Fruit Detection and Counting for Yield Analysis in Digital Agriculture

Sornalakshmi K[a], Sayan Majumder and Yash Khandelwal
Department of Data Science and Business Systems,
Faculty of Engineering and Technology,
SRM Institute of Science and Technology, Kattankulathur Campus, 600023, India

Keywords: Fruit Detection and Counting, Computer Vision, You Only Look Once (YOLO).

Abstract: For the cause of evolution of agriculture to its next stage, Artificial Intelligence and Data driven approaches will play a major role in the development of agricultural practices that as per our vision would offer numerous economic, environmental and social benefits. Digital/Precision Agriculture is providing more benefits since the state-of-the-art ICT tools are used for better decision-making process. The other benefits include enhanced productivity in yield, reduced environmental footprints and better resource management. Our solution uses the adoption of Computer Vision and real time monitoring of plants, studying their respective conditions and their autonomous cultivation and harvesting patterns. The proposed system uses YOLO v8 algorithm for the detection of fruits from the Kaggle fruit detection data set and Mango YOLO dataset for four different fruits and returning the count of fruits in the image. The fruits were detected and counted from the images of the respective trees having various other parts like branches, leaves and flowers. Also the images from two data sets were combined to create four classes of fruits. The proposed system uses YOLOv8 and YOLO-NAS for detection and counting. Our results recorded an average confidence score of 92% for fruit detection and recall score of 0.97 for counting considering situations like un-ripe fruit and overlapping of objects. Our model was able to successfully count the accurate number of fruits in the test images with critically overlapping fruit counts in a test environment with a Tesla T4 GPU.

1 INTRODUCTION

Precision agriculture or Digital agriculture is growing in countries like India to increase the food supply to match the growing food demand. The different support systems in digital agriculture provides information required by farmer for timely decision making. Growing plants or crops in controlled environments like poly houses or green houses is also gaining popularity because of the ability to predict the growth and control various environmental parameters. Computer vision is one major tool that could be used in digital agriculture for many activities like plant growth, disease and pest damage surveillance using a variety of images like RGB images, hyper spectral images and aerial surveillance images. The ability of an expert system to identify the type of disease in a crop or growth stage in a crop are done with the help of computer vision techniques like image classification, object detection and segmentation. Many recent works have summarized the challenges in applying the computer vision problems in literature to practical scenarios and the respective future directions(Xiao et al., 2023). In this work, we combine three data sets with lot of background information like branches, leaves and create four classes of images – apples, bananas, oranges and mangoes. We apply the YOLO v8 classifier to detect the fruits and count the fruits in test images.

2 RELATED WORK

The authors in (Mishra et al., 2013) have detected and counted the gerbera flowers from the images in polyhouse. The flower and background regions are segmented from the images. The flowers are defined

[a] https://orcid.org/0000-0002-3579-3384

K., S., Majumder, S. and Khandelwal, Y.
Fruit Detection and Counting for Yield Analysis in Digital Agriculture.
DOI: 10.5220/0012881200004519

In *Proceedings of the 1st International Conference on Emerging Innovations for Sustainable Agriculture (ICEISA 2024)*, pages 50-57
ISBN: 978-989-758-714-6

using the HSV (Hue Saturation Value) color space techniques. The flower was then extracted using thresholding techniques. The other work that detects and counts fruits are discussed in Table 1.

