Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 14th International Scientific and Practical Conference. Volume 1, 81-85

Rapid Prototyping of Pear Detection Neural Network with YOLO Architecture in Photographs

Sergejs Kodors
Institute of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
sergejs.kodors@rta.lv

Marks Sondors
Institute of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
ms20094@edu.rta.lv

Gunārs Lācis
Institute of Horticulture (LatHort)
Dobele, Latvia
gunars.lacis@llu.lv

Edgars Rubauskis
Institute of Horticulture (LatHort)
Dobele, Latvia
edgars.rubauskis@llu.lv

Ilmārs Apeināns
Institute of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
ilmars.apeinans@rta.lv

Imants Zarembo
Institute of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
imants.zarembo@rta.lv

***Abstract*. Fruit yield estimation and forecasting are essential processes for data-based decision-making in agribusiness to optimise fruit-growing and marketing operations. The yield forecasting is based on the application of historical data, which was collected in the result of periodic yield estimation. Meanwhile, the object detection methods and regression models are applied to calculate yield per tree. The application of powerful neural network architectures for rapid prototyping is a common approach of modern artificial intelligence engineering. Meanwhile, the most popular object detection solution is YOLO architecture. Our project team collected the dataset of fruiting pear tree photographs (Pear640) and trained YOLOv5m with mAP@0.5 95% and mAP@0.5:0.95 56%. The obtained results were compared with other YOLOv5-7.0 and YOLOv7 models and similar studies.**

Keywords: artificial intelligence, deep learning, smart horticulture, yield estimation.

INTRODUCTION

Pears are the third most economically important fruit crop globally [1], [2], reaching 25.7 million tons in 2021 [3]. Although it is not the most important fruit crop in Latvia, it forms a very important niche product with high added value, and the area of pear growing is about 200 ha [4]. Pear growing faces various challenges beyond our control: environmental conditions and their changes and biotic factors such as diseases and pests. The grower's options are choosing appropriate cultivars, adjusting cultivation technologies to reduce environmental impact, and obtaining the optimal yield. One of the solutions is the development of a yield model, in which by changing the parameters of the yield-forming components and the applied agrotechnical practice, it is possible to predict the potential yield and evaluate how a change in a specific agrotechnical technique could affect it. Considering pear orchards' longevity and induced effects' time lag, such a forecasting system can significantly assist growers in their decision-making process. Timely and accurate prediction of fruit yield is also of great economic importance to optimally plan post-harvest activities, storage facilities and sales. Developing such yield forecasting systems has been going on for a long time for various fruit plant species [5]. For example, for apples, the 'Bavendorf' yield forecast model is still recognized as the best one, which is based on such parameters as the characteristics of the analyzed trees (cultivar, rootstock, orchard age), orchard characteristics (slope, elevation and area), the fruit-set density in the given year and the average fruit mass at a harvesting time [6]. However, the disadvantage of all these forecasting systems is the need for high-quality, large-scale data because the accuracy of the developed model and the correspondence of the forecasted and real harvested yield depend on it [7]. Manual collection of such data is time-consuming and complex and may be affected by the subjectivity of the evaluator, which in turn may affect the

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2023vol1.7293

usefulness of the forecasting system. Therefore, tools that provide accurate and automated evaluation of characteristics involved in pear fruit development should be developed, including through pear imagining.

YOLO ("You Only Look Once") is a well known and popular object detection architecture, which was firstly presented in 2016 [8]. The advantage of YOLO is its unified architecture with real-time image processing speed. The previous solutions were based on a sliding-window approach or the region proposals.

YOLO is a complex system, which can have minor and major improvements. At this moment, there are many versions of YOLO architecture, which are developed by different authors. For example, YOLOv4 is based on CSPDarknet53 backbone, Spatial pyramid pooling (SPP) additional module, PANet path-aggregation neck, and YOLOv3 (anchor based) head [9]. YOLOv5 is relatively similar to YOLOv4, the focus layer can be mentioned as the difference [10]. Meanwhile, YOLOv5 has technological advantages such as PyTorch framework, code readability, easy configuration of environment and other user-friendly things [11]. Speaking about YOLOR, it was based on the new concept with a unified model, which connected explicit and implicit knowledge [12]. At the beginning of 2023, YOLOv7 [13] and YOLOv8 [14] can be mentioned as the youngest architectures. YOLOv7 presented Extended Efficient Layer Aggregation Network (E-ELAN). Meanwhile, the anchor-free model is presented in YOLOv8 for performance and accuracy improvement.

