Boyd-Scott Graduate Research Award
OFFICIAL ENTRY FORM

STUDENT INFORMATION

Name	Rashmi Sahu	**ASABE Member #**	M1061191
Mailing Address		**Email Address**	
Research Paper Title	Apple Flower Bud Detection for Robotic Bud Thinning Using Deep Learning Neural Networks		
M.S. or Ph.D.	M.S.	**Expected Date of Graduation (month/year)**	08/12/2023

I hereby attest that the information I have provided in this entry form is true and I meet ALL eligibility requirements for the Graduate Research Award Competition. I have read and understood the rules for the competitions. The paper I am submitting is based on work completed in partial fulfillment of the requirements for the M.S. or Ph.D. degree in Biological/Agricultural Engineering or other closely related engineering graduate degree.

Student's Name *Rashmi Sahu* *Date 03/14/2023*

GRADUATE PROGRAM INFORMATION

Major Professor's Name	Long He	**Major Professor's Email Address**	luh378@psu.edu
Dept Head's Name	Suat Irmak	**Dept Head's Email Address**	sfi5068@psu.edu
Department Name	Department of Agricultural and Biological Engineering		
University Name	Pennsylvania State University		

MAJOR PROFESSOR AND DEPARTMENT HEAD ENDORSEMENTS

I attest that the student named above is a member of ASABE, was enrolled in a graduate program in our department for at least four months between March 15 of this year and March 15 of the previous year and the paper being submitted is based on research completed for either M.S. or Ph.D. degree.

Major Professor's Name Long He Long He *Date* 3/15/2023

Department Head's Name Suat Irmak SuatIrmak *Date* 3/15/2023

Submit an electronic copy of your paper and completed and Official Entry Form in a PDF file and email to the attention of the ASABE Awards Administrator, awards@asabe.org, by March 15.

APPLE FLOWER BUD DETECTION FOR ROBOTIC BUD THINNING USING DEEP LEARNING NEURAL NETWORKS

Rashmi Sahu and Long He

The authors are **Rashmi Sahu,** ASABE student member, Graduate Research Assistant, Department of Agricultural and Biological Engineering, Pennsylvania State University, University Park, PA, USA; **Long He**, Assistant professor, Penn State Fruit Research and Extension Center (FREC), Department of Agricultural and Biological Engineering, Pennsylvania State University, University Park, PA, USA; **Corresponding author:** Rashmi Sahu, 290 University drive, Penn State Fruit Research and Extension Center (FREC), Biglerville, PA, 17307; e-mail: rps6056@psu.edu

Highlights

- Algorithms were developed for apple flower bud detection using deep neural network.
- Bud detection performances were compared with YOLOv4, YOLOv5, and YOLOv7 models.
- Models were also tested with two datasets and two labeling methods to improve the generalizability of models.
- Results showed the YOLOv4 model outperformed the YOLOv5 and YOLOv7 models on bud detection accuracy.

Abstract. *Crop load management practices such as mechanical pruning (hedging), chemical thinning, and mechanical thinning are mostly followed methods for apple crop load management due to their effectiveness in both cost and efficiency. However, these methods are non-selective and can lead to unpredictable fruit numbers per tree, or even damage leaf tissue. Nevertheless, achieving accurate and optimal target fruit numbers per tree is a challenging task. Robotic bud thinning is an alternative technique of crop load management that regulates fruit bud density in the tree canopy. Real-time flower bud detection in the natural environment is a key step for developing this robotic system to automatically remove flower buds. This study proposed a real-time bud detection model using the You Only Look Once (YOLO) v4 deep learning algorithm. The detection performance of YOLOv4 model was compared with those of YOLOv5 and YOLOv7 models. The results showed that under the same conditions, YOLOv4 performance better than YOLOv5 and YOLOv7 for buds' detection. The mean average precision (mAPs) of bud detection with YOLOv4 were 98.99% on dataset-1 (stereo image dataset) and 94.07% on dataset-2 (mobile images), which were 31.11% and 35.78% higher on dataset-1, and 19.07% and 28.86% higher on dataset-2 than the YOLOv5 and YOLOv7 algorithms, respectively for one class. The YOLOv4 results with one class (bud) showed a mean average precision (mAP) of 98.90%, F1 score of 96.00%, Recall of 98.00%, and precision of 93.00%. While the corresponding values for three classes (silver tip, green tip, tight cluster) are 84.70%, 82.00%, 86.00%, and 77.00% respectively. The proposed method shows great potential for the real-time rapid detection of the apple bud location and its growth stages in complex orchard scenarios. This model could lay the foundation for the machine vision unit of the robotic apple flower bud thinning system.*

Keywords. *computer vision, YOLO, robotic bud thinning, LED-Stereo Vision image acquisition.*

INTRODUCTION

Presently, pruning and thinning (chemical, mechanical, and hand thinning) techniques are mostly used for crop load management. However, achieving accurate target fruit numbers per tree is still a challenging task. Pruning has a risk of low-temperature injury and cannot optimize fruit density, chemical thinning results in unpredictable fruit numbers per tree. mechanical thinning reduces more flower, damage to leaf tissue and tree, lack of selectivity, and the risk of spreading disease, and hand thinning is expensive and laborious job.

