level phenotyping results.

(SAM), to significantly reduce the challenges of manual annotation. By leveraging SAM's zero-shot capabilities, this method efficiently transforms existing bounding box annotations into detailed, class-specific masks, reducing the need for extensive human intervention and making the data suitable for complex phenotyping and genotype analysis. The automated pixel-wise labeling method helps eliminate variations in annotators' perceptions of berries and clusters, reducing labeling inconsistencies that could compromise model training accuracy.

The SAM-based segmentation method still faces limitations due to potential ambiguity in prompts and challenging environmental conditions. When prompts are ambiguous or fail to accurately target the desired objects, the segmentation masks may mistakenly include non-fruit elements, such as leaves, branches, or trunks. However, in our work on blueberry detection, the method achieved over 92 % mIoU, demonstrating that bounding boxes serve as highly effective prompts. These failures mainly occur in situations with significant occlusion, especially when the occlusion exceeds 40 %, making it difficult for the model to accurately determine the primary object. Additionally, in low-contrast or poor lighting conditions, the boundaries of the blueberries may become unclear, leading to inaccurate segmentation and potential misclassification. Incorporating advanced text prompts that describe the desired fruit characteristics could be a helpful solution for reducing ambiguity in segmentation. While this approach has the potential to improve accuracy, it still requires human involvement to monitor and slightly modify the incorrect labels rather than completely replacing human expertise in the process. The training data must be as accurate as possible, as errors in the ground truth can propagate through the model and result in an unreliable performance. This result might mistakenly be attributed to the model's inadequacy rather than to inaccuracies in the annotation data.

4.3. *BerryNet: Cluster detection and fruit segmentation*

BerryNet provides substantial advantages for blueberry phenotyping, particularly in accurately segmenting blueberry fruit. It has demonstrated the potential to address challenges arising from the small size, reduced visibility, and minimal geometric cues associated with small objects (Chen et al., 2022). By incorporating architectural enhancements, such as a lower backbone layer, BiFPN for feature fusion, and the FasterNet Block for efficient spatial feature extraction, BerryNet offers a tailored solution to the complexities of detecting individual berries within clusters, where overlapping features and variable visibility often pose difficulties. However, BerryNet's cluster detection performance was less ideal than fruit segmentation, despite still outperforming other popular models. The difficulty in detecting fruit clusters arises from their inconsistent patterns, as clusters can exhibit various shapes and fruit distributions. For example, some clusters may be tightly grouped, grow along trunks, or be sparsely distributed with only a few fruits. Additionally, occlusions can obscure parts of the clusters, making detection even more challenging. This variability also affects manual labeling, as different annotators may label clusters inconsistently, introducing further difficulties into model training.

Additionally, its different model configurations provide flexible selections for specific applications. For example, the lightweight nano

