The International Society of Precision Agriculture presents the
15th International Conference on Precision Agriculture
26–29 JUNE 2022
Minneapolis Marriott City Center | Minneapolis, Minnesota USA

Real-time detection of picking region of ridge-planted strawberries based on YOLOv5s with a modified neck

Zixuan He[1], Safal Kshetri[1], Manoj Karkee[1], Qin Zhang[1]

[1]Department of Biological Systems Engineering, Center for Precision and Automated Agriculture Systems, Washington State University, Prosser, WA, USA

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract.

Robotic strawberry harvesting requires machine vision system to have the ability to detect the presence, maturity, and location of strawberries. Strawberries, however, can easily be bruised, injured, and even damaged during robotic harvest if not picked properly because of their soft surfaces. Therefore, it is important to cut or pick the strawberry stems instead of picking the fruit directly. Additionally, real-time detection is critical for robotic strawberry harvesting to adapt to the changing field environment quickly. In this study, first, a detection algorithm was created for accurately localizing strawberries and their picking regions based on object detection network (YOLOv5s). The neck of YOLOv5s was replaced with a feature pyramid network (FPN) from path aggregation net (PA-Net) to reduce the complexity in network structure. This YOLOv5s model with FPN (YOLOv5s-FPN) was used to detect three maturity levels (immature, nearly mature, mature) of strawberries. Then, the model was used to detect picking region in strawberry stems using strawberry bounding boxes detected in the previous step as the input. For comparison, the original YOLOv5s was trained with same environment and datasets. The results showed that YOLOv5s-FPN model achieved the mean average precision (mAP) of 92.3% based on testing strawberry canopy dataset. In immature, nearly mature, and mature classes, it achieved an average precision of 93.6%, 91.7%, and 91.7%, respectively. For picking region, it achieved a mean average precision of 82.8%. Compared to YOLOv5s, the YOLOv5s-FPN had smaller size of 12.0 Mb (85.7% of YOLOv5s) and faster detection speed of 36.5ms (83.7% of YOLOv5s) on image of resolution 640×640 pixels. However, the performance YOlOv5s-FPN was equally good compared to YOLOv5s (mAP in strawberry detection: 92.5%; mAP in picking region: 82.6%). The YOLOv5s-FPN developed in this study showed good potential as a means for providing real-time detection of strawberry locations and corresponding stem regions for robotic twisting or cutting of stem as a way to harvest strawberries.

Keywords. *YOLOv5, strawberry detection, picking region, deep learning*

Introduction

Strawberry robotic harvest is being sought as an alternative to manual harvest due to the aging workforce and decreasing immigrants (Delbridge, 2021). Strawberry detection, which is the first and one of the most significant tasks during robotic strawberry harvesting, provides the presence, location, and maturity of strawberries in the canopies under field conditions (He et al., 2021). Object detection methods based on CNNs are suitable for strawberry detection as they could find targets in the RGB image and provide the grading or classification on these detected targets (Zou et al., 2019). Machine vision systems applying region based convolutional neural networks (R-CNN) were increasingly used in developing robotic harvester and helped robots to accurately locate target fruits, as well as estimating fruit/crop maturity (e.g., Lamb et al., 2018; Chen et al., 2019).

Compared to the two-stage models such as R-CNN, faster-R-CNN, and Mask-R-CNN, You-Only-Look-Once (YOLO) models (Redmon et al., 2016) relate the detection results (e.g., bounding boxes and class probability) directly with a single feed-forward network, making them computationally much more efficient. YOLOv2 was improved greatly on detection accuracy and learning process from YOLO when an anchor was used, which was inspired by faster-R-CNN (Redmon et al., 2017). YOLOv3, with a complex Darknet53 as backbone, could predict more bounding boxes than YOLOv2 with the same input image (Redmon et al., 2018). In addition, with the introduction of spatial pyramid pooling and the path aggregation network (Liu et al., 2018), it was demonstrated that YOLOv4 can achieve an average precision (AP) of 43.5% on the MS COCO dataset and ~65 fps processing time on a Tesla V100 GPU, which was an improvement of 10% and 12%, respectively, compared to those of YOLOv3 (Bochkovsiy, et al., 2020). YOLOv5 has similar structure to YOLOv4 but it contains mosaic functions on data augmentation and auto learning bounding boxes anchors (Jocher et al., 2022). Some studies based on both YOLOv4 and YOLOv5 achieved promising results (Lu et al., 2022; Yan et al., 2021; Yu et., 2020) in detecting strawberry. However, strawberry, with their soft surfaces, could be bruised or injured during harvesting when the detection focuses on strawberry for robotic picking. Therefore, it might be better to find the picking points or regions in strawberry stems, besides detecting strawberries for robotic harvesting to avoid the fruit damage.

