2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
DOI: https://doi.org/10.13031/aim.202300883
Paper Number: 2300883

Blueberry Yield Estimation Through Multi-View Imagery with YOLOv8 Object Detection

Zhengkun Li[a], Changying Li[a], Patricio Munoz[b]
a. Bio-Sensing, Automation and Intelligence Lab, Agricultural and Biological Engineering, University of Florida, Gainesville, Florida, USA
b. Horticultural Science Department, University of Florida, Gainesville, Florida, USA

Written for presentation at the
2023 ASABE Annual International Meeting
Sponsored by ASABE
Omaha, Nebraska
July 9-12, 2023

ABSTRACT.

Accurately estimating blueberry yields is crucial for farmers aiming to optimize crop management practices and enhance agricultural profitability. However, the inherent challenges posed by blueberries growing in clusters and being frequently occluded by leaves or other fruits make direct counting of individual blueberries nearly impossible. Existing approaches rely on sampling a few plants or clusters to estimate yields based on expert knowledge or employ indirect regression analysis using yield-related features as inputs for predictive models. With recent advancements in deep learning technologies, there has been a growing interest in leveraging machine vision techniques to directly count fruits for yield estimation. In this paper, we propose a novel approach for blueberry yield estimation utilizing multi-view imagery in conjunction with the state-of-the-art YOLOv8 object detection framework. Our methodology involves a customized mobile platform equipped with a multi-camera sensing system that captures images of blueberry plants from three distinct views (top, left, and right) to ensure comprehensive coverage. We train a YOLOv8x model as the detector to accurately detect and localize individual blueberries within the images. Accounting for the overlapping information from the three views, we employ a regression model to estimate the total number of blueberries per plant. To evaluate the effectiveness of our approach, we compare single-view and multi-view methodologies and assess their estimation performance on 12 individual blueberry plants with varying genotypes. The multi-view imagery approach demonstrates promising results, exhibiting a mean absolute percentage error of 24.6% and an R2 value of 0.77. These figures represent a substantial improvement of 5.2% to 15.7% when compared to single-view approaches. Additionally, leveraging the predicted bounding boxes of blueberries, we are able to generate density maps that facilitate further phenotyping analysis. The methodology presented in this study holds significant potential for accurately and autonomously estimating blueberry fields, enabling the generation of high-resolution yield density maps, even at the individual plant level, with the aid of mobile robots.

Keywords. *Blueberry yield estimation, multi-view regression, blueberry detection, YOLO.*

1. Introduction

Blueberries, known for their popularity and nutritional value, are cultivated extensively in various regions worldwide. The global production of blueberries has exhibited a consistent upward trend, and prominent countries in this industry include the United States, Canada, Chile, and Mexico. Notably, the commercial blueberry sector in the United States has emerged as a frontrunner, boasting a substantial production volume. In the year 2021, the US commercial blueberry industry successfully harvested a staggering 774.1 million pounds of blueberries across an expansive area of 102,400 acres. This impressive accomplishment allowed the United States to lead in terms of total blueberry production volume. Additionally, the utilized production value of the US blueberry industry reached an impressive figure of $1.1 billion, signifying its economic significance. It is worth noting that this value accounted for 26% of the total berry crop value in the United States for the same year [1].

Accurate estimation of blueberry yields holds significant importance for blueberry farmers as it enables them to optimize crop management practices and enhance profitability in agricultural production[2]. Nevertheless, estimating blueberry yields poses challenges due to the unique growth pattern of the fruit. Blueberries tend to grow in clusters and are frequently obscured by leaves or other fruits on the plant, making it arduous to directly count individual berries.

In the past, two general approaches have been employed for estimating crop yields in blueberries: sampling a small number of plants or clusters to estimate the yield for the entire field based on expert knowledge, and indirect regression analysis that involves developing a predictive model using features indirectly related to yield as inputs [3]. Direct methods rely on collecting data from direct measurements of yield-forming generative organs, such as flowers, buds, or fruits. These measurements may encompass the number of these organs, their geometric dimensions, or weight. The data collection can be carried out manually or automatically using different types of stationary [4] or mobile ground platforms [5-7], as well as aerial platforms [8, 9]. RGB cameras are commonly mounted on these platforms to capture raw field data for further analysis [10]. Additionally, studies have shown the feasibility of using RGB-D, infrared, and hyperspectral cameras. Direct estimation methods generally require data to be collected in close proximity to the crops, allowing for more detailed observations, such as accounting for occlusion. However, direct estimation methods often face limitations in terms of sample size, which can lead to inaccurate yield estimates. To achieve accurate estimates of the entire crop yield using direct methods, a significant number of plants need to be measured, which can be time-consuming and labor-intensive. Furthermore, direct methods may be constrained by factors such as the timing of data collection, the availability of trained personnel to perform the measurements, and the cost of equipment.

In contrast to direct methods, indirect methods offer the advantage of providing yield estimates for larger areas without requiring extensive fieldwork. These approaches rely on data collected from remote sensors, such as satellites or drones [11, 12], which provide a wealth of information over a wide geographic region. However, indirect methods are subject to limitations, including data quality and the need for accurate sensor calibration. Indirect methods commonly utilize various features, including vegetation data (e.g., vegetation indices, canopy cover, plant height, biomass), historical and current weather data (e.g., temperature, precipitation, solar radiation), soil data (e.g., soil texture, pH, nutrient content), and agrotechnical processes [3]. For instance, Salvo et al. [13] estimated plant yields in commercial orchards by proposing models that related the number of available fruits for harvest to the number of flower buds and climatic variables. While these methods can offer valuable insights into blueberry yield estimation, prediction models often rely on abundant and diverse types of data from various sources. Therefore, managing large datasets, often referred to as Big Data [14], becomes increasingly important for storing, processing, sharing, and analyzing such information. Insufficient data may hinder the accuracy of indirect regression models in capturing the complex relationships between yield and other contributing factors.

Recent advances in machine learning and computer vision have led to the development of more sophisticated and accurate methods for estimating blueberry yield, both through direct and indirect approaches. There are two remarkable features: one is the regression approach which use machine learning, especially the neural network approach to achieve the higher accuracy of predictions; Another is blueberry detection based on deep learning, has great potential to replace manual counting for the number of blueberries [2, 15]. In the study conducted by Obsie et al. [16], the importance of bee species composition and weather factors in regulating wild blueberry agroecosystems was evaluated with computer simulation and machine learning algorithms. Similarly, Seireg et al. [17] used a combination of machine learning algorithms and meteorological data to predict wild blueberry yields with high accuracy. They employed stacking regression (SR) and cascading regression (CR) techniques with a novel combination of machine learning algorithms and used four feature engineering selection techniques to identify the best features for wild blueberry agroecosystems. Niedbala et al. [3] presents a high-accuracy model for blueberry yield prediction, using machine learning algorithms to find correlations between various factors at play, such as agronomic, climatic, soil and vegetation data. A total of 11 models were trained and evaluated, with Extreme Gradient Boosting algorithm performing the best, achieving a MAPE value of 12.48%. As for the blueberry detection with deep learning and computer vision, lightweight and accurate detector are the important aspect of studies. One-stage detector, the families of You Look Only Once (YOLO) [18, 19], are the most popular model in blueberry detection.