An alternative technique of crop load management is artificial spur extinction (ASE) also referred to as bud extinction or bud thinning, which includes manipulation of fruit bud density in the tree canopy. Bud thinning primarily focuses on reducing the total number of floral buds of fruiting branches to optimal levels. The main purpose of bud thinning is to manage tree canopy, improve fruit quality, and tree health, maximize yield, and regularity of production and manage biennial bearing and good economic return. Bud thinning at an early stage can redistribute the fruit buds over the tree to manage the crop load. Bud thinning removes buds just after bud break that can reserve resources and effectively regulate the nutrient supply in a tree to support the healthy initial growth of retained floral buds (Tabing et al., 2016).

Manual bud thinning can be labor-intensive, and automatic or robotic bud thinning is one of the alternative solutions. It is a selective crop load management technique in which an end-effector accurately removes the selected bud with the help of a cutting blade and scissors. In the robotic thinning systems, the vision system plays the role of eye as human eye to detect and localize bud. A fast, efficient, and robust vision system with real-time bud detection at different growth stages under natural orchard environments is the great significance for automation in robotic bud thinning. In recent years, researchers have tried deep learning methods based on convolutional neural networks (CNN) to effectively promoted recognition and positioning of fruit/vegetables for fruit/ vegetable harvesting robots and pruning robots, mainly because CNN has the potential to learn shallow and deep features of objects autonomously. To address the challenges of crop load management, several studies have been focused on using computer vision and 3D reconstruct technology towards branch detection and localization such as branch identification of tall spindle apple trees for robotic pruning (Adhikari & Karkee, 2011)**.**

Researchers have employed deep learning in many agriculture applications (Kamilaris & Prenafeta-Boldú, 2018), such as apple flowers detection (Dias et al., 2018; Tian et al., 2019; Wu et al., 2020), apple fruitlet detection before fruit thinning (Wang & He, 2021), apple fruits detection and counting (Koirala et al., 2019; Vasconez et al., 2020), real-time kiwifruit flower and bud detection for robotic pollination (Li et al., 2022), real-time apple detection for picking robot (Yan et al., 2021), apple branches identification (Majeed et al., 2018; Zhang et al., 2018). Additionally, 3D skeletons of apple trees used the 3D camera for the identification of pruning branches (Karkee et al., 2014), and apple bud classification (Xia et al., 2021).

With the rapid development of deep learning-based methods to detect objects with excellent performance, it has a powerful feature extraction ability to extract features from densely distributed target objects and the robustness of CNN makes it possible to recognize under complex environments such as an orchard (Li et al., 2022; Wu et al., 2020). The deep learning method has two-stage and one-stage detection methods, Fast R-CNN (Girshick et al., 2014), Faster R-CNN (Ren et al., 2015), and Mask RCNN come in the category of two-stage detection. Gao et al. (2020) applied Faster-RCNN with the VGG16 network for multi-class apple detection in the SNAP apple tree system. the overall mAP of the four classes was 0.879. Yu et al. (2019) proposed Mask R-

CNN to overcome the problems of poor universality and robustness of strawberry detection and categories (ripe or unripe fruit) in a non-structural environment. Compared to two-stage detection networks (Fast RCNN, Faster RCNN, Mask RCNN), one-stage detection network (YOLO, SSD) has higher detection accuracy and faster detection speed (Gao et al., 2020; Li et al., 2022; Li et al., 2021; Wu et al., 2020). YOLO is a unified model that uses an end-to-end neural network to detect and classify objects all at once, which provides fast and accurate object detection in real-time.

Li et al. (2022) applied YOLOv4 and YOLOv3 for kiwi flower and bud detection simultaneously and found that YOLOv4 achieve a better result in real-time kiwifruit flower and bud detection simultaneously. Wang & He (2021) proposed a fine-tuned YOLOv5s method for rapid and accurate detection of apple fruitlet using transfer learning with 8 ms per image as the detection time. Wu et al. (2020) proposed a channel-pruned YOLO v4 deep learning algorithm to achieve fast and accurate real-time apple flower detection with compressed model size. the model achieved 97.31% of mAP and 72.33 f/s detection speed. Wang et al. (2022) used the YOLOv4 network with the MobileNetV3 lightweight network for dense plums detection in a real and complex orchard environment.

Researchers were focused only on flower detection, and no study has been focused on apple flower bud detection at a very early stage. The challenges for apple bud detection can attribute to their tiny shape, varying sizes, appearance, similar color to branches, and complex orchard environment, which all made identification of the bud substantially difficult. To address these challenges, an effective deep learning network is required to be capable of detect these tiny buds in a complex environment with fast and accurate detection.