One of the major challenges in picking region detection study is differentiating picking region of mature strawberries from stems or vines of immature or nearly mature strawberries in the canopies. An approach to address this challenge could be detecting mature strawberries first and then finding their stems. Yun et al (2020) used a post-processing technique based on the image processing methods (shapes and color) to find picking points after conducting strawberry detection using Mask-R-CNN. For picking regions, object detection models based on YOLOv5 could also be used as a solution with higher efficiency and robustness. However, the recognition of strawberries and their picking regions in stems (based on strawberry canopy dataset with 3 classes and the picking region dataset with 1 class) would be different from the detection task (based on COCO dataset with 80 classes) of the original YOLOv5. It is important to decrease the complexity in structure of YOLOv5 to acquire a faster processing speed without affecting the performance of strawberry and picking region detection. Therefore, in this study, a real-time detection algorithm, based on YOLOv5s with a modified neck FPN, was proposed for detecting the strawberries and the picking regions in corresponding stems.

Methodology

In this study (flowchart shown in Fig 1), YOLOv5s with a modified neck FPN (YOLOv5s-FPN) was used for conducting strawberry and picking region detection. Initially, strawberry canopy dataset was labelled with three strawberry maturity classes: (i) immature, (ii) nearly mature, and (iii) mature. Next, the bounding boxes of the mature strawberries were cropped, and a dataset was generated for training the YOLOv5s-FPN to conduct detection on the picking region. Picking regions of strawberries in the corresponding stems were then labeled in the generated dataset. These datasets were divided randomly into training, validation, and testing datasets to train YOLOv5s-FPN and acquire trained weights for strawberry canopy and picking region separately. After training and testing, YOLOv5s-FPN was first used to detect strawberry using the trained weight for strawberry canopy dataset. The model, then, was used to detect the picking regions using the associated weight.

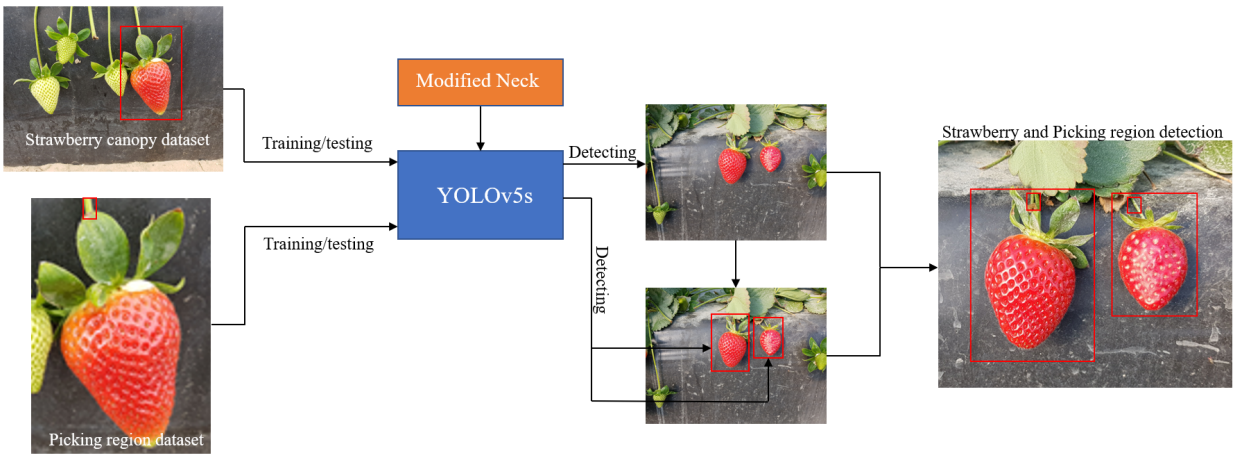

Fig 1. Flowchart of the strawberry detection and picking region detection from YOLOv5s with modified neck