The work in (Dorj et al., 2017) detected and counted the citrus fruits in an orchard. The RGB images are converted to HSV, apply thresholding and noise removal. For counting the fruits, the overlapping fruits were counted using watershed segmentation. The work in (Wan Nurazwin Syazwani et al., 2022) uses UAV (Unmanned Aerial Vehicle) top view images to detect and count pine apple crowns in a field. The images were preprocessed, segmented and the extracted features were then analyzed and matched to the features of pineapple to detect. The authors in (Turečková et al., 2022) use a 360 degree video form a polyhouse to acquire images of tomato plants. The image frames from the video are then processed under different resolution categories to inspect the performance of detection and counting. Since the frames are from a video image stitching metrics are also compared. The video based image frames of vertical wall fruiting method of apples are processed in (Li et al., 2023). The trunk and fruits are also detected. The displacements of references between consecutive video frames are used in predicting fruit positions and unique ids are assigned to fruits to avoid duplicate counting. A light weight object detection framework (Zeng et al., 2023) uses MobileVnet's module for the backbone network to avoid the requirement of heavy computational resources. The model is embedded on an app interface to. In (Mamat et al., 2023), the authors use computer vision, the various versions of YOLO to classify and auto annotate the ripening stages of the oil palm fruit images. The authors use multi-scale fusion and reuse at neck level for a light weight architecture to collect small target features and discard redundant features in far off small apples (Ma et al., 2023). The work in (Zheng et al., 2023) captures remote sensing images, converts it into a one orthomosaic tiff image. This image is fed into a Faster R-CNN network, with a ResNet-50 feature extraction back bone. The algorithm classifies images into four classes namely – ripe fruit, unripe fruit, flower and background classes. Multi-view duplicate removal was done using an improved FaceNet model to learn the geographical position of the strawberry. Later clustering is applied to remove duplicates and count the strawberries. In (da Silva et al., 2023), the authors aim at analyzing and providing computer vision AI solutions in the edge devices like mobile phones with limited connectivity and computational power. In detection YOLO was performing faster and in the classification task MobileNetV2 was performing better. In the recent work (Zhong et al., 2024), a light weight YOLO based on having skip and bidirectional connection module using the DarkNet53 architecture. So analyzing the state-of- the art work in fruit detection and counting so far, we have contributed the following i) Combining three different data sets to create a reference data set with images having significant background noise ii) Apply YOLO v8 on the combined data sets iii) Apply YOLO NAS on the data set.

Table 1 – Summary of recent research in fruit detection and counting.

Reference	Data Set	Fruit	Algorithm Used	Accuracy (%)	Image Type	Image Count	Augmentation	Resolution
(Dorj et al., 2017)	Custom collected	Citrus	HSV, Thresholding, Watershed segmentation	93	RGB	84	None	1824x1028
(Wan Nurazwin Syazwani et al., 2022)	Custom collected	Pine apple	ANN, SVM	94	UAV – RGB	1300	None	2704x1520
(Turečková et al., 2022)	Custom collected	Tomato	Faster R-CNN, ResNet-50	83	360 video	1997	None	Multiple 2448x4078 1469x2448 1333x735

(Li et al., 2023)	Custom collected	Apple	Yolo V4 Tiny	99 detection, 91 in counting	Video	800	None	416x416
(Zeng et al., 2023)	Custom collected	Tomato	Improved light weight YOLO v5	93 true detection rate	RGB	932	None	4032x3024
(Mamat et al., 2023)	Custom collected	Oil Palm	YOLO v3,v4 v5	98 in YOLO v5	RGB	400	Yes.	416x416
(Ma et al., 2023)	MineApple	Apples	Upgraded YOLO v7 Tiny	80.4	RGB	829	Yes	416x416
(Zheng et al., 2023)	Custom collected	Strawberry	Faster R-CNN	Average 97	Remote sensing	2415	None	536x712
(da Silva et al., 2023)	Custom collected	Citrus Fruits	YOLO, MobileNetV2	98	RGB	160	None	-
(Zhong et al., 2024)	ACFR Mango Dataset	Mango	Improved YOLO	96	RGB	1964	None	500x500

3 PROPOSED METHODOLOGY

3.1 Data Set

The data set uses images for three classes apples oranges and bananas from the Kaggle datasets (Tyagi, 2023) and (Kaggle, n.d.). The images for the Mango classes were obtained from (Koirala et al., 2019). The images were converted to 416x416. A total of 6210 images of all classes were used.