Schumann et al. [20] studied YOLOv3 as the detector to detect the green, red, and blue fruits to evaluate the maturity stage of wild blueberries. Their YoloV3-spp network performed the best with 91% recall and 28.3 ms inference time. Also, MacEachern et al. [21] compared YOLOv3, YOLOv4 and their variants for detecting ripeness stage in wild blueberries. These models were used to detect the 2-class (unripe and ripe) and 3-class (green, red and blue) blueberries. On average there was a 9.73% increase in mAP50 across all 2-class model and AP50 for the unripe class was superior to either the red or green. A non-linear power concave model is used to develop the regression yield prediction models, with YOLOv4-Small performing the best with a mean absolute error of 24.1%. Yang et al. [22] focus on developing a lightweight blueberry fruit recognition based on improved YOLOv5 through multi-scale and attention fusion NCBAM. Experimental results on a self-made blueberry dataset showed that the proposed method achieved a mean average precision (mAP) of 83.2%, which was 2.4% higher than the original network, indicating improved blueberry recognition accuracy. Besides, some studies used instance segmentation approach to count and evaluate blueberry cluster. Gonzalez et al. [23] use Mask R-CNN to quantification of blueberries in the wild using instance segmentation, which can achieve 75.9% detection accuracy. Ni et al. [24, 25] also use Mask R-CNN to segment individual blueberry in blueberry cluster images to extract fruit traits associate with harvestability and yield.

While recent advances in machine learning and computer vision have shown promising results in estimating blueberry yield, most research has focused on wild blueberry or small-scale blueberry clusters. Wild blueberries are often grown on low bushes that spread out in fields, and their fruits are typically visible from the top view. This makes it easier to capture images of the entire fruiting area and use computer vision techniques to count and estimate yield. However, commercial blueberry plants may have a different growth pattern, with fruits located at different heights on the bush. This can make it more challenging to accurately estimate yield using traditional imaging techniques. Capturing images with a single view may not be sufficient as a significant portion of fruits may remain hidden from view, resulting in a lower estimate of the yield. Furthermore, relying solely on detection in small-scale fruit cluster images may not provide a comprehensive understanding of the entire plant's yield potential. Additionally, using sampling techniques may introduce bias into the yield prediction models, leading to inaccurate estimates.

A potential solution to address these limitations could be to utilize multi-view imaging techniques. This approach involves capturing images of the blueberry plants from multiple angles, allowing for a more comprehensive view of the fruits' distribution across the plant. Additionally, incorporating the state-of-the-art computer vision and machine learning techniques could help in accurately detecting and counting the blueberry fruits. In this study, we present a potential approach for blueberry yield estimation through multi-view regression with the state-of-the-art YOLOv8 detector. The specific objectives are to (1) design a customized mobile platform integrated a multi-camera sensing and construct a blueberry dataset; (2) Compared the performance of YOLOv8 models with different configurations for detect small and dense object. (3) valid the effectiveness of blueberry yield estimation of multi-view regression.

2. Materials and Methods

2.1 Experiment field and multi-view image acquisition system

The experimental field was located at IFAS Plant Science Research and Education Unit (PSREU) in Citra, FL, USA (Figure 1). The field (29.408593° N, - 82.164398° W) grows blueberries with various genotypes for blueberry breeding. We collected the image data with our platform on the top second and third row. The distance between rows is 3 meters, and distance between the blueberry crops is 0.5 meter. For each row, blueberry plants are various in height from 0.4 to 2.0 miters.

Figure 1: Blueberry field (Satellite photo from Pix4Dcapture software).

Figure 2 shows our image acquisition systems are amounted on a mobile platform we developed in previous studies [26]. The platform configured four-steering and four-driving wheel, which is controlled with Robot Operation System (ROS). It can navigate autonomously with the dual RTK-GPS and also can be controlled with remote controller. In order to cover the

blueberry canopy and minimize occlusion as possible, the multi-view imaging system is customized to capture the blueberry image from top, left-side and right side. There are three different types of camera system and 9 cameras in total:

1) RGB system with three Panasonic Lumix G7 mirrorless cameras (Panasonic, Osaka, Japanese) which is used to collect 4592 x 3448 resolution color photographs or record the videos with the resolution of 1920 x 1080.
2) RGB-GPS system with three Raspberry Pi high quality 3MP cameras (Raspberry Pi Foundation, Cambridge, United Kingdom), which can communicate with mobile platform thorough the communication mechanism of ROS. Each Pi camera is controlled with a single Raspberry Pi separately and can publish the image topic with resolution of 1920 x 1080 and 30 FPS to the ROS MASTER. When the robot moves along the blueberry row, the robot states, especially the RTK-GPS data and Pi camera data will be synchronously recorded as ROSBAG.
3) RGBD-GPS system with RealSense D435i camera (Intel RealSense, California, United States), which is used to record RGB (1280 x 720, 15 FPS) and aligned depth image (1280 x 720, 15 FPS), as well as the RTK-GPS data from mobile platform. An external laptop is used to save all the data and communication with mobile platform through wired networks.

During data collection period, we collected all the data from the three camera systems with the robot speed of 0.2~0.4 m/s. In this paper, we build a blueberry dataset with the RGB images (three different resolutions) from the three cameras.

Figure 2. Multi-view imaging system is integrated with a ground robotic platform. The imaging system possesses three different types of camera system (RGB, RGB-GPS, RGBD-GPS) and 9 cameras in total. Each camera system can capture the blueberry image from top, left-side and right-side to minimize the occlusion by leaves and trunks.

2.2 Data preparation

Our Field-Blueberry dataset consists of three sections: HANDset, OPENset, and ROBOset. HANDset includes the RGB images captured manually with digital camera and cell phone in other blueberry field recent years ago or current season; OPENset is an open-source labeled dataset collected from Roboflow (https://roboflow.com/) and we selected high-quality images and modified the corresponding annotations; ROBOset includes RGB images captured from the multi-view acquisition system. Table 1 shows the datasets and description:

Table 1: Composition of our Field-Blueberry Dataset

Subdataset	# of images	# of annotated images	Resolution
HANDset	19460	215	Includes the RGB image randomly captured by digital camera or phones from 2018 to 2023. The resolution mainly contains 4896 x 3624, 4032 x 3024, 1920 x 1080, 1280 x 720.
OPENset	92	92	Collect from Roboflow and modified the annotations with lower resolution. Most of images focus on clusters rather than whole plants.
ROBOset	4742	50	Includes RGB images captured by our multi-view image acquisition system; The resolution includes 4592 x 3448, 1920 x 1080, 1280 x 720; All of the images focus on the individual blueberry plants.

Figure 3 (a) shows the histogram of blueberry count by image. Among the annotated images, most of images contains over 100 blueberries, which is ranging from 45 to more than 3000. Figure 3 (b)-(f) show several typical examples of our Field-Berry dataset. Figure 3b is an example of HANDset captured with digital camera, which possesses about 600 blueberries. We selected images that have good lighting and clear views as the annotation cases. Figure 4c shows the example from OPENset data, which has 202 blueberries in the image. OPENset only focus on blueberry clusters thus the ratios between the pixels of single blueberries and the total pixels of whole image are the highest among the three datasets. It makes the blueberry can be easier to be detected and observed. Figure 3 (d)-(f) are examples from ROBOset, which captured images from left, top and right view for one plant. Detected individual blueberry fruits in these images are more challenging due to its small size, dense clusters, occlusions, shadows, motion blur, and variability in shape and color. These factors can make it difficult for computer vision algorithms to accurately identify and locate blueberries in images.