Data pre-processing

An open-source (Pérez-Borrero et al., 2020) strawberry dataset with images (resolution: 640×480 pixels) as JEPG format was used for this study. The images in this dataset were shot at approximately 20 cm from the crop ridges, at about 35 ±10 cm height and angles of 25±10°. There were 1,300 RGB images selected for the strawberry canopy dataset. The grading on strawberry maturities (immature, nearly mature, and mature) followed the method developed by Barnes & Patchett (1976). The examples of maturity classes are shown in Fig 2. Labeling (Tzutalin, 2015) was used as the software tool for image annotation. The strawberry canopy dataset was separated into training dataset (1,000 images), validation dataset (150 images), and testing dataset (150 images). Similarly, picking region dataset was separated into training dataset (500 images), validation dataset (50 images), testing dataset (50 images).

Fig 2. Examples of maturity classes (mature, nearly mature, immature) of strawberries

Object detection model

YOLOv5 was reported to have high detection accuracy (reported highest mAP of 72.7% in COCO dataset) and detection speed lower than 30ms/image (resolution: 1280×1280) based on the V100 Tensor Core GPUs (Jocher et al., 2022). YOLOv5 consists of neck, head, and backbone. The backbone (CSPDarknet53) of YOLOv5 is used for extracting the features of objects. The neck (PA-Net) is used to produce feature scale and the head collects the feature from backbone. The head is applied to generate the output including scores and bounding boxes of the detected objects by applying the anchor boxes.

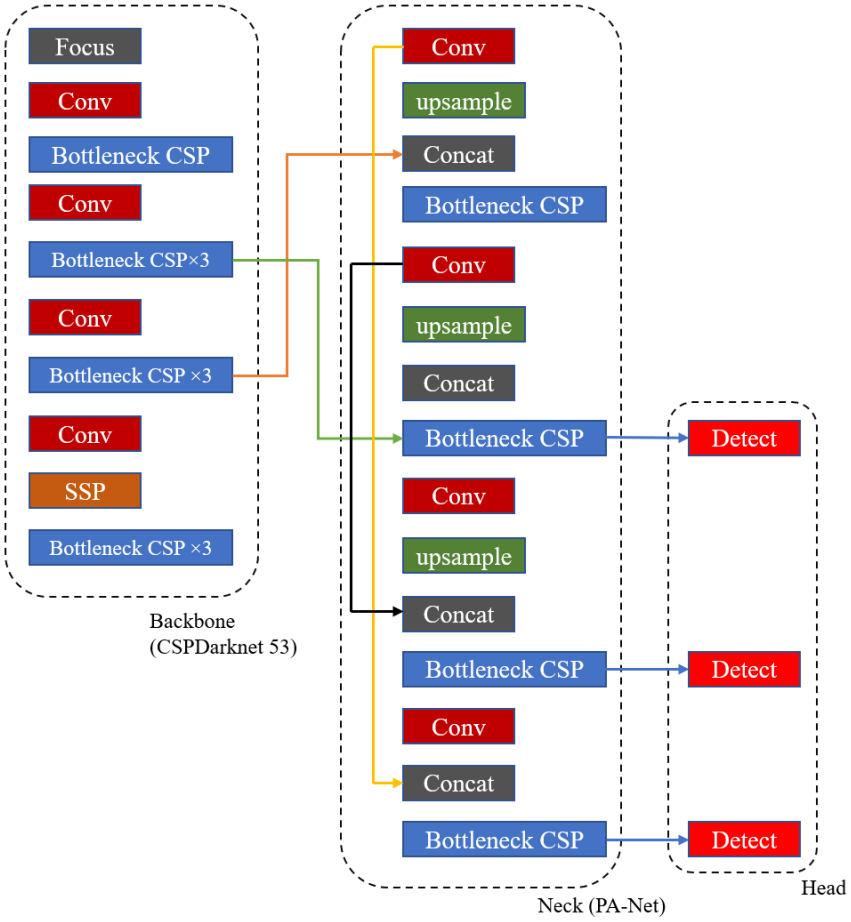

Fig 3. Structure of originalYOLOv5 network.