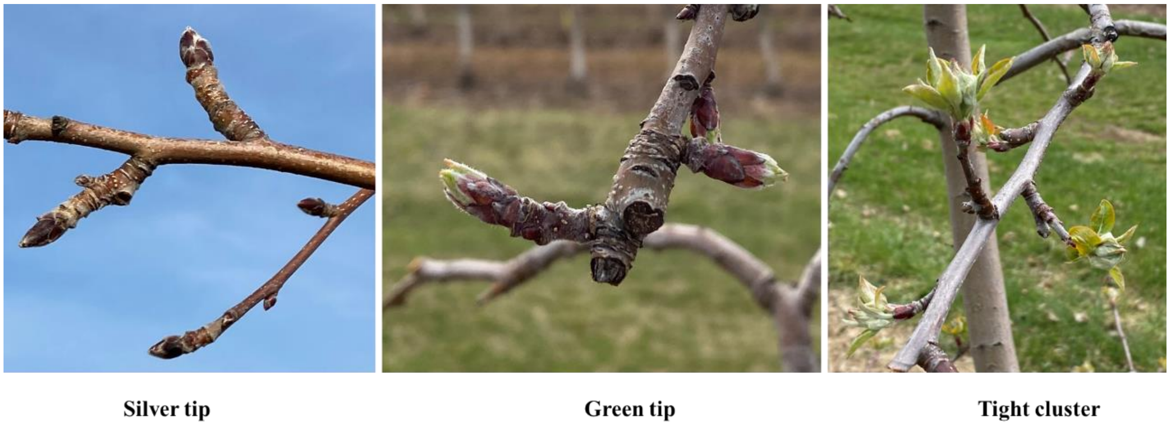
YOLO is a unified model that uses an end-to-end neural network to detect and classify objects all at once. In this study, state-of-the-art YOLOv4 and YOLOv5 were both employed for real-time bud detection at different growth stages. The YOLOv4 bud detection model was fine-tuned to improve the accuracy and detection speed. The specific objectives of this study were:

1) Employ the YOLOv4, YOLOv5, and YOLOv7 models to detect tiny buds from apple trees in orchard environment.
2) Compare the performance of three tested models on two different datasets (stereo-vision images dataset and mobile images dataset) and two labeling methods (one class and three classes).

MATERIALS AND METHODS

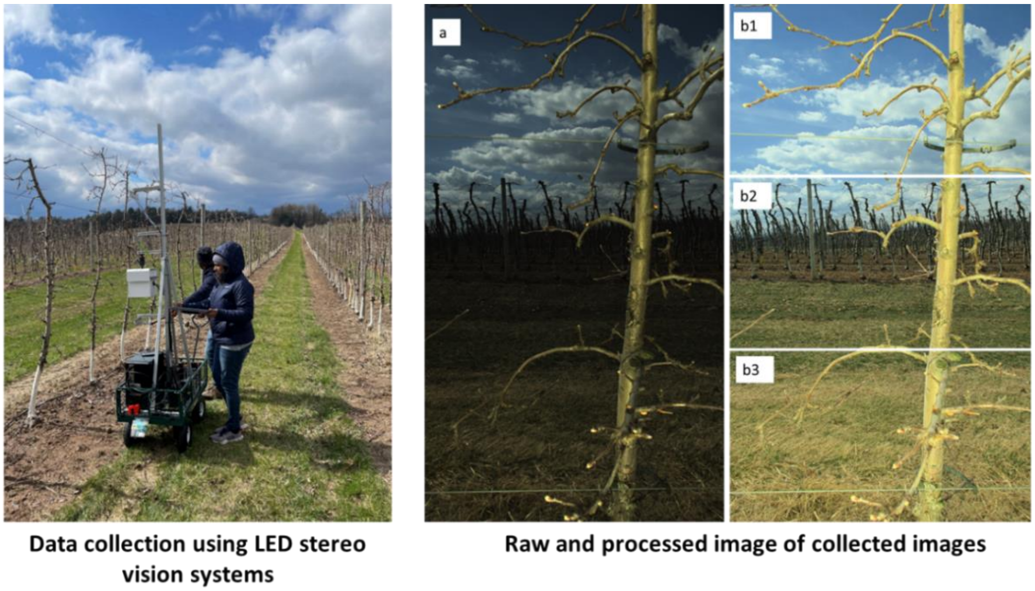
Image data acquisition

In this study, image acquisition was conducted using two different imaging methodologies, stereo vision camera and a mobile phone. The image data was collected from Penn State Fruit Research Extension Center, Pennsylvania, USA, from March 3 to April 3, 2022. The stereo vision image acquisition system includes two FLIR Blackfly S cameras (model BFS-U3-88S6C-C), mounted in a stereo configuration with a resolution of 4096*2160 pixels. This system saved images in Portable Network Graphics (PNG) format. The 12 Cree LEDs were attached to an active flash system for artificial illumination to capture images at constant illumination and subside the motion blur effect. To obtain distortion-free and intrinsic parameters camera images, both cameras were calibrated before capturing the images (Mirbod et al., 2020). The whole LED stereo vision image acquisition system was set up in a cart, and the cart was dragged between two rows at 1 m/s speed to collect the images at 3 Hz. The distance between the camera and the tree was 1 m during image acquisition. The dataset was collected at three bud growth stages, including silver tip, green tip, and tight cluster (Figure 1).



Silver tip Green tip Tight cluster

Figure 1. Different growth stages of apple bud after dormant



Data collection using LED stereo vision systems

Raw and processed image of collected images

Figure 2. Data collection and image preprocessing

Table 1. Outlines the datasets acquired at all three bud growth stages from both (stereo and mobile) imaging methodologies.