YOLO was already applied for yield estimation tasks. For example, Wnag et al. (2022) obtained next accuracy results for yield estimation of litchi fruits (mAP): YOLOv4 - 82.87%, YOLOv5s - 88.9%, YOLOv5 improved by them - 92.4% [15]. Meanwhile, Lyu et al. (2022) experimented with yield estimation of green citrus (mAP@0.5): YOLOv5 97.51% and improved YOLOv5-CS 98.23% [16]. One more example, banana detection solution was presented by Fu et al. (2022): YOLOv3 - 93% mAP and YOLOv4 93.69% mAP [17]. However, despite the existing experiments, there is a restricted number of studied cultivars that is mainly related to the limited number of datasets or their open access availability. Another knowledge gap is the most suitable YOLO model for rapid prototyping of yield estimation solutions, because each YOLO architecture traditionally provides different models with different size and accuracy (Pareto front).

Our study proposes two original things in this article: 1) open dataset with natural images of pear trees in the fruiting stage, which we called Pear640; 2) the accuracy comparison of YOLOv5-7.0 and YOLOv7 templates for the rapid prototyping of yield estimation models.

Additionally to our pear dataset Pear640, two open datasets were selected for experiment: grape dataset "WGISD" [18] and apple dataset "MinneApple" [19]. Considering to the YOLO architectures: YOLOv5n, YOLOv5m, YOLOv5l, YOLOv7 and YOLOv7-X models were selected.

The YOLOv5m was the most suitable model (trade-off solution) for rapid development in our experiment. YOLOv5m achieved accuracy mAP@0.5 95% and mAP@0.5:0.95 57%. YOLOv7 showed the worst result in the cases of Pear640 and MinneApple, but YOLOv5n showed significantly smaller accuracy in the case of WGISD. Meanwhile, the accuracies of YOLOv5m and YOLOv5l were relatively similar, therefore YOLOv5m was more suitable due to its smaller size (latency).

I. MATERIALS AND METHODS

A. *Pear640 Collection and Annotation*

Digital images of pear fruits were collected in the experimental site of the Institute of Horticulture (LatHort) with cultivars 'Suvenirs' and 'Mramornaya' on seedling rootstocks 'Kazraushu' with planting distances 4×5 m (500 trees per 1 ha). (Krimūnu parish, Dobeles district: 56.610169, 23.305956). Collection of fruit images of 'Suvenirs' and 'Mramornaya' was done at the end of August (105 days after full bloom) prior to the harvest.

The collection of digital images was carried out using a digital photo camera Nikon D40 (Image size: 3008×2000; 6.0 MP).

The collection of images was carried out in field conditions, in the orchard at the distance from the tree planting point 2.5 m (middle of alleyway). The whole canopy of trees was photographed as separate objects. The images were taken in front of the tree (tree trunk, planting point), perpendicularly the tree row from the west side of rows (the rows of pear trees oriented from north to south) around noon (11:00–13:00) at clear sky natural light conditions.

The dataset annotation process was performed manually using MakeSense annotation tool. The annotations consisted of bounding boxes around a pear in the photographs, indicating the location of the pears within the image. Then the annotated images 3008x2000 were automatically cropped out on 640x640 images with overlap 30% and validated manually, because YOLOv5 and YOLOv7 work with input size 640x640, but image reduction was not possible due to small bounding boxes, which could achieve size until 25x25. Once all the images were annotated, they were stored in a YOLO format. The result dataset is available in Kaggle repository under CC-BY license [20].

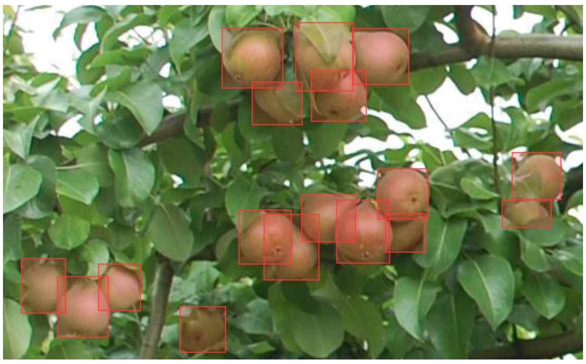

Fig. 1. Pear640 image example.