In this study, YOLOv5s was selected to achieve accurate and real-time detection of both

strawberries and their picking regions. As discussed below, the neck was then replaced from original structure to feature pyramid network (FPN).

Neck structure

In this study, PA-Net of YOLOv5s was replaced with FPN. The comparison between structures of PA-Net and FPN are shown in Fig 4. The main difference between FPN and PA-Net is the architecture direction. FPN has a top-down pathway to generate feature maps between level 3 to level 7 (Lin et al. 2017) while PA-Net is based on the structure of FPN and combines down-top and top-down pathway between level 3 and level 7 to preserve spatial information precisely (Liu et al., 2018). The structure of the YOLOv5-FPN is shown in Fig 5, which is simpler than the original one (Fig 3). In the neck structure of YOLOv5-FPN, FPN only uses the pathway from level 3 to level 5.

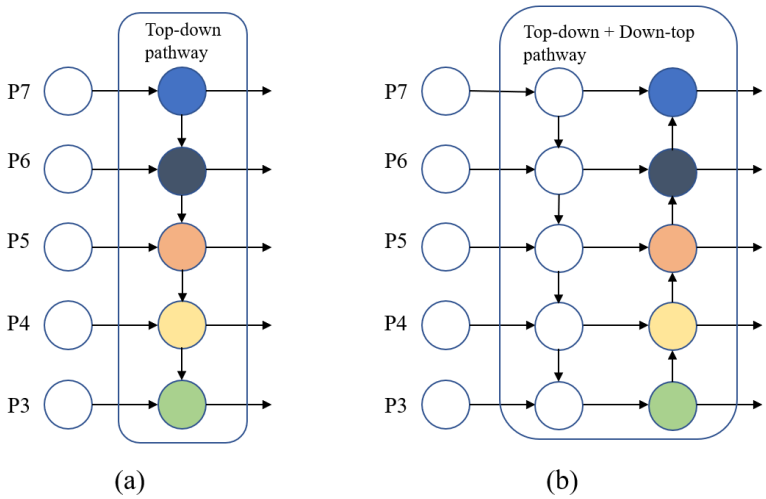

Fig 4. Feature network comparison: a) Top-down pathway in FPN; b) Top-down and down-top pathway in PA-Net