Datasets	Growth stages			Total
	Silver tip	Green tip	Tight cluster	
Stereo vision	820	960	882	2662
Mobile phone	250	380	220	850

Data construction

Image pre-possessing such as light enhancement and image divider was applied to improve the accuracy of identification. Dehazing, a pre-processing algorithm, was used to improve visibility in naturally degraded (by low-visibility weather) images. An example of raw and pre-processed acquired images is shown in figure 2. Since buds are very tiny objects, an image divider was used to divide images into three equal parts to help with image labeling (Figure 2 b1, b2, and b3). An example of raw and pre-processed acquired images is shown in figure 2. To obtain the ground truth for subsequent training, Makesense.ai, an image annotation tool, was employed to draw bounding boxes and classify categories manually for 2650 stereo images and 850 mobile images. At the time of labeling, it was ensured that the bud should be in the center of the bounding box.

The data were categorized into two parts, the first category is one class (Bud), where the silver tip and green tip stages merge. The second category is three classes (silver tip, green tip, and tight cluster). These Two categories were defined in the annotation tool to label buds in the images. A .txt annotation format is required to train the YOLO model. Therefore, each class and location of the images were annotated with their corresponding information and saved in .txt format. The whole bud dataset was partitioned into 7:2:1 ratio for training, testing, and validation. A Python-based open-source software makesense.ai has been used to annotate the target classes in images. The models have been trained with transfer learning by using the pre-trained weights. The detection model has been trained and tested in a local system on a single 16 GB NVIDIA GeForce RTX 2080 GPU.

YOLO network architecture

The YOLO (You Only Look Once) network is a one-stage object detection algorithm which makes it versatile for real-time object detection. The YOLO series, including YOLOv5, YOLOv4, and

YOLOv3, evolved from YOLO. YOLO employs end-to-end convolutional neural networks (CNN) to predict object position coordinates and classification, with a single pass of images into CNN making detection fast. It is based on the idea of segmenting an image into S*S square grid cells (Redmon and Farhadi, 2018). Each grid is responsible for predicting the boundary boxes for the target (Bochkovskiy et al., 2020). YOLO network mainly consists of three components, (1) the backbone, a deep convolution layers that extract features from input images, (2) the neck, which works as a feature aggregator that collects generated feature maps from different layers of the backbone, and (3) the head, it performs the prediction of the bounding box and the confidence score of that prediction.

YOLOv4

YOLOv4 uses CSP densenet53 as a backbone CSP stands for Cross-Stage-Partial connections. The second stage of this neural network is the neck to collect feature maps from the different stages of the backbone and aggregate them for send to the head. YOLOv4 uses a modified path aggregation network (PANet), which includes bottom-up path augmentation that allows better propagation between lower layers and the topmost feature. Then, Adaptive feature pooling is used to aggregate features from all feature levels for each proposal. SPP block added between the feature extractor and feature aggregator to generate fixed-size features regardless of the input size and to increase the receptive field without affecting network operation speed.

YOLOv4 uses a YOLOv3 (anchor-based) head, and the main function of the head is to predict the confidence score for each class and bounding box coordinates (x, y, w, h). YOLOv3 head is capable of generating three detection feature maps (large (16 x 16), medium (26 x 26), and small (52 x 52)) to perform multi-scale prediction. YOLOv4 also introduced Bag of Freebies (BoF) & Bag of Specials (BoS).

YOLOv5

YOLOv5 is an upgraded version of YOLOV3 by adding BottleneckCSP, mosaic, Focus, SPP, and PANet module (Wang et al., 2022). YOLOv5 implemented in PyTorch framework instead of Darknet. It has the same CSPDarknet53 backbone with a focus layer. It is also a lightweight model than YOLO v4. YOLOv5 has five different sizes v5n, v5s, v5m, v5l, and v5x based on simple to complex network structures (depths and widths). Although, more complex networks provide better detection but require high computation power.

YOLOv7

The YOLOv7 model is the latest version of the YOLO models. Extended efficient layer aggregation networks (E-ELAN), an extended version of the ELAN computational block enhance the learning ability of the model by using “expand, scramble, merge cardinality" without eradicating the original gradient path. E-ELAN alters only the computational block in the architecture, without changing the transition layer architecture.

Evaluation of the model performance

To validate the algorithm performance and robustness, Intersection Over Union (IOU), Precision, mean average precision (mAP), Recall, and F1 score, were used on the test dataset. The number of objects that were detected correctly and false positives generated can be determined by Intersection Over Union (IoU) metric. IOU is a metric that quantifies the degree of overlap between ground truth and predicted bounding box. The predicted bounding box is considered a good and acceptable detection if IoU scores more than 0.5. Otherwise, it is unacceptable.

$$IoU = \frac{|P \cap G|}{|P \cup G|}$$

where “P” represents the prediction bounding box and the “G” represents the ground truth bounding box. Precision measures the percentage of actually correct positive predictions. It measures the level of accuracy of the model prediction.

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

Where TP is the number of true positive cases, FP is the number of false positive cases, and FN is the number of false negative cases. Recall measures the percentage of actual positives out of all Ground Truths

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

F1-score is the harmonic mean of precision and recall. It calculates the balance between precision and recall.

$$F1 = 2 \times \frac{Precision \ \times Recall}{Precision + Recall}$$

AP calculates the area under the precision-recall curve for each class and at the different thresholds.