B. Comparison Datasets

WGISD dataset [18] is a dataset consisting of 300 images containing 4432 grape objects identified by bounding boxes (see Fig. 2). The dataset was used for several reasons. Firstly, it uses a format of annotation that is easy to adapt to the YOLO algorithms. Additionally, the dataset is diverse in terms of variety. The dataset consists of images taken of different grape varieties in different weather conditions. The images were not cropped, because objects were sufficiently large.

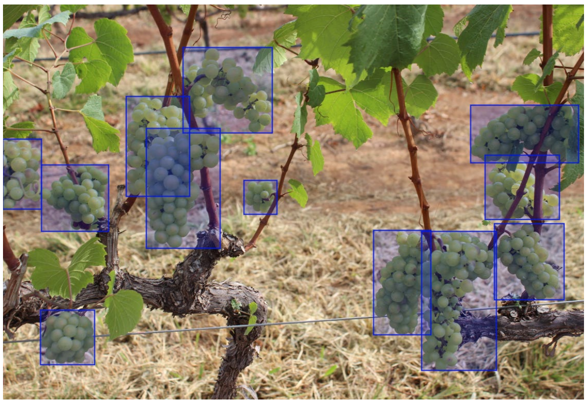

Fig. 2. WGISD image example.

MinneApple dataset [19] contains over 1000 images with over 41000 labelled instances of apples (see Fig. 3). This dataset was used for comparison for several reasons. Firstly, it contains a large number of images with also a lot of annotations. Additionally, the apple dataset is much more similar to pears than grapes. Images were cropped on 640x640 due to small objects. The annotation was not improved to save possibility to compare results with other experiments.

Experiment Design

In this experiment, we utilized YOLOv5 [21] and YOLOv7 [22] models, specifically YOLOv5n, YOLOv5m, YOLOv5l, YOLOv7 and YOLOv7-X.

The experiment was conducted on an NVIDIA RTX 2070 GPU, which provided sufficient performance for training and testing the models.

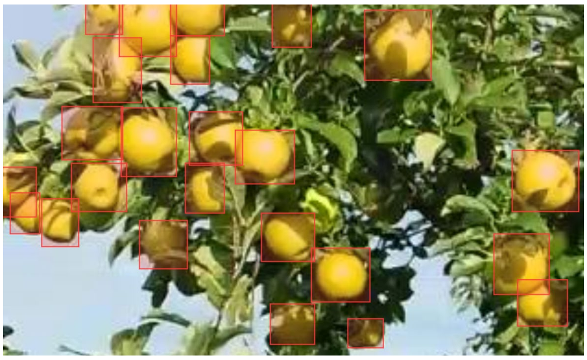

Fig. 3. MinneApple image example.

Pear640 dataset was randomly distributed across training, validation and test folders using the Python random shuffle method. This was repeated 5 times to generate five distinctive data splits, ensuring that the images in each data split contained the same images but with varying locations across the three folders. The data splits were 70% of the images would be in the train folder, 20% of the images would be in the validation folder and 10% images would be in the test folder. This was also then repeated for WGISD and MinneApple datasets respectively.

The augmentation was applied in the experiment: we modified scaling, mix-up and shearing to 0, rotation parameter to 30º and increased mosaic to 1, which takes multiple images and combines them into one.

For each data split, YOLOv5 and YOLOv7 models were trained independently, resulting in five trained models for one specific model type. Subsequently, the trained models were then tested with the test images on their respective data splits to evaluate their performance and ensure an extensive comparison.

RESULTS AND DISCUSSION

If results are analysed in the scope of one dataset, the type of YOLO model can increase accuracy until 8.2% (Tab. 1, median). In the case of our dataset (Pear640), it was not so impressive - only 4.1% (Tab. 3, median). The analysis of YOLO models among different datasets showed the more interesting distribution of obtained accuracies (see Tab. 2-3, Fig. 4). YOLOv7 showed the worst result in the cases of Pear640 and MinneApple, but YOLOv5n showed significantly smaller accuracy in the case of WGISD. Meanwhile, the accuracies of YOLOv5m and YOLOv5l were relatively similar, but YOLOv5m has smaller size (latency). Therefore, YOLOv5m was the most suitable model (trade-off solution) for the rapid development in our experiment.

YOLOv5m showed next accuracy: mAP@0.5 95% and mAP@0.5:0.95 56%, precision 93%, recall 90%.

Analysing results with WGISD, it can be mentioned, that Santos et al. (2020) trained Mask R-CNN with accuracy 71.9% mAP@0.5 [18]. Thomas et al. (2023) completed similar study applying YOLOv5 models for WGISD object detection. They obtained sufficiently similar results: YOLOv5n - 89.4%, YOLOv5m - 89.5%, YOLOv5l - 90.5% [23].