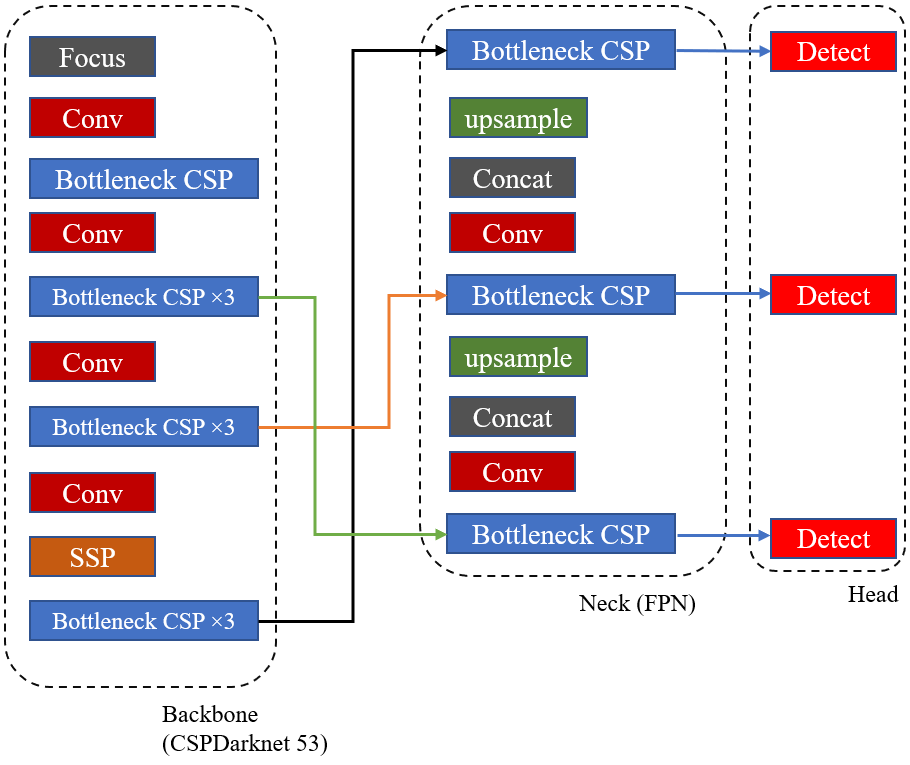

Fig 5. Modified structure of YOLOv5 with FPN

Networks Training

Training parameters

The whole training environment was based on Colab cloud platform, Google (GPU: Tesla PCIE 100; 32GB Ram). Firstly, the number of training epochs was set to 35 for the strawberry dataset and 30 for the picking region dataset with same batch size of 8. The momentum was set to 0.937 with decay weight of 0.005. Besides, the input image size for strawberry canopy dataset and picking region dataset were set to 640×640 pixel to keep the most of features of strawberries and corresponding picking regions. The image augmentations for the training dataset on Hue, Saturation and Value (HSV) were set to 0.015, 0.7, and 0.4 respectively to increase the robustness of YOLOv5-FPN. The number of training epochs for strawberry and picking region detection was set to 35 and 100 respectively. After the training, the best weights with highest mean average precision based on testing datasets were used for detection. A comparative study was conducted between YOLOv5s with original neck PA-Net and YOLOv5-FPN using the same training parameters.

Evaluation Metrics

Strawberry detection results were evaluated using recall (*R*), precision (*P*), *AP* and mean average precision for each class (mAP) with an intersection-over-union (*IOU*) of 50%. *P* and *R* are used for measuring the accuracy of overlap between the predicted and ground truth bounding boxes. The calculation equations are listed as follows:

$$IOU = \frac{|A \cup B|}{|A \cap B|} = \frac{\text{Area(I)}}{\text{Area(U)}} \tag{1}$$

$$\text{Precision } (P) = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

$$\text{Recall } (R) = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$AP = \sum_n (r_{n+1} - r_n) \max_{\tilde{r}: \tilde{r}^3 r_{n+1}} p(\tilde{r}) \tag{4}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{5}$$

where *A* is the area of predicted bounding boxes, *B* is the area of ground truth bounding boxes, Area (*I*) is the intersection of predicted and ground truth bounding boxes, and Area (*U*) is the union of predicted and ground truth bounding boxes. *TP* is the number of true positive objects detected, *FP* is the number of false objects detected, and *FN* is the number of objects falsely not detected as strawberries or picking regions. *AP* was used to show the performance of individual class. mAP was used to show the overall performance under different confidence thresholds. $p(\tilde{r})$ is the precision at recall $\tilde{r}$.

Results and Discussion

The training results on strawberry dataset and picking region dataset are shown in Fig. 6 and Fig. 7, respectively. The training results based on these datasets showed the mAP increased rapid in the early training epochs and became stable after epoch 10. The charts indicate the YOLOv5s-FPN models was trained without overfitting to training datasets in both cases (strawberry dataset and picking region dataset).