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

Where p, r is precision and recall respectively. The mean average precision is the average of AP with different IoU and all the classes.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_K$$

$$AP_K = the\ AP\ of\ class\ k$$

$$n = the\ number\ of\ classes$$

RESULTS AND DISCUSSION

In this study, images have been collected for different growth stages of apple buds from the apple orchard located in FREC at Penn State University, USA. The dataset 70% of images have been randomly chosen for the training dataset to train the proposed detection model, and 30% of images are selected for both validation and test datasets. To obtain better accuracy during training, the image size of the input dataset was set to 1280×1280 due to the tiny bud size. The parameters such as initial learning rate, number of channels, momentum value, decay regularization referred to the original parameter in the YOLOv4, YOLOv5, and YOLOv7. In total of 80000 training steps were selected for better analysis of the training process. The resolution of the input image is 1280 × 1280 for all bud detection model. As we discussed earlier, we have collected data from 2 devices. In order to verify the effectiveness of for apple flower bud detection model, three object detection algorithms and two datasets were compared in this study.

Comparisons of state-of-the-art models

The rapid and precise identification of buds will not only be helpful for bud count estimation on the tree branches, but also it will provide a technical reference to the robotic bud thinning system. Therefore, in this study, three object detection algorithms YOLOv4, YOLOv5, and YOLOv7 were compared to analyze the bud detection performance. To achieve the best results from each model, the image input sizes were 1280×1280 for all models. The test results from table 2 showed that the mAPs of YOLOv4, YOLOv5, and YOLOv7 algorithms with one class on dataset-1 were 98.99%, 75.50%, and 72.90%, respectively; the model sizes were 244 MB, 41 MB, and 71.8 MB, respectively.

Table 2. Comparison of P, R, F1-score, and mAP between state-of-the-art models: YOLOv4, YOLOv5, and YOLOv7 on dataset-1.

	Dataset-1 (Stereo Data)					
	1 Class			3 Classes		
	YOLOv4	YOLOV5	YOLOv7	YOLOv4	YOLOv5	YOLOv7
mAP	98.9	75.5	72.9	95.3	80.9	76
P	93.0	76.5	73.7	89	77.3	72.4
R	98.0	71.2	71.2	95	78.0	74.5
F1 score	96.0	73.8	72.4	92	77.6	73.4

Table 3. Comparison of P, R, F1-score, and mAP between state-of-the-art models: YOLOv4, YOLOv5, and YOLOv7 on dataset-2.

	Dataset-2 (Mobile Data)					
	1 Class			3 Classes		
	YOLOv4	YOLOv5	YOLOv7	YOLOv4	YOLOv5	YOLOv7
mAP	94.07	79	73	98.37	76.5	66.9
P	92	75.6	74.6	95	73.7	66.9
R	91	73.9	67.7	96	73.6	66.2
F1 score	91	74.74	70.98	95	73.65	66.55

Table 2 also indicates that the performance of YOLOv7 performed the worst compared to YOLOv4, YOLOv5 for one class with minimum P, R, and F1-score of 20.75%, 27.34%, and 24.56% respectively. While YOLOv5 demonstrates better performance compared to YOLOv7 with a 1.83% increase in F1-score and a 3.56% increase in mAP, respectively. However, YOLOv4 provided superior results compared to YOLOv5 with 21.56%, 37.60%, 30.17%, and 31.11% increases in P, R, F1-score, and mAP, respectively. To summarize, the YOLOv4 with one class outperforms other state-of-the-art models in terms of detection accuracy, which makes it a promising model for high-performance real-time bud detection.

Performance of models with one class and multiple classes

To verify the performance of the model in detecting buds at different categories of growth stages. The images were collected at three different growth stages. These growth stages are silver tip, green tip, and tight cluster. To verify the generalization ability of the model data were annotated in two categories, the first category is one class (Bud), where the silver tip and green tip stages are merged. The second category is three classes where silver tip, green tip, and tight cluster are separated. The models detection results on dataset-1 with both categories are shown in figure 3. From the figure 3, it can be seen intuitively that the recognition accuracy for both categories is different. The specific results of the comparison between the two categories for dataset-2 are shown in table 3. YOLOv4 algorithm mAP reached 98.99% with one class and 95.25% with three classes on dataset1. The YOLOv5m and YOLOv7 mAP reached 75.5%, and 72.9% with one class, and 80.90%, and 76.0% respectively with three classes on the same dataset. It can be seen from table 3 that the YOLOv4 with one class provided more accurate results as compared to the three classes on dataset-1. According to table 4, the APs with YOLOv5 in the tight cluster, silver tip, and green tip were 2.50%, 20.17%, and 23.1% lower than those with YOLOv4. Moreover, the YOLOv7 AP of silver tip and green tip were 25.18%, 32.47%, and the tight cluster are 3.73% lower than the YOLOv4 respectively. Which meant that YOLOv4 achieved better detection results than the other state of the art models.