Considering to MinneApple dataset, its authors proposed achieved accuracy equal to 77.5% mAP@0.5 by using Mask RCNN method [19]. Meili et al. (2022) presents BFP Net model, which provides 84.6% mAP@0.5 accuracy [24]. Meanwhile, Li et al. (2021) compared exactly YOLO models: YOLOv4 CspdarkNet53 - 90.53% and YOLOv5s - 80.11% mAP@0.5 [25]. In our experiment, the better results were obtained for all YOLOv5 models, but YOLOv7 provided close results. The better results can be explained by the image crop on 640x640, that was intuitive for our team. At the same time, other authors mentioned problems with small objects.

Experiments with pear detection were completed by other authors too. Sun at al. (2023) proposed the modified

YOLOv5 model called YOLO-P, which was obtained by completing redevelopment of the backbone part for orchard picking robots. YOLO-P achieved 97.6% mAP@0.5 and 39.4% volume improvement and was tested on pear dataset [26]. Li et al. (2022) presented another modified YOLOv5 model called YOLOv5s-FP, which was tested on a pear dataset. It achieved 96.12% mAP@0.5. The modification was oriented to increase image processing speed [27].

Summarising, the yield detection accuracy 90% mAP@0.5 is a relatively good achievement at this moment, which can be obtained in the rapid development stage. Considering to our experiment, YOLOv5m is preferable model for the rapid development, because YOLOv7 was unstable, YOLOv5l provided similar results to YOLOv5m, but YOLOv5n in the case of WGISD showed accuracy smaller than 80%. Speaking about studies related to yield monitoring, artificial intelligence engineers try to optimise the backbone of YOLO architecture to minimise latency for edge solutions (unmanned aerial vehicles, unmanned ground vehicles, fruit pickers, mobiles, etc.). However, it may be more suitable to simply retrain YOLO backbone on the huge dataset of rural content domain comparable with ImageNet and COCO collections. That is challenging at this moment, because there are too few public agriculture datasets as well as selected categories must be well planned.

CONCLUSIONS

In this article we presented our public dataset Pear640, which is available in Kaggle under CC-BY licence.

Completing the pilot experiments directed to develop yield estimation solutions, we wanted to identify the suitable architecture and model for the rapid development of fruit detection. Our pilot experiment showed that YOLOv5m is a preferable model for the rapid development of yield estimation solutions. The best trained YOLOv5m model showed the following results for the Pear640 dataset: mAP@0.5 95%, mAP@0.5:0.95 56%, precision 93%, recall 90%.

TABLE 1 EXPERIMENT RESULTS WITH WGISD

YOLO	**Test Dataset WGISD (mAP@0.5)**				
	v5n	***v5m***	***v5l***	***v7***	***v7-X***
min	0.799	0.833	0.842	0.880	0.861
mean	0.831	0.877	0.879	0.899	0.902
median	0.825	0.890	0.889	0.888	0.907
max	0.881	0.918	0.930	0.933	0.932

TABLE 2 EXPERIMENT RESULTS WITH MINNEAPPLE

YOLO	**Test Dataset MinneApple (mAP@0.5)**				
	v5n	***v5m***	***v5l***	***v7***	***v7-X***
min	0.885	0.894	0.894	0.802	0.812
mean	0.891	0.903	0.905	0.881	0.865
median	0.890	0.904	0.907	0.896	0.883
max	0.896	0.909	0.914	0.912	0.914

TABLE 3 EXPERIMENT RESULTS WITH PEAR640

YOLO	**Test Dataset Pear640 (mAP@0.5)**				
	v5n	***v5m***	***v5l***	***v7***	***v7-X***
min	0.927	0.932	0.936	0.834	0.737
mean	0.935	0.941	0.940	0.915	0.874
median	0.938	0.940	0.940	0.928	0.897
max	0.943	0.951	0.942	0.955	0.945