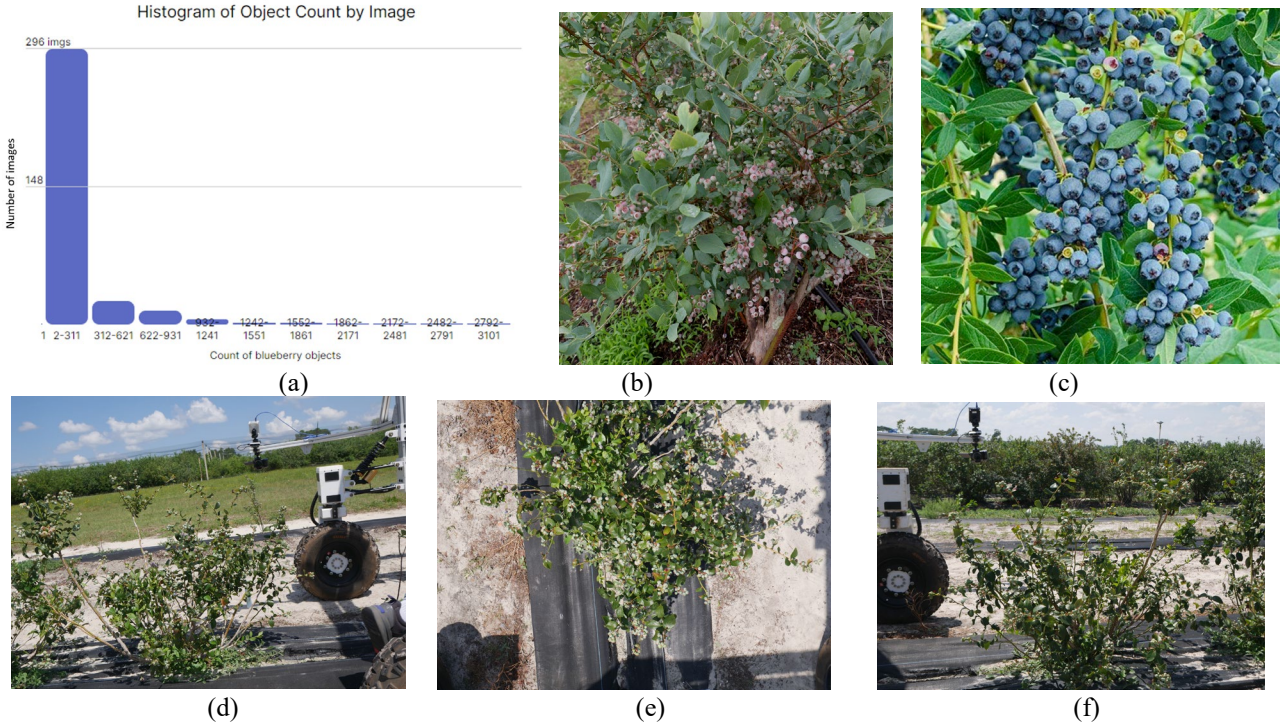


Figure 3. Example of Field-Blueberry dataset: (a) is the histogram of Blueberry Count by Images; (b) is from HANDset with 600 blueberries (labels) in the image; (c) is from OPENset with 202 blueberries (labels) in the image; (d)-(f) are from ROBOset, which captured images from left, top and right view for one plant.

Annotating each individual blueberry in the image is pretty time-consuming and labor-extensive. In this case, a browser-based interactive annotation tool powered by Roboflow (https://roboflow.com/) is used to realize model-assisted annotation to improve the annotation efficiency. Firstly, we trained the YOLOv8 model based on the OPENset images (collected form Roboflow Universal) and few images annotated by ourselves. The first trained YOLOv8 is deployed to realize the model-assisted annotation through setting the confidence and overlap of the predicted bounding box. As more annotations were collected, we retrained the YOLOv8 model again as the annotation assistant. We repeated the process three times, and the annotation assistant based on YOLOv8 finally can detect 40%~70% bounding boxes as the pre-annotation.

In a field situation where illumination changes significantly due to different weather conditions, it can be challenging to collect high-quality data with minimal motion blur with mobile platform. Setting the brightness levels to a range of -15% to +15% and a blur of up to 0.5 as the augmentation approach can be effective to provide a reasonable level of normalization. Considering the blueberry detection belongs to small and dense object task, we also use two augmentation approaches on blueberry dataset: mosaic[27] and cropping[28]. Mosaic technique can be used to stitch together multiple images of blueberries, which can increase the overall resolution of the image and provide a more detailed view of the individual blueberries. Cropping can be used to focus on individual blueberries or clusters of blueberries, allowing for more accurate analysis of these small and dense objects.

2.3 YOLOV8 detector

In 2015, the debut of YOLO [18], or You Only Look Once, shook the world of computer vision as its system was capable of real-time object detection with astounding accuracy and speed. Since then, YOLO has undergone several iterations of improvements in prediction accuracy and efficiency, eventually culminating in the release of its latest family member: YOLOv8 [19].

YOLOv8 was released in January 2023 by Ultralytics (https://github.com/ultralytics/ultralytics), the company that developed YOLOv5. It achieved the state-of-the-art performance, featuring an object detection model for P5 640 and P6 1280 resolutions, as well as a YOLACT-based instance segmentation model [29]. The model also includes different size options with N/S/M/L/X scales, similar to YOLOv5, to cater to various scenarios. Besides, more range of vision AI tasks, including pose estimation, tracking, and classification, are added and still updated.

Since, at the time of this writing, there was no paper about YOLOv8, we need insights into the architectural decisions compared to the rest YOLO versions. Considering YOLOv8 and YOLOv5 are from the same research group with the similar programming style, the difference between the two networks models is emphasized during the comparation.

1) **Model structure design**: The most significant changes in the model lay in the head module. The head module has been changed from the original coupling structure to the decoupling one, and its style has been changed from YOLOv5's Anchor-Based to Anchor-Free. The removal of the objectness branch and the retention of only the decoupled classification and regression branches stand as the major differences. Additionally, the regression branch now employs integral form representation as proposed in the Distribution Focal Loss [30]. Another important change is that the backbone network and neck module are based on the YOLOv7 ELAN design concept [31], replacing the C3 module of YOLOv5 with the C2f module. C2f module possesses more skip connection and additional split operations with higher abilities in feature extraction. Here are other changes different from YOLOv5: the kernel of the first convolutional layer has been changed from 6x6 to 3x3; Removed 2 convolutional connection layers from neck module; The block number has been changed from 3–6–9–3 to 3–6–6–3.

2) **Loss calculation** consists of 2 parts: the classification and regression, without the objectness loss in the YOLOv5. The classification branch still uses BCE loss and the regression branch employs both Distribution Focal Loss and CIoU Loss. The 3 Losses are weighted by a specific weight ratio. Given the superiority of dynamic assignment strategies, the YOLOv8 algorithm directly incorporates the one employed in TOOD's TaskAlignedAssigner [32].

3) **Training strategy**: YOLOv8's data augmentation is similar to YOLOv5, whereas it stops the Mosaic augmentation in the final 10 epochs as proposed in YOLOX [33]. Except mosaic, other augmentations are also used during training process, including random affine, random perspective, CopyPaste, Letterbox, center Cropping, mix-up, albumentations, Random HSV, and Random Flip.

4) **Model inference process**: The inference process of YOLOv8 is almost the same as YOLOv5. The only difference is that the integral representation bounding box in Distribution Focal Loss needs to be decoded into a regular 4-dimensional bounding box, and the subsequent calculation process is the same as YOLOv5.