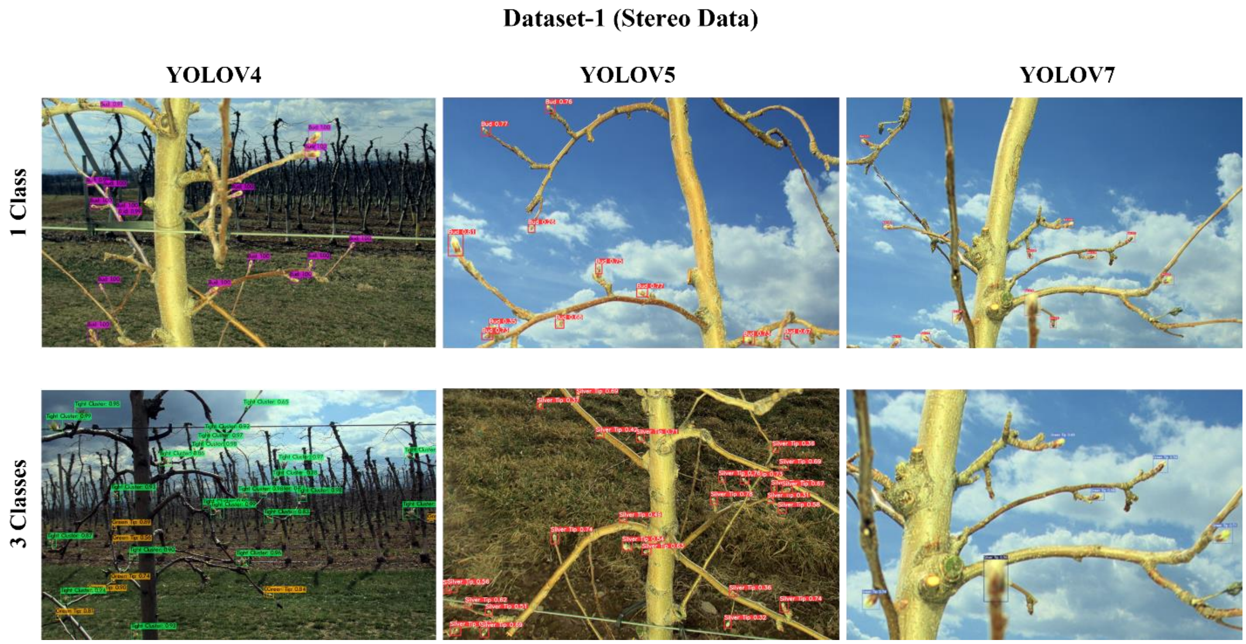


Figure 3. Detection results of different growth stages of flower bud on datasets-1 from YOLOV4, YOLOv5, YOLOv7 models.

The test result also revealed that the AP of the tight cluster class was higher than the silver tip or green tip in all models, the reason for this is the tight cluster size is larger than the silver tip and green tip. Besides, sometimes it is difficult to differentiate between silver tip and green tip. Hence, the model had a poor ability to distinguish between the silver tip and green tip, being the cause of the simultaneous decline in the AP indicators of both. However, YOLOv4 achieves reasonable Ap in the silver tip and green tip detection over the YOLOv5 and YOLOv7. Thus, it is apparent from the previous comparison that the YOLOv4 model significantly outperforms YOLOv5 and YOLOv7 in terms of overall performance. Which makes it effective and feasible model for accurate and fast bud detection.

Table 4. Comparison of average precision results of YOLOv4, YOLOv5, and YOLOv7 on Stereo and mobile datasets.

Average Precision Results						
	Dataset-1 (Stereo Data)			Dataset-2 (Mobile Data)		
	YOLOv4	YOLOv5	YOLOv7	YOLOv4	YOLOv5	YOLOv7

Silver tip	93.96	75.00	70.30	96.77	62.40	47.00
Green Tip	93.74	72.10	63.30	99.18	74.90	65.50
Tight Cluster	98.06	95.60	94.40	99.15	92.20	88.20

Generalizability of state-of-the-art models

Another image dataset (dataset-2) was used to demonstrate the generalizability of YOLOv4, YOLOv5, and YOLOv7. The models detection results on dataset-2 with both categories are shown in figure 4. Detection results of YOLOv4, YOLOv5, and YOLOv7 on dataset-1 and dataset-2 are shown in table 3. The mAPs of YOLOv4, YOLOv5, and YOLOv7 on dataset-2 with one class (94.00%, 79.00%, and 73.00%, respectively) were slightly lower than those on dataset-1. F1-scores have been compared of these models to evaluate the efficiency detection performance which shows 5.2%, 1.34%, and 1.98% declination in YOLOv4, YOLOv5, YOLOv7 respectively on dataset-2 with 1 class over dataset-1. Overall dataset-1 showed better performance on all the models as compared to dataset-2.

However, the AP of all silver tip, green tip, and tight cluster detection with YOLOv4 on dataset 2 was 3%, 5.8%, and 1.1% higher than that of dataset-1, while the AP of silver tip, and tight cluster by YOLOv5 on dataset-2 was 62.4%, 92.2% which was 20.2%, 3.68% lower than that of dataset-1. As shown in table 4, the mAP achieved by YOLOv4 with 1 class was 5.23% higher than YOLOv5 on dataset-2, while the mAP of three classes achieved by YOLOv4 on dataset-2 was 3.27% higher than dataset-1. These results interpret that the detection performance of YOLOv4 on dataset-1 was better than that of dataset-2. The reason for the variation in the results of the same model on different datasets could be the dataset-2 resolution was lower than dataset-1 and the images in dataset-2 were also less.