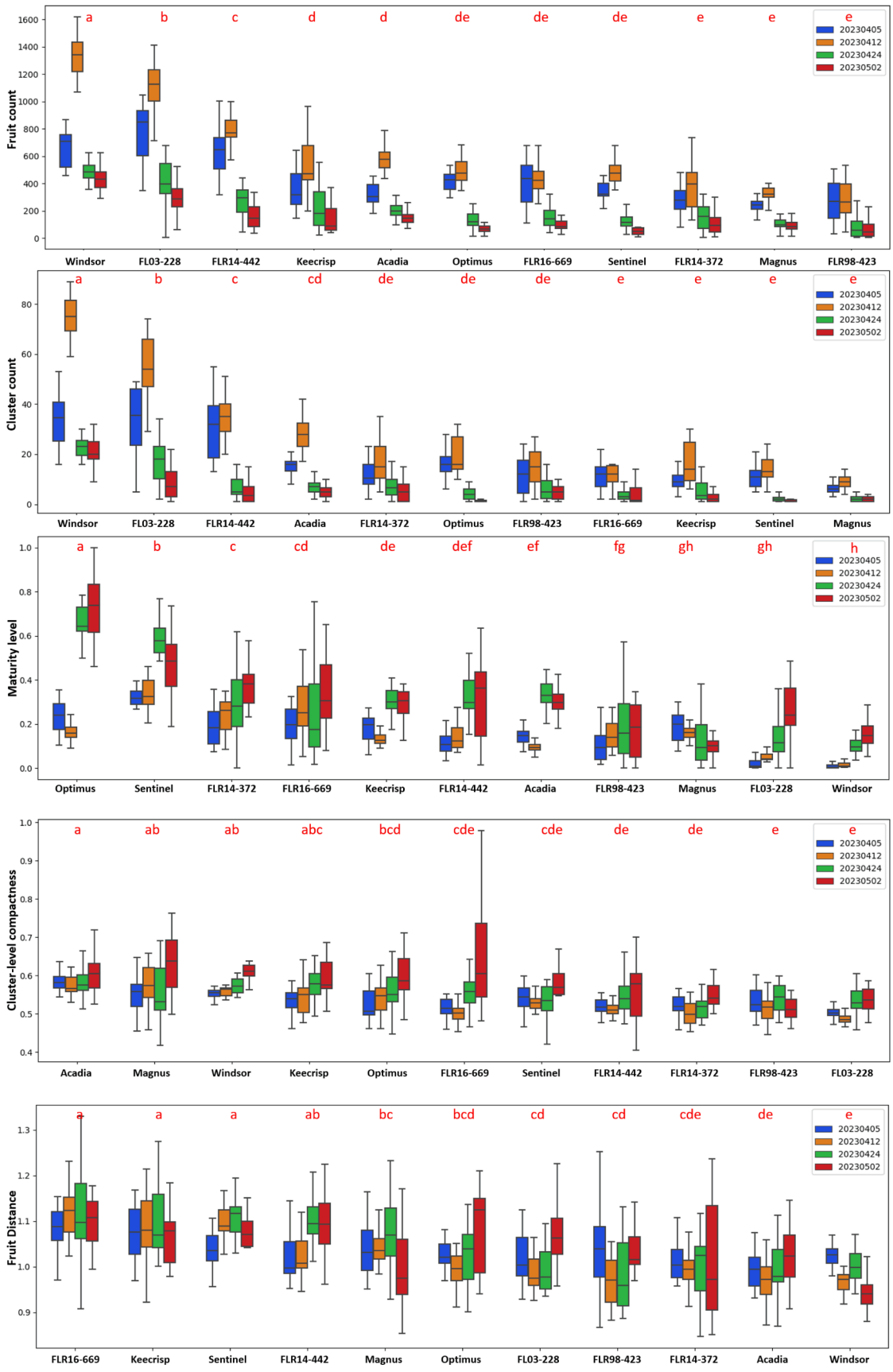


Fig. 19. Illustration of the phenotypic traits analysis among different blueberry genotypes over four weeks based on our proposed pipeline. From top to bottom, there are the traits comparison of fruit count, cluster count, maturity level, cluster-level compactness, and fruit distance. A total of 11 genotypes were analyzed during the fruiting period from April to May with four times data collection.

version is suitable for real-time applications on mobile or edge devices such as the NVIDIA Jetson single-board computers for achieving real-time in-field phenotyping. While the large version enables more accurate prediction but requires higher computational resources, which is suitable for offline analysis.

4.4. Phenotypic traits extraction

The phenotypic trait extraction method shows standardized, imaging-based approaches potentially replacing human involvement in data sampling and analysis. Yield traits demonstrated a strong correlation with predicted berry and cluster counts, both in single-view and multi-view regression models, indicating that this method could be deployed for blueberry yield estimation tasks. However, crop yield is also influenced by individual fruit weight, which varies across different genotypes. Thus, considering fruit size is equally important. Current sensors, such as stereo or depth cameras, face limitations in terms of resolution and image quality in outdoor scenarios, making it difficult to capture accurate and stable depth data in field situations, especially when adapting robotic systems. Challenges arise due to the complex plant structure, varying lighting conditions, and motion during image capture. One potential solution is to apply the canopy and artificial lighting to provide more accurate illumination. The centimeter-level depth accuracy of these sensors makes it difficult to precisely detect and measure individual berries, which typically have diameters of only 1 to 2 cm. While commercial LiDAR systems are not affected by lighting conditions, their resolution is often insufficient to distinguish blueberries with precision. Multi-sensor fusion could be a potential approach to enhance imaging stability by combining stereo cameras with LiDAR or high-resolution RGB cameras to compensate for individual sensor limitations. Additionally, incorporating adaptive lighting solutions, such as flash or polarized light, can enhance image quality by mitigating varying lighting conditions, leading to more stable and accurate measurements of individual berries during robotic scanning. The use of artificial lighting also enables the robotic system to operate effectively at night, extending its functionality beyond daylight hours and ensuring consistent imaging performance regardless of ambient lighting. This enhancement not only improves the accuracy of berry detection but also increases the system's operational flexibility and efficiency.