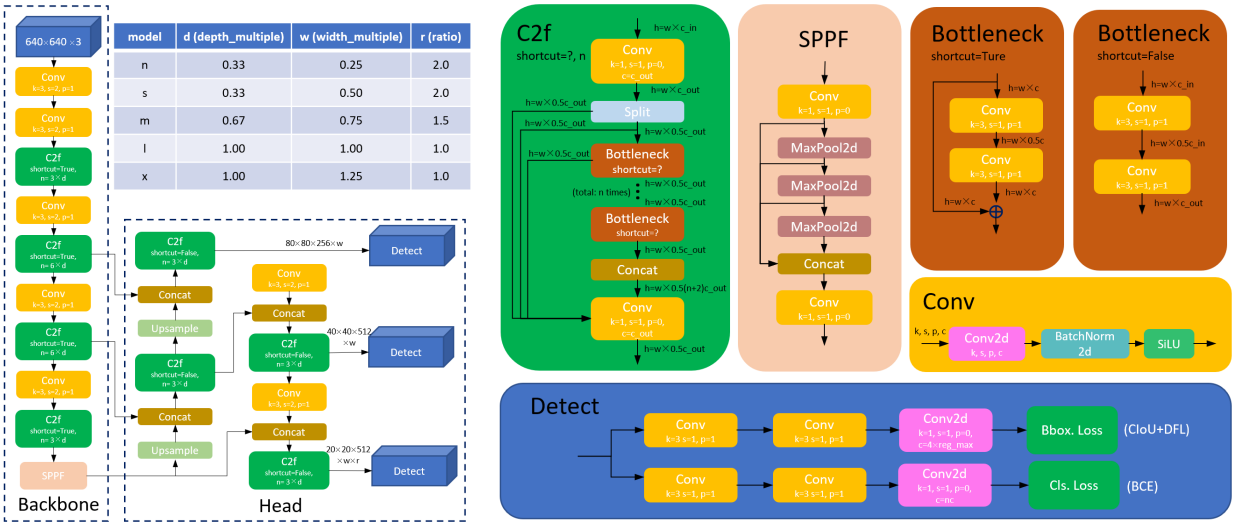


Figure 4. The architecture of YOLOv8 detector. (a) The network consists of two main parts: backbone and head. The backbone of the YOLOv8 model is a CSPDarknet53 feature extractor, which is followed by a novel C2f module instead of the traditional YOLO neck architecture. The Head module has been updated to the current mainstream decoupled structure, separating the classification and detection heads, and switching from Anchor-Based to Anchor-Free. (b) the architecture of the modules used in YOLOv8, including C2f, SPPF, Bottleneck, conv and detect module.

2.4 Blueberry yield estimation with multi-view Imagery

Figure 5 shows the pipeline of blueberry yield estimation for an individual blueberry plant using a multi-view imaging acquisition system. The pipeline starts with the three views' images of a blueberry plant (plot or row) being cropped manually

to create three cropped images, which based on time stamps and spatial information of the blueberry field. The cropped images are then inputted into a YOLOv8 detector. In this case, the YOLOv8 detector is trained and fine-tuned with our datasets (describe in 2.2) and then used to detect blueberries within the three cropped images. The YOLOv8 detector is then used to detect the blueberries within the three images, and the number of detected blueberries from the three views are used to predict the blueberry yield with a regression model. The locations of the detected bounding boxes are also used to generate a density map that contains the plant-level spatial distribution information of the blueberry yield.

There are three main components within the whole pipeline includes YOLO detector module, multi-view regression, and density map generation module.

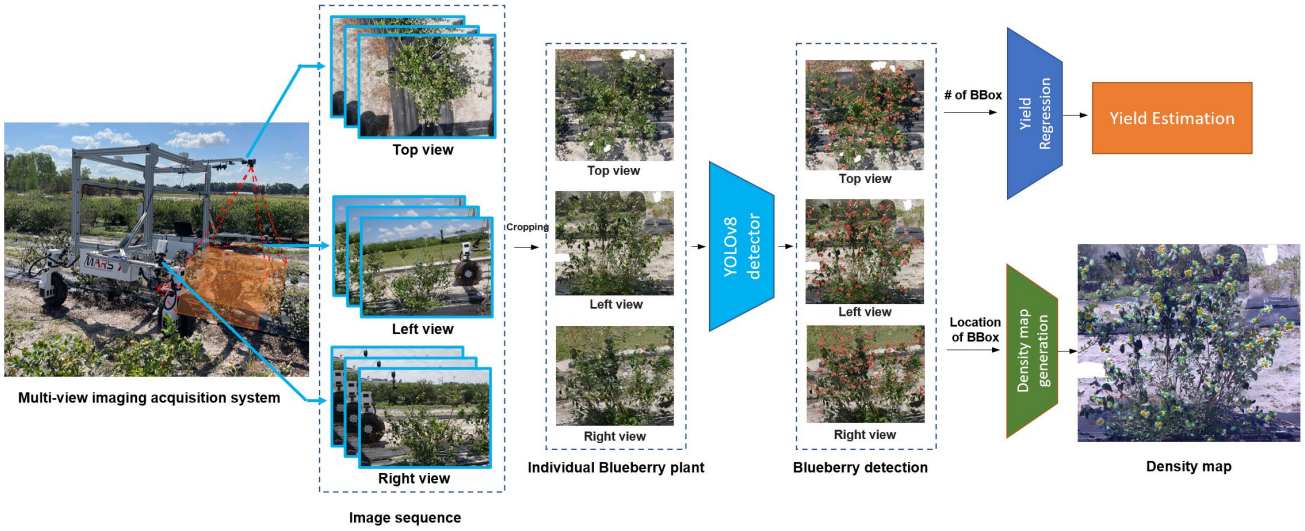


Figure 5. Illustration of pre-processing of yield estimation for individual plant. The three image sequences captured with our multi-view imaging acquisition system are used to generate the corresponding cropping image of three-views individual blueberry plant. The cropping processing relies on the time stamp and spatial information of blueberry field. Then, the YOLOv8 detector is used to detect the blueberries within the three images. The numbers of detected blueberries from the three views are used to predict the blueberry yield with regression model. Meanwhile, the locations of detected bounding boxes are used to generate the density map, which contains the plant-level spatial distribution information of the blueberry yield.

YOLO detector module: YOLO detector is used to generate the predicted bounding box of blueberries, and the detailed model can be seen in 2.3 section. Considering our pipeline is offline with less constrains of inference speed, the highest accuracy model YOLOv8x was selected for the detection task. We compared the YOLOv8 with different configurations, as well as YOLOv5, and finally evaluate the model performance.

Multi-view regression module: The number of detected blueberries from the three views is then used to predict the blueberry yield using a regression model. The regression model takes the number of YOLO-based predicted blueberries as inputs, and outputs an estimate of the blueberry yield. This yield estimate can then be used to guide harvest management decisions or to plan future production. So far, we directly use the linear regression model owing to the limited sampling pairs. In the future, we compared several typical regression models including linear, non-linear, and Multiple layer perception.

Density map generation module: In addition to yield prediction, the density map of blueberry images is generated based on the detected bounding boxes after YOLOv8 detector. We followed the similar idea as proposed by Zhang et.al[34] and Li et.al [35] for the objects counting of small and dense objectives. Firstly, a white grayscale image with the same size of input image is generated and all the predicted bounding boxes are drawn on the image with the reverse color. Then, gaussian filters with geometry-fixed kernel were used to blur each object prediction to generate predicted density maps (grayscale). The size of kernel was determined based on the exploratory data analysis is performed on the training set to analyze the average scale for blueberries. Finally, the grayscale of density maps was converted to fake color images and blended with original RGB image with a specific weight. The density map can provide the proposed attention region of blueberries for farmers or researchers.

3. Experiments and Results

3.1 multi-view imaging system

The field mapping requires a reliable and efficient data collection system to capture high-quality image sequences with high-resolution and less noise (e.g., blur). In this study, a multi-view camera system was used to capture images in the blueberry field with consumer-grade cameras. Specifically, the highest-resolution images (4592 x 3448) captured with

Panasonic Lumix G7 cameras.