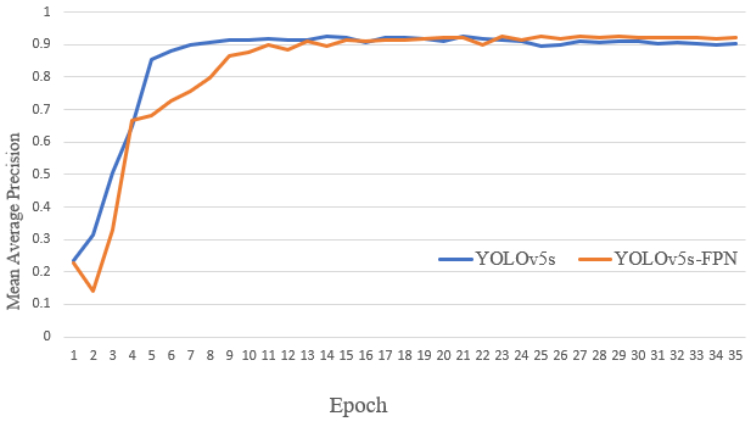

Fig 6. Training results based on strawberry canopy dataset

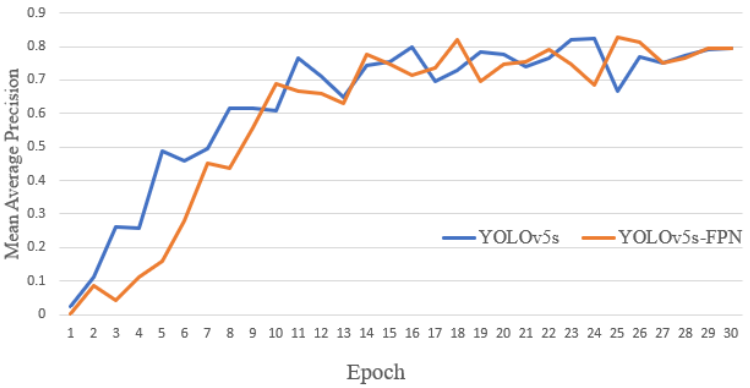

Fig 7. Training results based on picking region dataset

Based on the test dataset (150 images), the results (Table 1) showed that YOLOv5s-FPN model achieved a mAP of 92.3%, which was only marginally lower than the same achieved with original YOLOv5s (92.5%). The results from Table 2 showed that *AP* achieved with YOLOv5-FPN in mature and nearly mature classes have nearly same performance with YOLOv5s while the original YOLOv5s had slightly better performance in immature class, which was 0.9% over YOLOv5s-FPN. The YOLOv5s-FPN, however, had simpler structure with 6.0×10^6 parameters and smaller size of 12.0 Mb (85.7% of original size of YOLOv5s), which apparently decreased the average processing time from 17ms to 14ms during strawberry canopy detection. Similarly, the results based on a test dataset (50 images) showed that YOLOV5s and YOLOv5s-FPN had nearly the same performance with mAP of 82.6% and 82.8% respectively while the processing time of YOLOv5s-FPN was reduced from 14.0 ms to 10.4 ms. Moreover, the overall processing time of YOLOv5s-FPN for strawberry and picking region detection was 36.5ms whereas YOLOv5s had a slower speed of 43.6ms. The YOLOv5s-FPN with lighter weight could satisfied the requirement of real-time detection for robotic strawberry harvesting.

Table 1. Overall performance in strawberry detection in canopy images

Object detection models	P(%)	R(%)	mAP@.5(%)	Processing time@640× 640 (ms)	Number of Parameters	Size (Mb)
YOLOv5s	88.0	86.7	92.5	17.0	7.0×10^6	14.0
YOLOv5s - FPN	87.2	88.6	92.3	14.0	6.0×10^6	12.0

Table 2. Network performance in detecting individual strawberry classes in canopy images

Object detection	AP (%)

models			
	Immature	Nearly mature	Mature
YOLOv5s	94.5	91.4	91.7
YOLOv5s - FPN	93.6	91.7	91.7

Table 3. Overall performance in picking region detection

Object detection models	P(%)	R(%)	mAP(%)	Processing time @640×640
YOLOv5s	74.9	83.8	82.6	14.0
YOLOv5s-FPN	87.9	74.0	85.6	10.4

The main source of errors in strawberry and picking region detection using YOLOv5s-FPN was the occlusion of objects of interest by leaves or other parts in the canopies, which influenced the performance of the model. Besides, the color and shape of dead leaf are similar to the mature strawberry which could also result in incorrect detection. There have been a few studies in the past using deep learning approaches (e.g., YOLOv4, Mask-RCNN), which showed similar results in strawberry detection. It might be difficult, however, to compare the performance of previous methods (Table 4) with ours because of different datasets used. There are only few studies conducted to find picking region based on the color or shape of the strawberries and stems. In table 4, the performance of different methods used for strawberry and picking point detection are listed. Huang et al. (2017) investigated picking points based on color and shape with accuracy of 84%. Yu et al. (2019, 2020) applied image processing methods based on shape and color to locate picking points after detecting strawberry using deep learning approaches. The results (Yu et al., 2020) showed an identification rate of 84.35% in finding picking points. The image processing methods to find picking points might be influenced by varying lighting conditions. In contrast, 500 picking region images with different lighting conditions were used to train YOLOv5s-FPN, which resulted in robust detection performance despite the variation in outdoor lighting conditions. Although the performance might be influenced from the computational environments, in this study, the proposed method took only 34.5 ms to execute the entire process of strawberry and picking region detection, which is faster than other methods listed in table 4.