The maturity estimation performance was satisfactory, achieving an MAE of around 5 %, making it a viable alternative to manual assessments by breeders in breeding programs or on commercial farms. By accurately predicting fruit maturity, growers can optimize harvest schedules, improve labor efficiency, and reduce the likelihood of picking fruit that is either underripe or overripe. The next step can consider integrating field mapping and GPS location data, the robotic system can provide maturity levels for individual plants, enabling precise genotype comparisons and generating time-series reports on maturity changes. This capability assists breeders in selecting early-mature genotypes and determining the optimal harvest time, significantly improving decision-making in blueberry breeding programs or enhancing profitability for commercial farms.

The two harvesting-related metrics, cluster-level compactness and fruit distance, offer valuable insights into local and global compactness, aiding in harvesting decisions. However, as occlusion and the 3D shape of clusters can distort calculations, cluster-level compactness is limited by 2D imaging to capture accurate true compactness. Overlapping berries and leaves often leads to underestimation of density, and the accuracy of these metrics is further impacted by the effectiveness of the cluster detection algorithms. Errors in detecting clusters can exacerbate inaccuracies, reducing the reliability of the compactness metric. To overcome these challenges, a manipulator equipped with a camera to scan 3D images of fruit clusters could provide a more accurate solution by generating spatially precise representations of fruit clusters. This would enable more accurate compactness measurements but would reduce the throughput of data collection and require advanced scanning techniques to capture multiple viewpoints around the plant.

5. Conclusions

Our work demonstrated the potential of an in-field robotic mobile system for *in situ* high-throughput phenotyping (HTP) of blueberry fruit traits, making a significant advancement from traditional manual assessments. This robotic system, integrated with an automated labeling method using the Segment Anything Model (SAM), reduced the burden of manual annotations and minimizes human errors, enhancing data reliability for blueberry phenotyping. The customized deep learning model BerryNet effectively detected clusters and segmented berries in the images, providing phenotypic traits such as yield, maturity, and compactness. These traits not only benefit breeders and growers in selecting high-yield and early-mature genotypes for higher market profits but also provide decision support for mechanical and hand harvesting. Future research directions could include more effectively fusing the three-view images using a stereo-imaging approach and estimating the fruit size. The MARS-Phenobot system enables autonomous data collection and continuous monitoring of fruit-related phenotypic traits, providing valuable information for breeding programs and crop management.

Four datasets mentioned in this paper have been released in Kaggle:

Dataset	Link
Blueberry Fruit Detection	https://www.kaggle.com/datasets/zhengkunli3969/blueberry-detection-dataset
Blueberry Maturity Classification	https://www.kaggle.com/datasets/zhengkunli3969/blueberry-maturiety-classification
Blueberry Pixel-wise Segmentation	https://www.kaggle.com/datasets/zhengkunli3969/blueberry-segmentation-with-segment-anything-model
Blueberry Cluster Detection	https://www.kaggle.com/datasets/zhengkunli3969/blueberry-cluster-detection

CRediT authorship contribution statement

Zhengkun Li: Writing – original draft, Visualization, Validation, Software, Investigation, Data curation. **Rui Xu:** Writing – review & editing, Methodology, Investigation, Data curation. **Changying Li:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Patricio Munoz:** Writing – review & editing, Resources, Methodology, Funding acquisition. **Fumiomi Takeda:** Writing – review & editing, Methodology. **Bruno Leme:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the University of Florida IFAS LIFT AI seed grant, Hatch Project (FLA-ABE-006451), and the University of Florida blueberry breeding program.