3.2 Object Detection

Detection of objects in this task entails pinpointing the position of all objects in the image. Our model is working with the anchor box method. The said process starts with the formation of a number of predefined anchor boxes that delineates the complete input image. Compared to that, each anchor-box undergoes two types of predictions by the network. During the very beginning it infers whether the proposed box has positive reference either one of the specified object classes. The second task performance is also object recognition. For that, the box is annotation. In this stage, the network tries to move and reshape the box to become closer to the ground truth location of the objects to detect.

3.3 Fruit Detection

In Deep Learning for fruit detection processing is based on the following models: object detection and segmentation via SSD, R-CNN, Faster R-CNN with VGG-16 as a backbone, Inception ResNet. We found these models performing remarkably well in estimation results, which is proved through repetition of recent methods.

Besides, *the* reaction speed of neural networks is significant too from the view of their utilization. Secondly, these networks generally do not have scaling capability with large dataflow in terms of volume or time as a requirement for real time monitoring. The YOLOv8 pretrained on the COCO dataset has already learnt to detect and classify features relevant to fruits because COCO dataset has apple and orange instances along with some irrelevant yet expected features.

Because of the fact that we had two approaches – model *centric* and data-centric, within the framework of model centric method we detect the feature contribution of each hidden layer and remove the kernel of convolution that outputs the non-fruit signals and classes. Finally, we get to the successful point where these shared low level features don’t decrease for the other fruit classes. on the other hand, pruning higher layers doesn't affect the detection of the fruit. We took approach of theory that requires using Fruit Detection Dataset from Kaggle and MangoYOLO. more emphasis on these specific classes while creating a new data configuration file (data.yaml) and defining the number of class as 4 being apples, oranges , bananas and mangos. Fine-tuning the model using transfer learning technique from our last model checkpoint gave us pretty good results and a deployable model for low response time and quality output provided the environment was powered by NVidiaTesla T4 GPU.

We repeated the experiments using a groundbreaking object detection *foundational* model YOLO-NAS, developed by Deci AI as a part of their SuperGradient project.

For an overview, YOLO-NAS employs quantization-aware blocks and selective quantization for optimal performance. The model, *when* converted to its INT8 quantized version, experiences a minimal precision drop, a significant improvement over other model. YOLO-NAS is easily available via ultralytics or supergradient package and provides features like sophisticated training and quantization and AutoNAC optimization and pre-training.

3.4 Yield Counting

For counting objects, it is necessary to use the correct pretrained model for our case it was yolov8n.pt. YOLOv8 has been an absolute breakthrough along with YOLO-NAS in the field of modern computer vision tasks like real time detection, monitoring and counting. This technology is offered by a python package "ultralytics", coupled with ByteTrack which is a tool for tracking objects which provides various options such as SORT, DeepSort, FairMOT. It has its own repository available open source, but for our case we would be using its python package. After tracking comes the counter which requires an API named supervision which utilizes the ByteTrack to track the objects and simultaneously count, these two application components would run autonomously where Supervision will have a dependency on ByteTrack. In case of real time counting, using these technologies would be our recommendation, from a still image yolov8 should suffice the task due to the selected model's inbuilt feature that displays the object count in the terminal. The object detection process flow using YOLO is given in Figure 1.

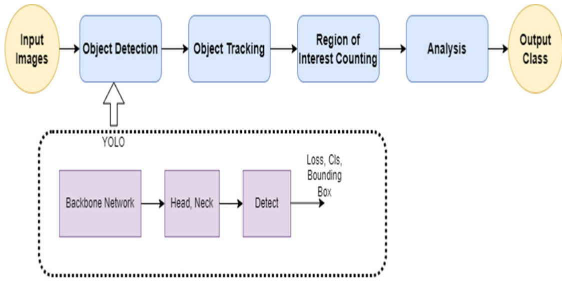


Figure 1: System Design for Detection and Yield Analysis.

4 RESULTS AND DISCUSSION

4.1 Kaggle Fruit Detection Dataset

We trained YOLOv8 with the dataset containing 600 images for 3 categories : apple , banana and orange. The below figures show precisely what we obtained out of it.