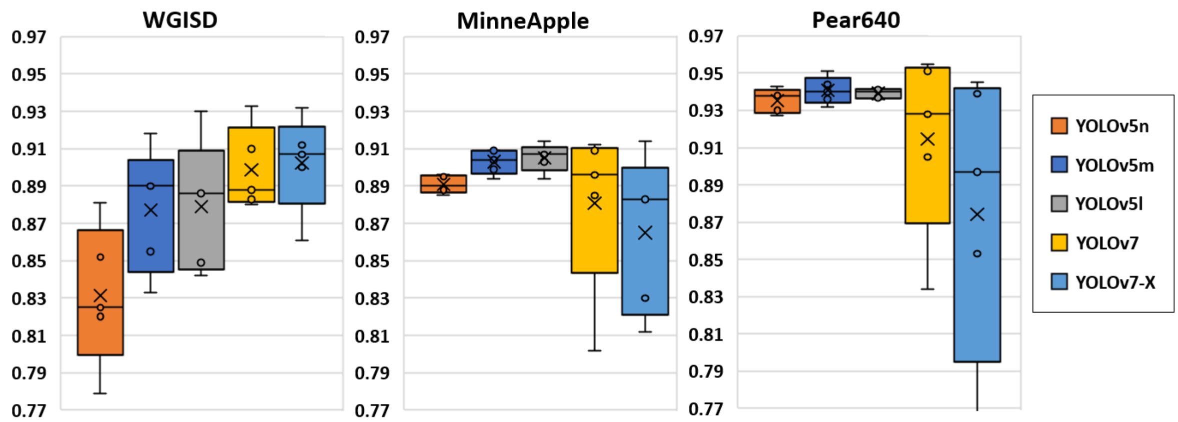

Fig. 4. Box-plot diagram of YOLO model accuracy (mAP@0.5).

ACKNOWLEDGEMENT

This research is funded by the Latvian Council of Science, project "Development of autonomous unmanned aerial vehicles based decision-making system for smart fruit growing", project No. lzp-2021/1-0134.

REFERENCES

[1] G. J. Silva, T.M. Souza, R.L. Barbieri and A.C. de Oliveira, "Origin, domestication, and dispersing of pear (Pyrus spp.)," Adv. Agricult., 2014. Available: https://doi.org/10.1155/2014/541097

[2] A. Zarei, J. Erfani-Moghadam and H. Jalilian, "Assessment of variability within and among four Pyrus species using multivariate analysis," Flora, vol. 250, pp. 27-36, 2019. Available: https://doi.org/10.1016/j.flora.2018.11.016

[3] FAOSTAT, *Food and agriculture data*. [Online] Available: http://www.fao.org/faostat/[Accessed March 21, 2023].

[4] G. Lācis, I. Kota-Dombrovska, K. Kārkliņa, B. Lāce, "Genetic diversity and relatedness of Latvian Pyrus germplasm assessed by a set of SSR markers," Proceedings of the Latvian Academy of Sciences. Section B, vol. 76, no. 4 (739), pp. 438–447, 2022. Available: https://doi.org/10.2478/prolas-2022-0068

[5] H. Cheng, L. Damerow, Y. Sun and M. Blanke, "Early Yield Prediction Using Image Analysis of Apple Fruit and Tree Canopy Features with Neural Networks," *J. Imaging*, vol. 3, no. 1, p. 6+, Jan. 2017. Available: https://doi.org/10.3390/jimaging3010006

[6] F. Winter, "Modelling the biological and economic development of an apple orchard," in *Proceedings of the III International Symposium on Research and Development on Orchard and Plantation Systems*, *Montpellier, May 21-26, 1984*, *France*. Place of publication: Acta Hortic, 1986. Available: https://doi.org/10.17660/ActaHortic.1986.160.40

[7] R. Črtomir, C. Urška, T. Stanislav, S. Denis, P. Karmen, M. Pavlovič, V. Marjan, "Application of Neural Networks and Image Visualisation for Early Forecast of Apple Yield," Erwerbs-Obstbau, vol. 54, pp. 69–76, 2012. Available: https://doi.org/10.1007/s10341-012-0162-y

[8] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA. Place of publication: IEEE, 2016. Available: https://doi.org/10.1109/CVPR.2016.91

[9] A. Bochkovskiy, C.-Y. Wang, H.-Y. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv, Apr 23, 2020. [Online] Available: https://arxiv.org/pdf/2004.10934.pdf [Accessed: Mar 17, 2023]

[10] U. Nepal, H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs," *Sensors*, vol. 22, no. 2, p. 464, Jan. 2022, Available: https://doi.org/10.3390/s22020464

[11] P. Jiang, D. Ergu, F. Liu, Y. Cai, B. Ma, "A Review of Yolo Algorithm Developments," Procedia Computer Science, vol. 199, pp. 1066-1073, 2022. Available: https://doi.org/10.1016/j.procs.2022.01.135