Table 4. performance comparison of four different methods

Authors	Methods	Performance	Computational Environment	Speed (image resolution)
Huang et al., 2017	Color and shape	Accuracy (picking points): 84%	No report	No report
Yu et. al 2019	Mask-RCNN (strawberry) + color/shape (picking region)	Recognition rate(strawberry): 98.41% No report in picking region	CPU: i7-8700k GPU: NVIDIA 1080	130ms (640×480) in strawberry detection
Yu et al., 2020	YOLO3 based model(strawberry) + color/shape (picking region)	mAP (strawberry): 94.3% Identification rate (picking point): 84.35%	CPU: i7-8700k GPU: NVIDIA 1080	55ms (640×480) in strawberry detection
Ours	YOLOv5s-FPN	mAP (strawberry): 92.5% mAP (picking region): 82.8%	CPU: i7-1180H GPU: NVIDA 3070	34.5ms (640×640) including strawberry detection and picking region detection

Conclusion and Future Work

Picking region detection is challenging for object detection methods since there can be many areas with same or similar features in the canopy causing errors during detection. In this study, YOLOv5s-FPN was used for real-time strawberry and picking region detection in corresponding stems to assist the robotic strawberry harvesting. YOLOv5s-FPN was used to detect strawberries first and then detect corresponding picking regions. With only top-down path, FPN reduced the network size (85.7% of the original YOLOv5s) and resulted in faster detection speed (83.9% of the original). This study showed that a simplified structure based on YOLOv5s has high potential to support robotic strawberry harvest with improved speed and performance.

For both strawberry canopy and picking region datasets, more images could be added at varying lighting conditions in the future to increase the robustness of the object detection model. Data augmentation on training and testing datasets, and field evaluation of the model will also be included in the future work. Furthermore, point cloud data could be used for providing more accurate information on picking region of the strawberry to robotic harvesting systems.

Acknowledgements

This research was supported in part by the National Science Foundation (NSF; award# 1924640), and Washington State University (WSU). The first author of the work was also supported by the China Scholarship Council (CSC). Any options, findings, and conclusions expressed in this publication are those of authors and do not reflect any view from NSF, WSU, or CSC.

References

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., & He, Y. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sensing, 11*, 1584.

Delbridge, T. (2021). Robotic strawberry harvest is promising but will need improved technology and higher wages to be economically viable. *California Agriculture, 75*.

He, Z., Karkee, M., & Upadhayay, P. (2021). Detection of strawberries with varying maturity levels for robotic harvesting using YOLOv4. *2021 ASABE Annual International Virtual Meeting*, (p. 1).

Huang, Z., Wane, S., & Parsons, S. (2017). Towards automated strawberry harvesting: Identifying the picking point. *Annual Conference Towards Autonomous Robotic Systems*, (pp. 222–236).

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., . . . Minh, M. T. (2022, February). ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference. *ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference*. Zenodo. doi:10.5281/zenodo.6222936

Lamb, N., & Chuah, M. C. (2018). A strawberry detection system using convolutional neural networks. *2018 IEEE International Conference on Big Data (Big Data)*, (pp. 2515–2520).

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 2117–2125).

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 8759–8768).

Lu, S., Chen, W., Zhang, X., & Karkee, M. (2022). Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. *Computers and Electronics in Agriculture, 193*, 106696.

Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 7263–7271).

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

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

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*, 1619.

Yu, Y., Zhang, K., Liu, H., Yang, L., & Zhang, D. (2020). Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot. *IEEE Access, 8*, 116556–116568.

Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural

environment based on Mask-RCNN. *Computers and Electronics in Agriculture, 163*, 104846. doi:https://doi.org/10.1016/j.compag.2019.06.001

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055*.