Figure 2: Input Image

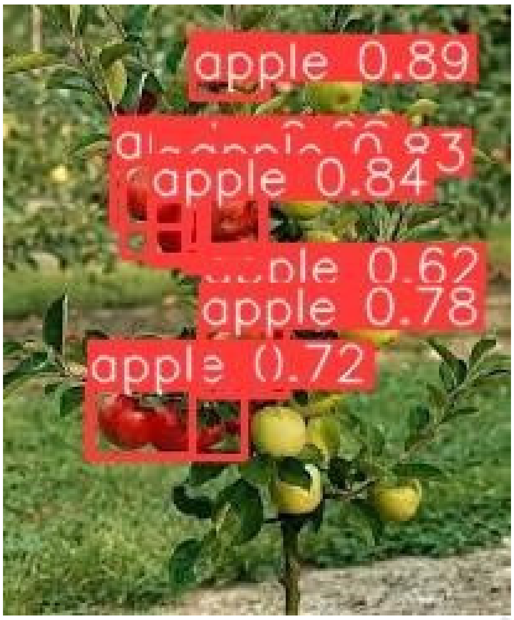


Figure 3: Output Image.

them with the confidence scores associated with each fruit.

For counting *results* YoloV8 automatically counts all the items present in its knowledge boundary. The below figure shows the result output for YOLO v8 which counts the number of fruits in the image as shown in Fig 4.

```
image 1/1 /content/apple_tree.jpeg: 224x192 11 apples, 56.6ms
Speed: 0.7ms preprocess, 56.6ms inference, 2.2ms postprocess per image at shape (1, 3, 224, 192)
Results saved to runs/detect/train3
```

Figure 4: Counting Result of YOLO v8.

The below graphs Fig 5 show how the model improved its knowledge development with each epoch over 50 epochs.

Here's the model evaluation results:

Recall: 0.931, 0.592, 0.646 for each category of fruits.

Mean Average Precision @ 50% object overlap: 0.97, 0.772, 0.573.

Mean Average Precision @ 95% object overlap: 0.796, 0.486, 0.39

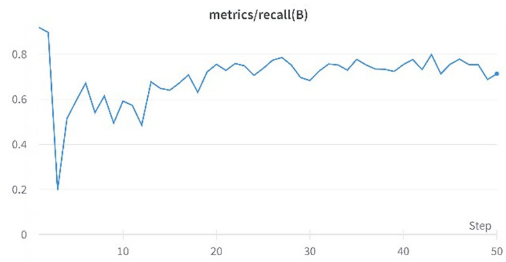


Figure 5a: Recall of YOLO v8.

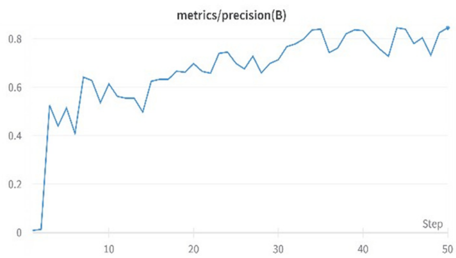


Figure 5b: Recall of YOLO v8.

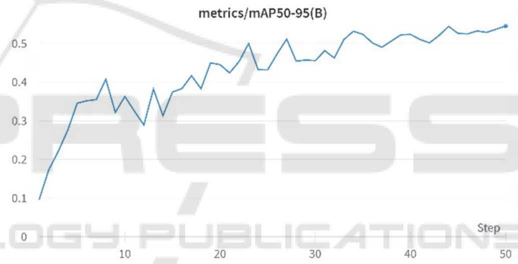
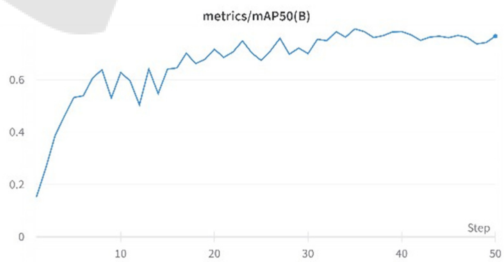


Figure 5c: Mean Average Precision at 95% Object Overlap of YOLO v8.