Due to the complexity and variability of the blueberry plants and the environmental conditions, automatic alignment among three views is pretty challenging with current open algorithm, e.g., Oriented FAST and Rotated Brief (ORB) algorithm. Thus, automatic cropping pipeline is pretty difficult to achieve the satisfied result, and finally we cropped the images of individual plants with artificial supplementation.

Figure 6 illuminates the feature matching among different views with Oriented FAST and Rotated Brief (ORB) algorithm. ORB is basically a fusion of FAST key-point detector and BRIEF descriptor with many modifications to enhance the performance, which is almost two orders of magnitude faster than SIFT and SURF. Fig 6 (a) and (b) are the right-top views matching and top-left views matching, respectively. Obviously, there are many mismatching pairs of features points, for example, the feature points of the camera extension or robot's wheel module mismatch to the blueberry canopy. Even through fine-tune the parameters of distance filters, the image alignments among three cameras are still failed.

There are other potential approaches to generate the images contains individual plants, plots or partial row, for example, using image stitching software or algorithm to generate the panorama images. Several software, such as Microsoft Image Composite, Hugin and PTGui can be used to generate the panorama for the image sequences of one view and then align other panoramas to crop the desired region. One of concern is the software may generate "fake" pixel to fill certain region. Besides, the expert knowledge is required for the satisfied image stitching. In addition, current image stitching algorithms are more robust that benefit from Deep learning approaches. For instance, LoFTR [36], Detector-Free Local Feature Matching with Transformers, can produce high-quality matches even in indistinctive regions with low-textures, motion blur, or repetitive patterns.

(a) (b)

Figure 6. Illustration of image matching with Oriented FAST and Rotated Brief (ORB) algorithm. Actually, there are many mismatching pairs of features points obviously. For example, the feature points on the robot mistakenly match to the points of blueberry plant. (a) matching result between right view and top view; (b) matching result between top view and left view.

3.2 YOLOv8 Detector Performance

We trained YOLOv8 model with the Field-Blueberry dataset (consists of HANDset, OPENset, and ROBOset) on Ultralytics framework. The original dataset was augmented six times applying different augmentations such as brightness adjustment (-15% and +15%), blurring (Up to 0.5px), bounding box adjustment (brightness: 15% and +15%), crop (0~20%), and mosaic. Augmented dataset was divided into training and valid dataset with the ratio of 8:2. In addition, we annotated more 39 images collected with our multi-view imaging platform on 4/12/2023 as the test dataset to evaluate the performance of model.

Considering the blueberry fruits are pretty dense and small in the image, we compared the performance influence with different inputting image size and different scale model. Performance of YOLOv8 model was tested on 39 test set images, including Precision, Recall, F1 score, and AP, as shown in table 2. The size of the model has a significant impact on the detection performance. Generally, larger models can provide better detection results. However, the computation required by larger models is more significant, which can lower the FPS. YOLOv8x with the 1280 image input shows the best detection performance with the mAP@0.5 of 77.3% and the FPS is pretty low (4.3 FPS) on Tesla T4 GPU inference. The performance from larger model to small model decrease slightly but the inference speed can improve significantly. For example, compared with the largest model, the smallest YOLOv8n (with the size of 1280) only decrease 2.2% of mAP@0.5 but only fourteenth of inference speed averagely were used.

It was observed that the input resolution of the images significantly affects the performance of the object detection models. The YOLOv8 group with 1280 input achieved higher performance (~10-17%) than the group with 640 input. But the difference gradually decreased with the larger model. This difference could be attributed to the fact that the objects in our dataset were small and densely packed, and high resolution was necessary to capture their important features. However, high resolution also increases computation time and decreases FPS, which should be taken into account.

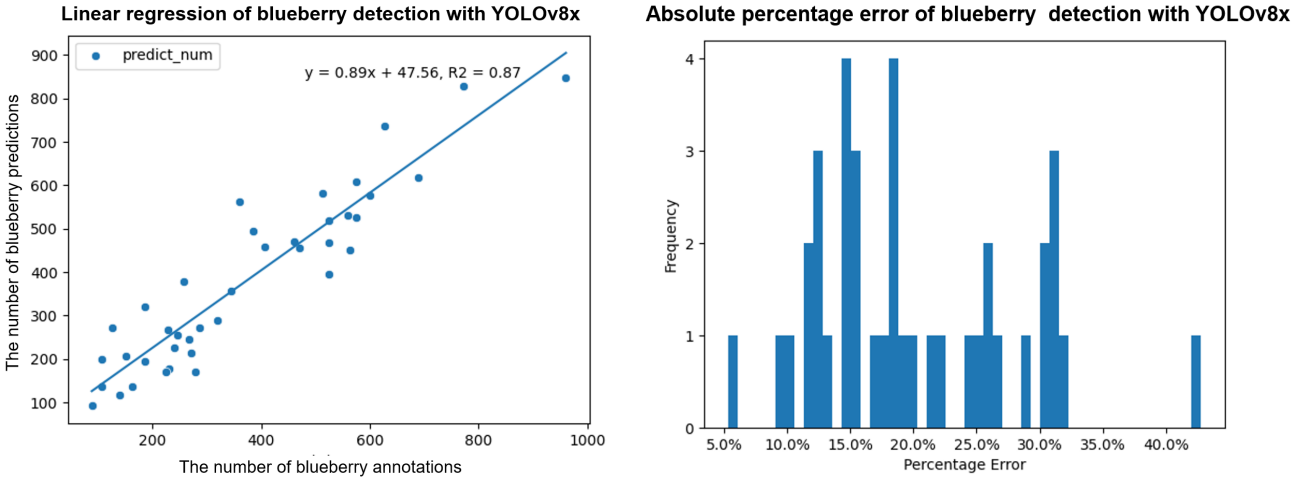
Besides, the yolov5x models were compared with yolov8x model. The result shows yolov8 performs better than its previous version yolov5 on our dataset. Especially with the small inputting image size (640), the mAP@0.5 increase 7% with smaller model parameters.

Table 2: Performance comparation of YOLOv8 with different configuration

YOLO model	Model size	size(pixels)	Precision	Recall	mAP@0.5	mAP@0.5-0.95	Params (M)	FLOPs (B)	Speed Tesla T4 (ms)
YOLOV5	YOLOv5x	640	78.9	46.5	62.0	29.5	86.7	205.7	46.1
		1280	87.7	62.7	76.6	47.5			
YOLOv8	YOLOv8n	640	69.2	45.6	58.4	30.5	3.2	8.7	10.6
		1280	86.1	61.2	75.1	46.8			16.0
	YOLOv8s	640	78.6	50.5	65.7	35.9	11.2	28.6	11.7
		1280	86.9	63.1	76.5	48.7			34.4
	YOLOv8m	640	79.5	53.2	68.0	38.2	25.9	78.9	23.9
		1280	87.1	63.0	76.6	48.8			82.5
	YOLOv8l	640	80.9	54.5	68.9	39.2	43.7	165.2	37.7
		1280	87.8	62.7	76.6	48.5			125.0
	YOLOv8x	640	81.0	54.1	69.0	39.3	68.2	257.8	58.5
		1280	**88.3**	**63.3**	**77.3**	**49.0**			232.3

Based on the results of our analysis, it is clear that the choice of object detection model depends on the specific application requirements. For real-time applications where the interface speed and FPS are critical factors, we recommend using YOLOv8n or YOLOv8s with the image size of 1280. These models have smaller sizes and lower computation requirements, which allows for faster detection speed and higher FPS. For offline applications where detection speed is not a critical factor, we recommend using YOLOv8x. This model has a larger size and higher computation requirements, which may lead to lower FPS. However, it can provide more accurate detection results, especially when working with high-resolution images or small and dense detection tasks.