Data availability

We have shared the data via a Github link.

References

Aguilera, C.A., Figueroa-Flores, C., Aguilera, C., Navarrete, C., 2023. Comprehensive analysis of model errors in blueberry detection and maturity classification: identifying limitations and proposing future improvements in agricultural monitoring. Agriculture 14 (1), 18.

Bai, X., Li, Z., Li, W., Zhao, Y., Li, M., Chen, H., Wei, S., Jiang, Y., Yang, G., Zhu, X., 2021. Comparison of machine-learning and casa models for predicting apple fruit yields from time-series planet imageries. Remote Sens. 13 (16), 3073.
Chang, Y.K., Zaman, Q., Farooque, A.A., Schumann, A.W., Percival, D.C., 2012. An automated yield monitoring system II for commercial wild blueberry double-head harvester. Comput. Electron. Agric. 81, 97–103.
Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., & Xu, J. (2019). MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.
Chen, J., Kao, S.-h., He, H., Zhuo, W., Wen, S., Lee, C.-H., & Chan, S.-H. G. (2023). Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J., Kumar, V., 2017. Counting apples and oranges with deep learning: a data-driven approach. IEEE Rob. Autom. Lett. 2 (2), 781–788. https://doi.org/10.1109/lra.2017.2651944.
Chen, G., Wang, H.T., Chen, K., Li, Z.J., Song, Z.D., Liu, Y.L., Chen, W.K., Knoll, A., 2022. A survey of the four pillars for small object detection: multiscale representation, contextual information, super-resolution, and region proposal. Ieee Trans. Syst. Man Cybernet.-Syst. 52 (2), 936–953. https://doi.org/10.1109/tsmc.2020.3005231.
Fang, Y., Yang, S., Wang, X., Li, Y., Fang, C., Shan, Y., Feng, B., & Liu, W. (2021). Instances as queries. Proceedings of the IEEE/CVF international conference on computer vision.
Gai, R., Gao, J., Xu, G., 2024. HPPEM: a high-precision blueberry cluster phenotype extraction model based on hybrid task cascade. Agronomy 14 (6), 1178.
Gonzalez, S., Arellano, C., Tapia, J.E., 2019. Deepblueberry: quantification of blueberries in the wild using instance segmentation. IEEE Access 7, 105776–105788. https://doi.org/10.1109/access.2019.2933062.
Gui, B., Bhardwaj, A., Sam, L., 2024. Evaluating the efficacy of segment anything model for delineating agriculture and urban green spaces in multiresolution aerial and spaceborne remote sensing images. Remote Sens. (Basel) 16 (2), 414.
Gutierrez, S., Wendel, A., Underwood, J., 2019. Ground based hyperspectral imaging for extensive mango yield estimation. Comput. Electron. Agric. 157, 126–135. https://doi.org/10.1016/j.compag.2018.12.041.
Haydar, Z., Esau, T.J., Farooque, A.A., Zaman, Q.U., Hennessy, P.J., Singh, K., Abbas, F., 2023. Deep learning supported machine vision system to precisely automate the wild blueberry harvester header. Sci. Rep. 13 (1), 10198.
Haydar, Z., Esau, T.J., Farooque, A.A., Bilodeau, M.F., Zaman, Q.U., Abbas, F., Yaqoob, N., 2024. Assessing UAV-based wild blueberry plant height mapping-a consideration for wild blueberry harvester automation. Smart Agric. Technol. 8, 100456.
He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. Proceedings of the IEEE international conference on computer vision.
He, L., Fang, W., Zhao, G., Wu, Z., Fu, L., Li, R., Majeed, Y., Dhupia, J., 2022. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. Comput. Electron. Agric. 195, 106812.
Jocher, G., Chaurasia, A., Qiu, J., 2024. Ultralytics YOLO (Version 8.3.13) [Computer software]. Retrieved October 2024 from https://docs.ultralytics.com/models/yolo11/.
Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Fang, J., Michael, K., Montes, D., Nadar, J., & Skalski, P. (2022). ultralytics/yolov5: v6. 1-tensorrt, tensorflow edge tpu and openvino export and inference. *Zenodo*.