(d)

Figure 5d: Mean Average Precision at 50% Object Overlap of YOLO v8.

As the next step, we implemented the YOLO NAS algorithm for the fruit detection task on the same data set with three classes. The output of YOLO NAS is given below in Figure 6 with 15 epochs and we get an recall score of 76 percentage maximum. The conclusion is that YOLO NAS requires that data has to be more since the model is more detail oriented and requires higher end GPU for improving the accuracy.

```
Testing: 100%|██████████| 28/28 [00:23<00:00,  1.32it/s]
{'PPYoloELoss/loss_cls': 1.1682425,
 'PPYoloELoss/loss_iou': 0.4165411,
 'PPYoloELoss/loss_dfl': 0.5608495,
 'PPYoloELoss/loss': 2.1456327,
 'Precision@0.50': 0.036405790597200394,
 'Recall@0.50': 0.7687849998474121,
 'mAP@0.50': 0.40779444575309753,
 'F1@0.50': 0.06894689053297043}
Testing: 100%|██████████| 28/28 [00:23<00:00,  1.19it/s]
```

Figure 6: Performance metrics of YOLO NAS.

4.2 Mango YOLO Dataset

The dataset consists of 1730 annotated images of Mango Trees with fruits. The background information such as leaves branches are present in the images. The sample batch of mango detection is shown in the figure below, The data set was trained for 50 epochs, along with the three categories of images in the fruit detection dataset. YOLO v8 was applied for the integrated data set.

The below figure 7 represents the detection output of YOLO v8 on the Mango data set alone.

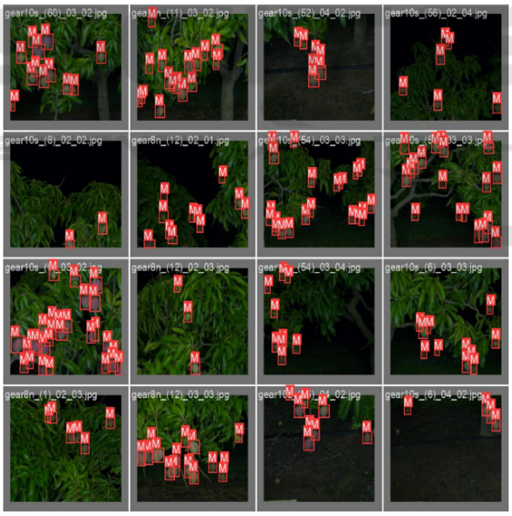


Figure 7: YOLO V8 performance on Mango dataset.

The graphs in Fig 8 are the results obtained for four classes combined as a single data set for four classes – apple, banana, orange from the fruit detection dataset and M (Mango) from the MangoYOLO dataset.

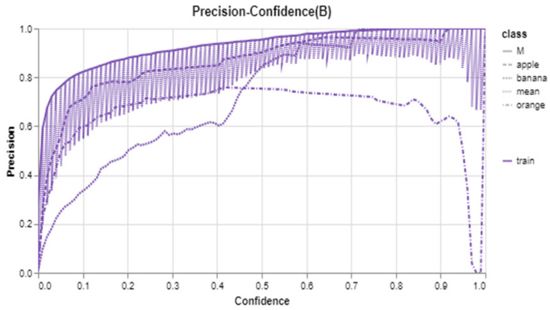


Figure 8a: YOLO V8 Precision vs Confidence Score on combined dataset.

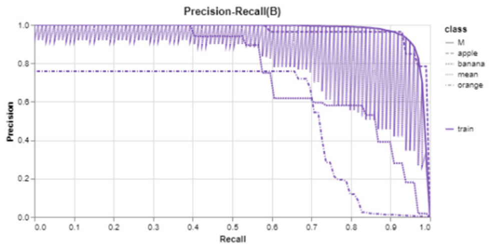


Figure 8b: YOLO V8 Recall vs Confidence Score on combined dataset.