In our experiment, the YOLOv8x with highest accuracy was used to detect the blueberries. Figure 7 (a) shows the linear regressions between the number of predicted blueberries (both true positives and false positives) and the annotated blueberries. It can be used to assess the accuracy of the YOLOv8 model in detecting blueberries, as a higher correlation between the predicted and annotated blueberries indicates better performance. The high R-squared values of 0.87 for the regressions indicate that the YOLOv8 model performs well in detecting blueberries accurately. The positive slope of the regression lines suggests that the model tends to underestimate the number of blueberries. Figure 7 (b) shows the mean absolute percentage error (MAPE) of blueberry detections on a field dataset. This metric can be used to evaluate the overall accuracy of the YOLOv8 model in detecting blueberries. The MAPE among most of the detection result is lower than 25% and focus on lower than 20%. Overall, considering the testing images includes the whole plants with pretty dense and small blueberries, the results suggest that the YOLOv8 model is a promising approach for detecting blueberries in this application.



(a) (b)

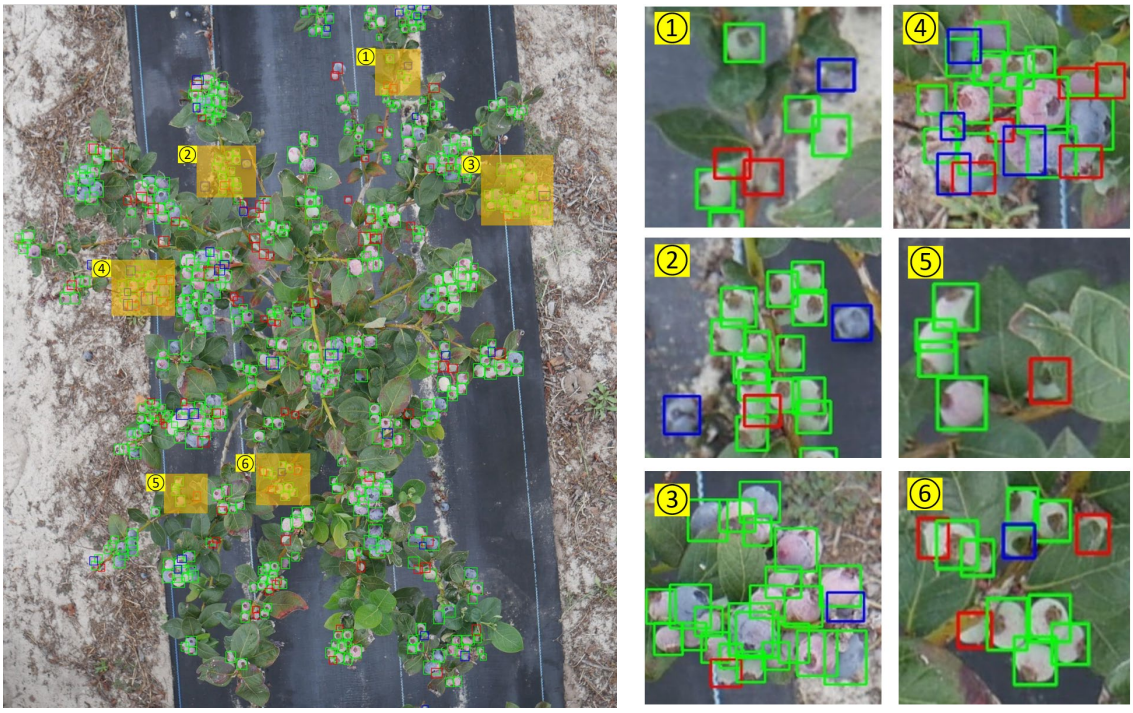
Figure 7. Performance evaluation of YOLOv8 model on test dataset. (a) Linear regressions between the number of predicted blueberries and annotated blueberries (TP+FP); (b) The absolute percentage error of blueberry detections on 39 testing images.

Based on the visualization in Figure 8, it shows that the YOLOv8x detector was able to detect majority of the blueberries in the field dataset, with green boxes representing true positive predictions. However, there were also instances of false

positive (blue boxes) and false negative (red boxes) predictions, indicating errors and missing detections. Figure 8 (a) shows a case of YOLOv8x detector result and the green regions are zoom-in in the figure 8 (b). Overall, 85.67% blueberries in this image were predicted (green) correctly, but still exits 15.58% of mistaken detections (blue) and 6.29% of missing detections (red).

One common issue observed in the failed detections was the presence of clustered objects that overlapped in the image, making it difficult for the detector to distinguish between them. In such cases, the detector may mistakenly detect part of one object as part of another or miss some objects entirely. Another issue that contributed to failed detections was the presence of shadows or low contrast areas, which made it difficult for the detector to accurately identify the object boundaries. This can lead to false negative predictions, where the object is missed entirely. In addition, these blueberries dropped from the plants are not regarded as the detection objectives. But actually, the detector can detect all' the blueberries wherever they are in the plant or dropped from plants. When dealing with side views of objects, especially in scenarios where rows or clusters of objects are present, the influence of neighboring rows can indeed lead to increased mistaken detections. This occurs because the overlapping or occluding objects in the neighboring row can affect the visibility and appearance of the objects in the target row, making it challenging for the detector to accurately identify and distinguish them.

To improve detection accuracy in such cases, it may be helpful to preprocess the images to enhance contrast and separate clustered objects before feeding them into the detector. Additionally, fine-tuning the detector on a larger and more diverse dataset may improve its ability to distinguish between overlapping objects and handle low-contrast areas. Another important aspect is to set a canopy for the mobile platform for more stable imaging environment with less illumination variations. It also can isolate the influence of neighboring rows, which benefit the detection of side views.



(a) (b)

Figure 8. Visualization of detection result using YOLOv8x detector on field dataset. The green, blue, red boxes denote true positive (TP), false positive (FP), and false negative (FN) predictions, respectively. (a) visualization of one predicted top-view image, in which the number of annotated, predicted (TP+FP), correct (TP), error (FP) and missing detections (FN) are 572, 526, 490, 82 and 36, respectively. (b) are zoom-in regions of the corresponding yellow regions in the left figure, containing some cases of failed detections.

3.3 Comparation of single-view detections and multi-view regression

To evaluate the performance of single-view and multi-view regression in blueberry detection, we conducted a manual harvesting process on 12 blueberry plants with different genotypes. During the harvest, we counted all potential blueberries, regardless of their ripeness. Subsequently, for each harvested plant, we carefully selected the corresponding image frames and manually cropped them to focus solely on one plant. These individual plant images were then annotated with bounding boxes to establish the ground truth for the annotations. We utilized the YOLOv8x model to predict the number of blueberries in the cropped images. In essence, for each harvested blueberry plant, we generated data pairs consisting of the manually counted blueberry numbers, the annotated numbers from three different views, and the predicted numbers from those same three views. This approach allowed us to collect comprehensive data for further analysis and comparison.

Figure 9(a) Compared the visibility of blueberries in three different views (top-view, left-view, and right-view), the visibility is measured by the ratio of annotated bounding boxes in the single-view image to manual counting. It appears that the top-view provides the highest coverage, with an average ratio of 52.7%, indicating that it captures more blueberries compared to the other views. The left-view and right-view have slightly lower visibility, with ratios of 47.3% and 38.1%, respectively. This suggests that the top-view is more effective in detecting blueberries.