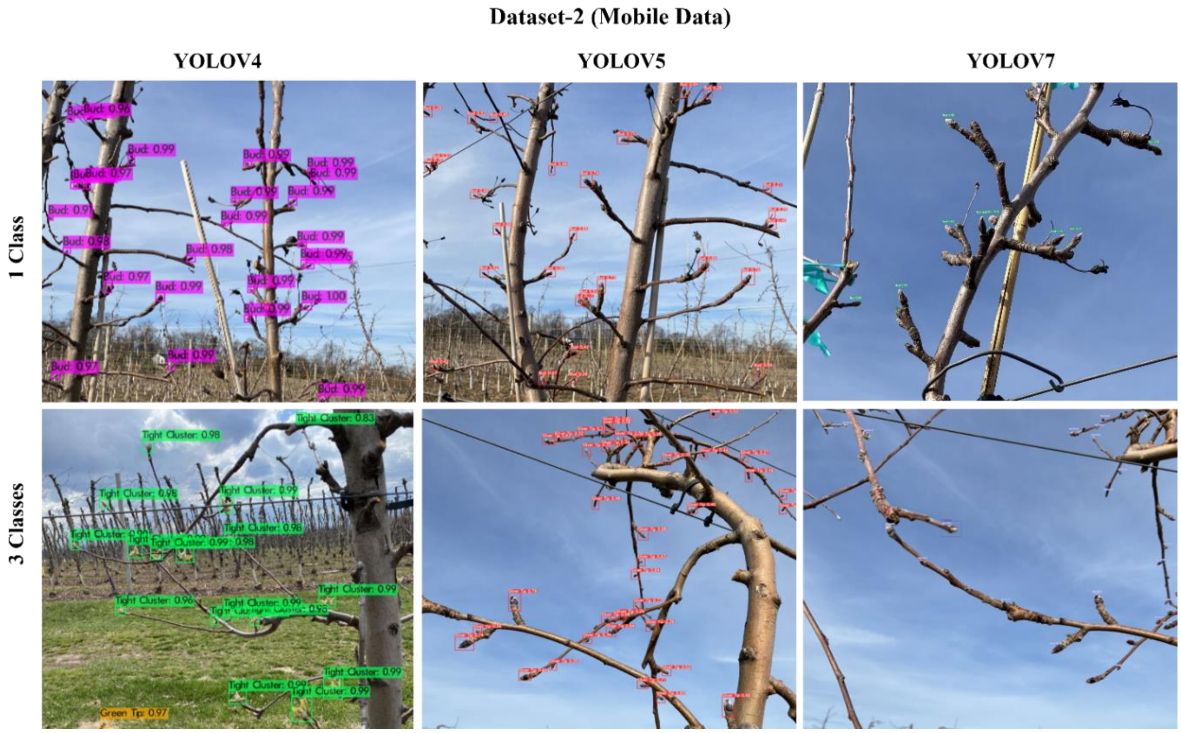


Figure 4. Detection results of different growth stages of flower bud on datasets-2 from YOLOV4, YOLOv5, YOLOv7 models.

In summary, most of the previous studies have developed algorithms for apple fruits, flowers, and branches detection for different purposes. However, there are no such studies on apple flower bud detection at an early stage. The YOLOv4 model used in this study can be used as a vision system in an automated apple bud-thinning robot and will provide bud location guides to the end-effector. For early crop load management, robotic bud thinning could be the potential solution to reduce labor requirements and production costs in a long term. Although it is notable that the YOLOv4 achieved considerable results since its mAP was the highest among the two contrasted algorithms, which represents the model has excellent target detection ability. However, the size of the YOLOv4 model is 244 MB, which is relatively large in terms of recognition of objects and may rise the deployment cost in the embedded devices of the vision system of the bud thinning robot. In the future, the vision system would also need the branch diameter and bud count algorithm development for decision-making for bud adjustment. In the next step, in order to adjust the number of buds on a branch, a bud removal end-effector will integrate with the manipulator and vision system.

CONCLUSION

State-of-the-art deep learning models have attained promising detection accuracy of agricultural objects in natural environment. Apple flower bud detection method based on YOLOv4-with transfer learning was proposed in this study to obtain good detection performance. The study compared the detection performance of YOLOv4 with YOLOv5, and YOLOv7 networks with different datasets and classes. The results showed that under the same conditions, the YOLOv4 algorithm achieved the highest accuracy and generalization ability among the three algorithms for the detection of flower buds in different datasets, which met the requirement of detection time for robotic bud thinning. Furthermore, it is found that YOLOv4 significantly achieved the most satisfactory detection results in different bud growth stage detection followed by YOLOv5 and YOLOv7 in conditions of complex scenes, which signifies it has better performance on small objects. The YOLOv4 model can be potentially used for robotic bud thinning in the complex orchard environment. Future work will focus on the application of this model in the vision system with integration of a robotic bud thinning system to remove flower buds at different growth stages.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the SCRI, USDA-NIFA, and Penn State Fruit Research and extension center (FREC) for their support and funding for this project. This research is supported by USDA NIFA SCRI Grant No. 2020 51181 32197.