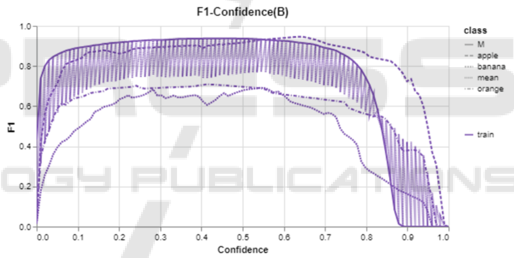


Figure 8c: YOLO V8 Precision vs Recall on combined dataset.

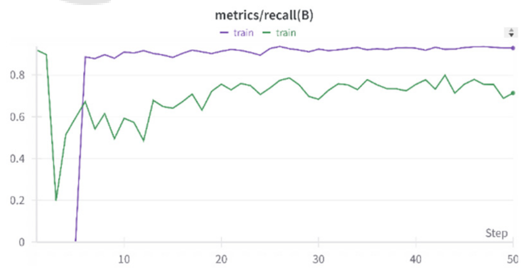


Figure 8d: YOLO V8 F1 Score vs Confidence Score on combined dataset

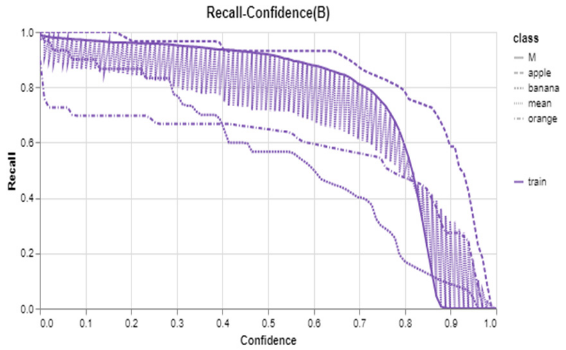


Figure 8e: YOLO V8 Recall for training and test on combined dataset.

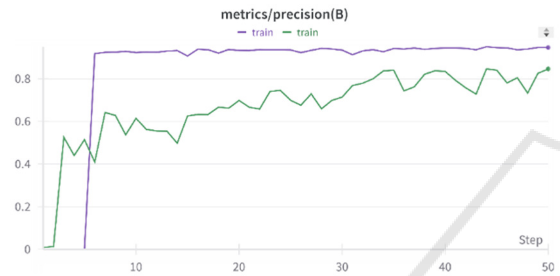


Figure 8f: YOLO V8 Precision for training and test on combined dataset.

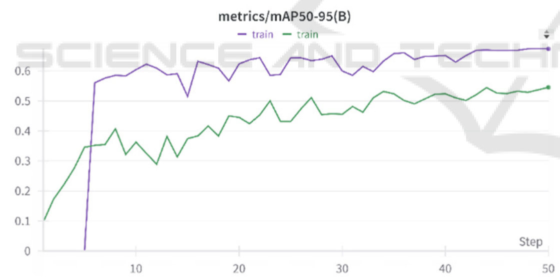


Figure 8g: YOLO V8 Mean Average Precision at 95% Object Overlap for training and test on combined dataset.

5 CONCLUSIONS

In this work, we have applied YOLO v8 to the multiple data sets from Kaggle fruit detection and Mango YOLO with four combined classes and obtained an accuracy of 92%. The we used the latest YOLO NAS, on the same data set to get a performance of 76%. We are able to conclude that though the data set had considerable background noise, the YOLO v8 model was able to detect and count efficiently. We got less performance with YOLO NAS. This could be because of data set size with more details and higher computational resource for more epochs have to be used. Our future work in to apply and improvise YOLO NAS for light weight fruit detection on edge devices.