However, even for the highest-coverage top view, only half of the blueberries can be seen. Single-view is pretty limited in its ability to capture the entire blueberry plant and may result in partial coverage, leading to a lower visibility ratio compared to the multi-view approach. Single-view images can only capture a portion of the plant, which makes it challenging to accurately detect and count all the blueberries. In contrast, by utilizing the information from multiple views, we can overcome these limitations and improve the overall accuracy and reliability of the detection system.

The integration of multi-view information in blueberry detection systems brings about substantial improvements in accuracy and robustness. By combining data from different views, these systems can effectively enhance coverage and mitigate the challenges posed by occlusions. One of the challenges is how to fuse the multi-view information with the overlap among cameras. It noted that there is not enough overlap among the three cameras in our imaging system (section 3.1), which means the incorporation of spatial information derived from multiple views is not doable technically. Another approach is to build the potential relationship among the real number of blueberries and the visible blueberries in the three-views. But the regression relationships only workable for the same imaging system (e.g., camera position and resolution) and it cannot conduct a general conclusion.

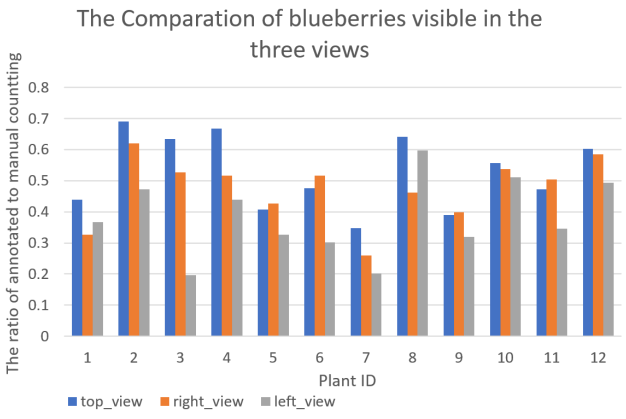
Considering the accessibility of technical route and the number of sampling pairs, we directly used the multiple linear regression approach to build the relationship among the predicted blueberry number of individual plants and the yolov8-based predicted blueberry number of three views, with the equation (1):

$$y_{multi-view} = a_0 + a_1 x_{top-view} + a_2 x_{left-view} + a_3 x_{right-view} \quad (1)$$

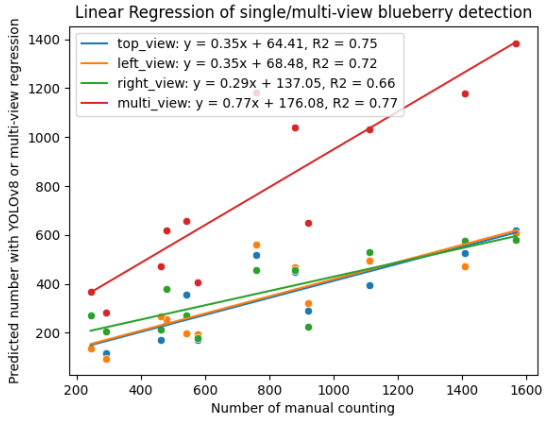
Where $y_{multi-view}$ is the predicted blueberry number of individual plants (with manual counting) and $x_{top-view}$, $x_{left-view}$, $x_{right-view}$ are the predicted blueberry number of three views (predicted numbers with YOLOv8x). a_1, a_2, a_3 are the weight of each view, a_0 is the intercept. As for the result, $a_0 = -6.47$, $a_1 = 1.11$, $a_2 = 0.73$, $a_3 = 0.45$ with the R2=0.772. The weight represents the importance of components, which keep the consistence with the blueberry visible of three views.

Figure 9(b) conducted linear regression analysis to estimate the blueberry detection using either a single-view or multi-view approach. The results indicate that the multi-view regression approach achieves the highest R2 value of 0.77, indicating a reasonably good correlation between the predicted blueberry numbers from the multi-view approach and the manual counting. This suggests that incorporating information from multiple views can improve the accuracy of blueberry detection compared to relying on a single view alone. Among the single-view approach, the top view achieves the strongest correlation with the R2 value of 0.75 that only slightly decrease than multi-view approach.

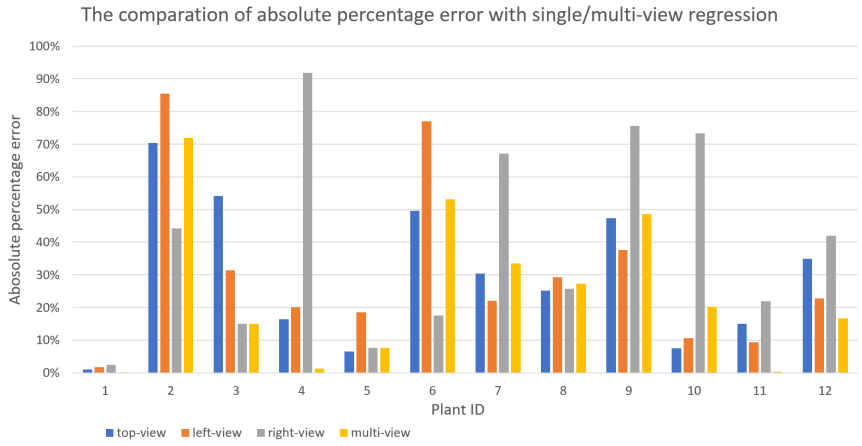
Figure 9(c) illustrated comparison of absolute percentage error among the four regression approaches revealed into the accuracy of estimating blueberry numbers. Among the 12 experimental plants, the multi-view approach emerged as the most precise, exhibiting a mean absolute percentage error of 24.63%. This result highlights the effectiveness of utilizing multiple views in enhancing the estimation accuracy. Following closely behind, the top-view approach demonstrated a mean absolute percentage error of 29.86%, signifying its reliable performance in estimating blueberry quantities. Similarly, the left-view approach achieved a mean absolute percentage error of 30.5%, indicating its overall competence, although slightly higher than the top-view approach. Overall, these findings emphasize the advantages of incorporating multi-view perspectives for blueberry detection, showcasing the superior accuracy achieved by the multi-view regression approach in estimating blueberry numbers.



(a)



(b)



(c)

Figure 9. Regression result of blueberry detection with YOLOv8. (a) comparation of blueberries visible in the three views. The visibility is defined by the ratio of annotated number of bounding boxes in the single-view image to the manual counting. Single-view images only can cover partial plants, making the ratio < 1. Among the three views, top-view can cover most region with the average ratio of 52.7%, and the left-view and right-view are 47.3 and 38.1%, respectively. (b) Linear regression of single or multi-view blueberry detection estimation. The multi-view regression approach achieves the highest R^2 (0.77) with the blueberry number of manual counting. (c) Comparation of absolute percentage error among 4 regression approach on the 12 experimental plants. The multi-view approach shows the lowest mean absolute percentage error of 24.63% among the 12 plants, followed by top-view with 29.86% and left-view with 30.5%.

3.4 Density map generation

With the YOLOv8x detector, the predicted blueberries can be used for generating the density map. Figure 10 shows a case of density map of individual plants from three views. The darker color shows higher blueberry density. The density map generated by the proposed system can be used to provide the proposed attention region of blueberries for farmers or researchers. This information can be used for yield prediction, quality assessment, and to optimize farming practices such as pruning, fertilization, and irrigation. Additionally, the density map can help farmers and researchers to identify areas with low blueberry densities, which may indicate disease or pest problems or suboptimal growing conditions. By using the density map, farmers and researchers can take appropriate actions to address these issues and improve the overall yield and quality of the blueberry crop.

(a) **(b)** **(c)**

Figure 10. Illustration of blueberry density map from three views: (a) left-view, (b) top-view, (c) right-view. The regions with darker color shows the higher blueberry density.