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., & Lo, W.-Y. (2023). Segment anything. *arXiv preprint arXiv:2304.02643*.
Kolahdouzan, M., Shahabi, C., 2004. Voronoi-based k nearest neighbor search for spatial network databases. Proceedings of the Thirtieth International Conference on Very Large Data Bases-Volume 30.
Li, Z., Li, C., & Munoz, P. (2023). Blueberry Yield Estimation Through Multi-View Imagery with YOLOv8 Object Detection. 2023 ASABE Annual International Meeting.
Li, H., Lee, W.S., Wang, K., 2014. Identifying blueberry fruit of different growth stages using natural outdoor color images. Comput. Electron. Agric. 106, 91–101.
Li, Y., Wang, D., Yuan, C., Li, H., Hu, J., 2023. Enhancing agricultural image segmentation with an agricultural segment anything model adapter. Sensors 23 (18), 7884.
Liu, Y., Zheng, H., Zhang, Y., Zhang, Q., Chen, H., Xu, X., Wang, G., 2023. "Is this blueberry ripe?": a blueberry ripeness detection algorithm for use on picking robots. Front. Plant Sci. 14, 1198650.
Lyu, C., Zhang, W., Huang, H., Zhou, Y., Wang, Y., Liu, Y., Zhang, S., & Chen, K. (2022). Rtmdet: An empirical study of designing real-time object detectors. *arXiv preprint arXiv:2212.07784*.
MacEachern, C.B., Esau, T.J., Schumann, A.W., Hennessy, P.J., Zaman, Q.U., 2023. Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks. Smart Agric. Technol. 3, 100099.
Morgan, K. L. (2022). Market Trends for US Berry Crops: Implications for Florida Blueberry, Blackberry, and Raspberry Producers: FE1123/FE1123, 11/2022. *EDIS*, *2022*(6).
Mudassar, B.A., Mukhopadhyay, S., 2019. Rethinking convolutional feature extraction for small object detection. BMVC.
Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R.J., Fredes, C., Valenzuela, A., 2020. A review of convolutional neural network applied to fruit image processing. Appl. Sci. 10 (10), 3443.
Nguyen, K. D., Phung, T.-H., & Cao, H.-G. (2023). A SAM-based solution for hierarchical panoptic segmentation of crops and weeds competition. *arXiv preprint arXiv: 2309.13578*.
Ni, X.P., Li, C.Y., Jiang, H.Y., Takeda, F., 2020. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. Hortic. Res. 7 (1), 14. https://doi.org/10.1038/s41438-020-0323-3.
Ni, X.P., Li, C.Y., Jiang, H.Y., Takeda, F., 2021. Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits. ISPRS J. Photogramm. Remote Sens. 171, 297–309. https://doi.org/10.1016/j.isprsjprs.2020.11.010.
Niedbała, G., Kurek, J., Świderski, B., Wojciechowski, T., Antoniuk, I., Bobran, K., 2022. Prediction of Blueberry (Vaccinium corymbosum L.) yield based on artificial intelligence methods. Agriculture 12 (12), 2089.
Patrick, A., Li, C., 2017. High throughput phenotyping of blueberry bush morphological traits using unmanned aerial systems. Remote Sens. (Basel) 9 (12), 1250.
Pham, T.Q., Van Vliet, L.J., Schutte, K., 2006. Robust fusion of irregularly sampled data using adaptive normalized convolution. EURASIP J. Adv. Signal Process. 2006, 1–12.
Qu, H., Zheng, C., Ji, H., Barai, K., Zhang, Y.-J., 2024. A fast and efficient approach to estimate wild blueberry yield using machine learning with drone photography: flight altitude, sampling method and model effects. Comput. Electron. Agric. 216, 108543.
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition.
Ren, S. Q., He, K. M., Girshick, R., & Sun, J. (2015, Dec 07-12). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.*Advances in Neural Information Processing Systems* [Advances in neural information processing systems 28 (nips 2015)]. 29th Annual Conference on Neural Information Processing Systems (NIPS), Montreal, CANADA.
Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.