REFERENCES

da Silva, J. C. F., Silva, M. C., Luz, E. J. S., Delabrida, S., & Oliveira, R. A. R. (2023). Using Mobile Edge AI to Detect and Map Diseases in Citrus Orchards. *Sensors*, *23*(4). doi: 10.3390/s23042165

Dorj, U. O., Lee, M., & Yun, S. seok. (2017). An yield estimation in citrus orchards via fruit detection and counting using image processing. *Computers and Electronics in Agriculture*, *140*, 103–112. doi: 10.1016/j.compag.2017.05.019

Kaggle. (n.d.). *Fruit Images for Object Detection*. Retrieved from https://www.kaggle.com/datasets/mbkinaci/fruit-images-for-object-detection

Koirala, A., Walsh, K., Wang, Z., & McCarthy, C. (2019). *MangoYOLO data set.* Retrieved from https://acquire.cqu.edu.au/articles/dataset/MangoYOLO_data_set/13450661

Li, T., Fang, W., Zhao, G., Gao, F., Wu, Z., Li, R., Fu, L., & Dhupia, J. (2023). An improved binocular localization method for apple based on fruit detection using deep learning. *Information Processing in Agriculture*, *10*(2), 276–287. doi: 10.1016/j.inpa.2021.12.003

Ma, L., Zhao, L., Wang, Z., Zhang, J., & Chen, G. (2023). Detection and Counting of Small Target Apples under Complicated Environments by Using Improved YOLOv7-tiny. *Agronomy*, *13*(5). doi: 10.3390/agronomy13051419

Mamat, N., Othman, M. F., Abdulghafor, R., Alwan, A. A., & Gulzar, Y. (2023). Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. *Sustainability (Switzerland)*, *15*(2). doi: 10.3390/su15020901

Mishra, M. K., Institute of Electrical and Electronics Engineers., Computer Society of India., & GLA University. Department of Computer Engineering & Applications. (2013). *Proceedings of the 2013 International Conference on Information Systems and Computer Networks (ISCON) : March 09 -10, 2013, Mathura, India*. IEEE.

Turečková, A., Tureček, T., Janků, P., Vařacha, P., Šenkeřík, R., Jašek, R., Psota, V., Štěpánek, V., & Komínková Oplatková, Z. (2022). Slicing aided large scale tomato fruit detection and counting in 360-degree video data from a greenhouse. *Measurement: Journal of the International Measurement Confederation*, *204*. doi: 10.1016/j.measurement.2022.111977

Tyagi, L. (2023). *Fruit Detection Dataset*. Kaggle. doi: 10.34740/KAGGLE/DSV/4922010

Wan Nurazwin Syazwani, R., Muhammad Asraf, H., Megat Syahirul Amin, M. A., & Nur Dalila, K. A. (2022). Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alexandria Engineering Journal*, *61*(2), 1265–1276. doi: 10.1016/j.aej.2021.06.053

Xiao, F., Wang, H., Xu, Y., & Zhang, R. (2023). Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review. In Agronomy (Vol. 13, Issue 6). MDPI. doi: 10.3390/agronomy13061625

Zeng, T., Li, S., Song, Q., Zhong, F., & Wei, X. (2023). Lightweight tomato real-time detection method based on improved YOLO and mobile deployment. *Computers and Electronics in Agriculture*, *205*. doi: 10.1016/j.compag.2023.107625

Zheng, C., Liu, T., Abd-Elrahman, A., Whitaker, V. M., & Wilkinson, B. (2023). Object-Detection from Multi-View remote sensing Images: A case study of fruit and flower detection and counting on a central Florida strawberry farm. *International Journal of Applied Earth Observation and Geoinformation*, *123*. doi: 10.1016/j.jag.2023.103457

Zhong, Z., Yun, L., Cheng, F., Chen, Z., & Zhang, C. (2024). Light-YOLO: A Lightweight and Efficient YOLO-Based Deep Learning Model for Mango Detection. *Agriculture (Switzerland)*, *14*(1). doi: 10.3390/agriculture14010140