REFERENCES:

Adhikari, B., & Karkee, M. (2011). 3D reconstruction of apple trees for mechanical pruning. *American Society of Agricultural and Biological Engineers Annual International Meeting*

2011, ASABE 2011, *1*, 303–318. https://doi.org/10.13031/2013.38139

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. https://doi.org/10.48550/arxiv.2004.10934

Dias, P. A., Tabb, A., & Medeiros, H. (2018). Multispecies Fruit Flower Detection Using a Refined Semantic Segmentation Network. *IEEE Robotics and Automation Letters*, *3*(4), 3003–3010. https://doi.org/10.1109/LRA.2018.2849498

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M., & Zhang, Q. (2020). Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. *Computers and Electronics in Agriculture*, *176*(May), 105634. https://doi.org/10.1016/j.compag.2020.105634

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/J.COMPAG.2018.02.016

Karkee, M., Adhikari, B., Amatya, S., & Zhang, Q. (2014). Identification of pruning branches in tall spindle apple trees for automated pruning. *Computers and Electronics in Agriculture*, *103*, 127–135. https://doi.org/10.1016/j.compag.2014.02.013

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, *162*, 219–234. https://doi.org/10.1016/J.COMPAG.2019.04.017

Li, G., Suo, R., Zhao, G., Gao, C., Fu, L., Shi, F., Dhupia, J., Li, R., & Cui, Y. (2022). Real-time detection of kiwifruit flower and bud simultaneously in orchard using YOLOv4 for robotic pollination. *Computers and Electronics in Agriculture*, *193*(December 2021), 106641. https://doi.org/10.1016/j.compag.2021.106641

Li, H., Li, C., Li, G., & Chen, L. (2021). A real-time table grape detection method based on improved YOLOv4-tiny network in complex background. *Biosystems Engineering*, *212*, 347–359. https://doi.org/10.1016/j.biosystemseng.2021.11.011

Majeed, Y., Zhang, J., Zhang, X., Fu, L., Karkee, M., Zhang, Q., & Whiting, M. D. (2018). Apple Tree Trunk and Branch Segmentation for Automatic Trellis Training Using Convolutional Neural Network Based Semantic Segmentation. *IFAC-PapersOnLine*, *51*(17), 75–80. https://doi.org/10.1016/j.ifacol.2018.08.064

Mirbod, O., Choi, D., Heinemann, P., & Marini, R. (2020). *Towards Image-Based Measurement of Accurate Apple Size and Yield Using Stereo Vision Cameras Written for presentation at the 2020 ASABE Annual International Meeting Sponsored by ASABE*. 1–6.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.48550/arxiv.1506.01497

Tabing, O., Parkes, H. A., Middleton, S. G., Tustin, D. S., Breen, K. C., & Van Hooijdonk, B. M. (2016). Artificial spur extinction to regulate crop load and fruit quality of “Kalei” apple. *Acta Horticulturae*, *1130*, 273–277. https://doi.org/10.17660/ActaHortic.2016.1130.40

Tian, M., Chen, H., & Wang, Q. (2019). Detection and Recognition of Flower Image Based on SSD network in Video Stream. *Journal of Physics: Conference Series*, *1237*(3), 032045. https://doi.org/10.1088/1742-6596/1237/3/032045

Vasconez, J. P., Delpiano, J., Vougioukas, S., & Auat Cheein, F. (2020). Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. *Computers and Electronics in Agriculture*, *173*, 105348. https://doi.org/10.1016/J.COMPAG.2020.105348

Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, *210*, 271–281. https://doi.org/10.1016/j.biosystemseng.2021.08.015

Wang, L., Zhao, Y., Liu, S., Li, Y., Chen, S., & Lan, Y. (2022). Precision Detection of Dense Plums in Orchards Using the Improved YOLOv4 Model. *Frontiers in Plant Science*, *13*(March), 1–14. https://doi.org/10.3389/fpls.2022.839269

Wang, Z., Jin, L., Wang, S., & Xu, H. (2022). Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. *Postharvest Biology and Technology*, *185*(November 2021), 111808. https://doi.org/10.1016/j.postharvbio.2021.111808

Wu, D., Lv, S., Jiang, M., & Song, H. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Computers and Electronics in Agriculture*, *178*(August), 105742. https://doi.org/10.1016/j.compag.2020.105742

Wu, L., Ma, J., Zhao, Y., & Liu, H. (2021). Apple detection in complex scene using the improved yolov4 model. *Agronomy*, *11*(3). https://doi.org/10.3390/agronomy11030476

Xia, X., Chai, X., Zhang, N., & Sun, T. (2021). Visual classification of apple bud-types via attention-guided data enrichment network. *Computers and Electronics in Agriculture*, *191*(September), 106504. https://doi.org/10.1016/j.compag.2021.106504

Yan, B., Fan, P., Lei, X., Liu, Z., & Yang, F. (2021). A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sensing*, *13*(9), 1–23. https://doi.org/10.3390/rs13091619

Yu, Y., Zhang, K., Zhang, D., Yang, L., & Cui, T. (2019). Optimized faster R-cnn for fruit detection of strawberry harvesting robot. *2019 ASABE Annual International Meeting*. https://doi.org/10.13031/aim.201901123

Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X., & Gao, Z. (2018). Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN). *Computers and Electronics in Agriculture*, *155*, 386–393. https://doi.org/10.1016/J.COMPAG.2018.10.029