Schumann, A. W., Mood, N. S., Mungofa, P. D., MacEachern, C., Zaman, Q., & Esau, T. (2019). Detection of three fruit maturity stages in wild blueberry fields using deep learning artificial neural networks. 2019 ASABE Annual International Meeting.
Swain, K.C., Zaman, Q.U., Schumann, A.W., Percival, D.C., Bochtis, D.D., 2010. Computer vision system for wild blueberry fruit yield mapping. Biosyst. Eng. 106 (4), 389–394.
Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
Tan, K., Lee, W.S., Gan, H., Wang, S., 2018. Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes. Biosyst. Eng. 176, 59–72.
Terven, J., & Cordova-Esparza, D. (2023). A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond. *arXiv preprint arXiv:2304.00501*.
Tripathy, P., Baylis, K., Wu, K., Watson, J., & Jiang, R. (2024). Investigating the Segment Anything Foundation Model for Mapping Smallholder Agriculture Field Boundaries Without Training Labels. *arXiv preprint arXiv:2407.01846*.
Van Beek, J., Tits, L., Somers, B., Deckers, T., Verjans, W., Bylemans, D., Janssens, P., Coppin, P., 2015. Temporal dependency of yield and quality estimation through spectral vegetation indices in pear orchards. Remote Sens. (Basel) 7 (8), 9886–9903.
Vasconez, J.P., Delpiano, J., Vougioukas, S., Cheein, F.A., 2020. Comparison of convolutional neural networks in fruit detection and counting: a comprehensive evaluation. Comput. Electron. Agric. 173, 105348.
Wang, X., Zhang, R., Kong, T., Li, L., Shen, C., 2020. Solov2: dynamic and fast instance segmentation. Adv. Neural Inf. Proces. Syst. 33, 17721–17732.
Williams, D., Macfarlane, F., Britten, A., 2024. Leaf only SAM: a segment anything pipeline for zero-shot automated leaf segmentation. Smart Agric. Technol., 100515
Williams, H., Ting, C., Nejati, M., Jones, M.H., Penhall, N., Lim, J., Seabright, M., Bell, J., Ahn, H.S., Scarfe, A., Duke, M., MacDonald, B., 2020. Improvements to and large-scale evaluation of a robotic kiwifruit harvester. J. Field Rob. 37 (2), 187–201. https://doi.org/10.1002/rob.21890.
Xu, R., Li, C.Y., 2022. A modular agricultural robotic system (MARS) for precision farming: concept and implementation. J. Field Rob. 23. https://doi.org/10.1002/rob.22056.
Yang, W. J., Ma, X. X., Hu, W. C., & Tang, P. J. (2022). Lightweight Blueberry Fruit Recognition Based on Multi-Scale and Attention Fusion NCBAM [Article]. *Agronomy-Basel*, *12*(10), 13, Article 2354. Doi: 10.3390/agronomy12102354.
Yang, C., Lee, W.S., Williamson, J.G., 2012. Classification of blueberry fruit and leaves based on spectral signatures. Biosyst. Eng. 113 (4), 351–362.
Yu, S., Liu, X., Tan, Q., Wang, Z., Zhang, B., 2024. Sensors, systems and algorithms of 3D reconstruction for smart agriculture and precision farming: a review. Comput. Electron. Agric. 224, 109229.
Zhang, C., Liu, L., Cui, Y., Huang, G., Lin, W., Yang, Y., & Hu, Y. (2023). A Comprehensive Survey on Segment Anything Model for Vision and Beyond. *arXiv preprint arXiv:2305.08196*.
Zhang, C., Marfatia, P., Farhan, H., Di, L., Lin, L., Zhao, H., Li, H., Islam, M. D., & Yang, Z. (2023). Enhancing USDA NASS cropland data layer with segment anything model. 2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics).
Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., et al., 2024. Detrs beat yolos on real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16965–16974.
Zheng, Z., Xiong, J., Wang, X., Li, Z., Huang, Q., Chen, H., & Han, Y. (2022). An efficient online citrus counting system for large-scale unstructured orchards based on the unmanned aerial vehicle. *Journal of Field Robotics*, *n/a*(n/a). https://doi.org/Doi: 10.1002/rob.22147.
Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as points. *arXiv preprint arXiv: 1904.07850*.