[12] C.-Y. Wang, I.-H. Yeh, H.-Y. M. Liao, "You only learn one representation: Unified network for multiple tasks," arXiv, May 10, 2021. [Online] Available: https://arxiv.org/pdf/2105.04206.pdf

[13] C.-Y. Wang, A. Bochkovskiy, H.-Y. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," arXiv, Jul 6, 2022. Available: https://arxiv.org/pdf/2207.02696.pdf

[14] J. Solawetz, "What is YOLOv8? The Ultimate Guide," Jan 11, 2023. [Online] Available: https://blog.roboflow.com/whats-new-in-yolov8/ [Accessed: Mar 17, 2023]

[15] L. Wang, Y. Zhao, Z. Xiong, S. Wang, Y. Li, Y. Lan, "Fast and precise detection of litchi fruits for yield estimation based on the improved YOLOv5 model," Frontiers in Plant Science, vol. 13, 2022. Available: https://doi.org/10.3389/fpls.2022.965425

[16] S. Lyu, R. Li, Y. Zhao, Z. Li, R. Fan, S. Liu, "Green Citrus Detection and Counting in Orchards Based on YOLOv5-CS and AI Edge System," Sensors, vol. 22, no. 2, p. 576+, 2022. Available: https://doi.org/10.3390/s22020576

[17] L. Fu, Z. Yang, F. Wu, X. Zou, J. Lin, Y. Cao, J. Duan, "YOLO-Banana: A Lightweight Neural Network for Rapid Detection of Banana Bunches and Stalks in the Natural Environment," Agronomy, vol. 12, no. 2, p. 391+, 2022. Available: https://doi.org/10.3390/agronomy12020391

[18] T.T. Santos, L.L. de Souza, A.A. dos Santos, S. Avila, "Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association," Computers and Electronics in Agriculture, vol. 170, 2020. Available: https://doi.org/10.1016/j.compag.2020.105247

[19] N. Häni, P. Roy and V. Isler, "MinneApple: A Benchmark Dataset for Apple Detection and Segmentation," in *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 852-858, April 2020. Available: https://doi.org/10.1109/LRA.2020.2965061

[20] lzp-2021/1-0134, *Pear640*, 2023. [Online]. Available: https://www.kaggle.com/datasets/projectlzp201910094/pear640 [Accessed: Mar 24, 2023]

[21] Ultralytics, *YOLOv5-7.0 GitHub repository*. [Online] Available: https://github.com/ultralytics/yolov5 [Accessed: Jan 30, 2023]

[22] *YOLOv7 GitHub repository*. [Online] Available: https://github.com/WongKinYiu/yolov7 [Accessed: Feb 20, 2023]

[23] T.A. Ciarfuglia, I.M. Motoi, L. Saraceni, M. Fawakherji, A. Sanfeliu, D. Nardi, "Weakly and semi-supervised detection, segmentation and tracking of table grapes with limited and noisy data, Computers and Electronics in Agriculture," vol. 205, 2023. Available: https://doi.org/10.1016/j.compag.2023.107624

[24] M. Sun, L. Xu, X. Chen, Z. Ji, Y. Zheng, W. Jia, "BFP Net: Balanced Feature Pyramid Network for Small Apple Detection in Complex Orchard Environment," Plant Phenomics, vol. 2022, 2022. Available: https://doi.org/10.34133/2022/9892464

[25] Y. Li, C. Lei, Z. Xue, Z. Zheng, Y. Long, "A Comparison of YOLO Family for Apple Detection and Counting in Orchards," World Academy of Science, Engineering and Technology, International Journal of Computer and Systems Engineering, vol. 15, no. 5, pp. 334 – 343, 2021. Available: https://publications.waset.org/10012056/a-comparison-of-yolo-family-for-apple-detection-and-counting-in-orchards

[26] H. Sun, B. Wang, J. Xue, "YOLO-P: An efficient method for pear fast detection in complex orchard picking environment," Frontiers in Plant Science, vol. 13, 2023. Available: https://doi.org/10.3389/fpls.2022.1089454

[27] Y. Li, Y. Rao, X. Jin, Z. Jiang, Y. Wang, T. Wang, F. Wang, Q. Luo, L. Liu, "YOLOv5s-FP: A Novel Method for In-Field Pear Detection Using a Transformer Encoder and Multi-Scale Collaboration Perception," *Sensors*, vol. 23, no. 1, p. 30+, Dec. 2022, Available: https://doi.org/10.3390/s23010030