4. Conclusions

In conclusion, the proposed machine vision approach for blueberry yield estimation using multi-view regression with the YOLOv8 detector shows promising results. The multi-view Imagery approach achieves promising results, with a mean absolute percentage error of 24.6% and R2 of 0.77, which is 5.2%~15.7% higher than single-views approaches. This method offers a more accurate and efficient way of estimating blueberry yields compared to traditional methods, which rely on expert knowledge or indirect regression analysis. The use of mobile robots, orthomosaics and panorama technology can also facilitate the estimation of the yield of the whole blueberry field. The density maps generated from the predicted bounding

box of blueberries provide additional information for further phenotyping analysis. This approach has potential to optimize crop management practices and increase profitability in the agricultural production of blueberries.

Acknowledgements

This study is supported by the University of Florida IFAS LIFT AI seed grant.

Reference

[1] K. L. Morgan, "Market Trends for US Berry Crops: Implications for Florida Blueberry, Blackberry, and Raspberry Producers: FE1123/FE1123, 11/2022," *EDIS,* vol. 2022, no. 6, 2022.

[2] L. He *et al.*, "Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods," *Computers and Electronics in Agriculture,* vol. 195, p. 106812, 2022.

[3] G. Niedbała, J. Kurek, B. Świderski, T. Wojciechowski, I. Antoniuk, and K. Bobran, "Prediction of Blueberry (Vaccinium corymbosum L.) Yield Based on Artificial Intelligence Methods," *Agriculture,* vol. 12, no. 12, p. 2089, 2022.

[4] J. P. Vasconez, J. Delpiano, S. Vougioukas, and F. A. Cheein, "Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation," *Computers and Electronics in Agriculture,* vol. 173, p. 105348, 2020.

[5] H. Williams *et al.*, "Improvements to and large-scale evaluation of a robotic kiwifruit harvester," (in English), *Journal of Field Robotics,* Article vol. 37, no. 2, pp. 187-201, Mar 2020, doi: 10.1002/rob.21890.

[6] S. Sun *et al.*, "Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 160, pp. 195-207, 2020.

[7] S. Gutierrez, A. Wendel, and J. Underwood, "Ground based hyperspectral imaging for extensive mango yield estimation," (in English), *Computers and Electronics in Agriculture,* Article vol. 157, pp. 126-135, Feb 2019, doi: 10.1016/j.compag.2018.12.041.

[8] Z. Zheng *et al.*, "An efficient online citrus counting system for large-scale unstructured orchards based on the unmanned aerial vehicle," *Journal of Field Robotics,* vol. n/a, no. n/a, 2022, doi: https://doi.org/10.1002/rob.22147.

[9] S. W. Chen *et al.*, "Counting Apples and Oranges With Deep Learning: A Data-Driven Approach," (in English), *Ieee Robotics and Automation Letters,* Article vol. 2, no. 2, pp. 781-788, Apr 2017, doi: 10.1109/lra.2017.2651944.

[10] K. C. Swain, Q. U. Zaman, A. W. Schumann, D. C. Percival, and D. D. Bochtis, "Computer vision system for wild blueberry fruit yield mapping," *Biosystems Engineering,* vol. 106, no. 4, pp. 389-394, 2010.

[11] X. Bai *et al.*, "Comparison of machine-learning and casa models for predicting apple fruit yields from time-series planet imageries," *Remote Sensing,* vol. 13, no. 16, p. 3073, 2021.

[12] J. Van Beek *et al.*, "Temporal dependency of yield and quality estimation through spectral vegetation indices in pear orchards," *Remote Sensing,* vol. 7, no. 8, pp. 9886-9903, 2015.

[13] S. Salvo *et al.*, "An estimate of potential blueberry yield using regression models that relate the number of fruits to the number of flower buds and to climatic variables," *Sci. Hortic.,* vol. 133, pp. 56-63, 2012.

[14] A. Cravero, S. Pardo, S. Sepúlveda, and L. Muñoz, "Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review," *Agronomy,* vol. 12, no. 3, p. 748, 2022.

[15] P. Maheswari, P. Raja, O. E. Apolo-Apolo, and M. Pérez-Ruiz, "Intelligent fruit yield estimation for orchards using deep learning based semantic segmentation techniques—a review," *Frontiers in Plant Science,* vol. 12, p. 684328, 2021.

[16] E. Y. Obsie, H. Qu, and F. Drummond, "Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms," *Computers and Electronics in Agriculture,* vol. 178, p. 105778, 2020.

[17] H. R. Seireg, Y. M. Omar, F. E. Abd El-Samie, A. S. El-Fishawy, and A. Elmahalawy, "Ensemble machine learning techniques using computer simulation data for wild blueberry yield prediction," *IEEE Access,* vol. 10, pp. 64671-64687, 2022.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[19] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond," *arXiv preprint arXiv:2304.00501,* 2023.

[20] A. W. Schumann, N. S. Mood, P. D. Mungofa, C. MacEachern, Q. Zaman, and T. Esau, "Detection of three fruit maturity stages in wild blueberry fields using deep learning artificial neural networks," in *2019 ASABE Annual International Meeting*, 2019: American Society of Agricultural and Biological Engineers, p. 1.

[21] C. B. MacEachern, T. J. Esau, A. W. Schumann, P. J. Hennessy, and Q. U. Zaman, "Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks," *Smart Agricultural Technology,* vol. 3, p. 100099, 2023.

[22] W. J. Yang, X. X. Ma, W. C. Hu, and P. J. Tang, "Lightweight Blueberry Fruit Recognition Based on Multi-Scale and Attention Fusion NCBAM," (in English), *Agronomy-Basel,* Article vol. 12, no. 10, p. 13, Oct 2022, Art no. 2354, doi: 10.3390/agronomy12102354.

[23] S. Gonzalez, C. Arellano, and J. E. Tapia, "Deepblueberry: Quantification of Blueberries in the Wild Using Instance Segmentation," (in English), *Ieee Access,* Article vol. 7, pp. 105776-105788, 2019, doi: 10.1109/access.2019.2933062.

[24] X. P. Ni, C. Y. Li, H. Y. Jiang, and F. Takeda, "Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield," (in English), *Hortic. Res.-England,* Article vol. 7, no. 1, p. 14, Jul 2020, doi: 10.1038/s41438-020-0323-3.

[25] X. P. Ni, C. Y. Li, H. Y. Jiang, and F. Takeda, "Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits," (in English), *Isprs Journal of Photogrammetry and Remote Sensing,* Article vol. 171, pp. 297-309, Jan 2021, doi: 10.1016/j.isprsjprs.2020.11.010.

[26] R. Xu and C. Y. Li, "A modular agricultural robotic system (MARS) for precision farming: Concept and implementation," (in English), *Journal of Field Robotics,* Article; Early Access p. 23, 2022, doi: 10.1002/rob.22056.

[27] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934,* 2020.

[28] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep CNNs," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 30, no. 9, pp. 2917-2931, 2019.

[29] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9157-9166.

[30] X. Li, W. Wang, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11632-11641.

[31] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696,* 2022.

[32] C. Feng, Y. Zhong, Y. Gao, M. R. Scott, and W. Huang, "Tood: Task-aligned one-stage object detection," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021: IEEE Computer Society, pp. 3490-3499.

[33] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430,* 2021.

[34] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 589-597.

[35] C. Li, T. Yang, S. Zhu, C. Chen, and S. Guan, "Density map guided object detection in aerial images," in *proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 190-191.

[36] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "LoFTR: Detector-free local feature matching with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8922-8931.