Estimation of Crop Yield using Deep Learning for Precision Agriculture

Stelios P. Neophytides*[a, b], Giorgia Guerissi[c], Michalis Mavrovouniotis[a,b], Marios Tzouvaras[a,b], Fabio Del Frate[c], Diofantos G. Hadjimitsis[a,b]
[a]ERATOSTHENES Centre of Excellence, Franklin Roosevelt 82, 3012, Limassol, Cyprus.[b]Department of Civil Engineering and Geomatics, Cyprus University of Technology, Archiepiskopou Kyprianou, 3036, Limassol, Cyprus.
[c]Tor Vergata University of Rome, Via Del Politecnico 1, 00133, Rome, Italy.
*stelios.neophytides@eratosthenes.org.cy

ABSTRACT

Precision agriculture is the application of correct amount of fertilizers and water pesticide to achieve higher agricultural productivity. Furthermore, under the framework of precision agriculture is the automated estimation of yield with advanced technologies including Artificial Intelligence (AI) and Remote Sensing (RS). The use of RS has advanced crop yield estimations and predictions in recent years. However, to validate RS-based models it is important to perform in-situ exercises such as fruit counting, which is a time-consuming task that increases the production costs. Drones, robots, and in-situ cameras in combination with AI algorithms are widely used to efficiently address these issues. The recent advancement in computational resources and power available has enabled the utilization of Deep Learning AI models. One of the best-performing models for object detection is the You-Only-Look-Once (YOLO). In this study, the YOLOv5s is used for object detection, which is the second smallest and fastest YOLOv5 architecture, on two different benchmark datasets collected from AgML. The first dataset consists of 1730 images of mango trees in Australia during night, and the second dataset consists of 6512 images of wheat heads collected from different regions around the world. The main objective of this work is to demonstrate the capabilities of light AI models for object detection and to evaluate their performance, which will serve as a benchmark for future comparison with the on-board environment.

Keywords: Remote sensing, artificial intelligence, crop yield, agriculture, object detection

1. INTRODUCTION

Agriculture is one of the largest and most important industries in Cyprus and worldwide [1]. Crop yield estimation underpins planning strategies to fulfill projected demands of human population under the constraints of food security. Precision agriculture (PA) is a farming method that uses data and modern technologies to maximize crop yield through decision-making applications. A key goal of PA is to improve yield estimation, which has traditionally been estimated with manual sampling or other indirect methods. Remote Sensing (RS) has been found important in PA and yield estimation tasks during the last decades, due to the daily volume of data it generates [4]. However, validating such RS-based models for yield estimation requires further enhancement with in-situ work, e.g., fruit counting. Those field exercises require many manhours. Estimating yield with higher accuracy helps farmers, stakeholders, and relevant governmental bodies for a better planning to satisfy the needs of consumers. Moreover, yield assessments can enhance the market projections and more precise financial management of agricultural market[9].

The use of data-driven application in farming using Artificial Intelligence (AI) has shown promising performance to efficiently validate RS models [3]. For example, data derived from unmanned aerial vehicles (UAVs), robots and cameras installed in fields are combined with Deep Learning (DL) models that are able to perform object detection tasks and, consequently, perform fruit counting [5], [6], [7]. Two primary categories exist in the subject of object detection AI models, i.e., one-stage and two-stage models. Two-stage models are typically based on the architecture of region-based Convolutional Neural Network (R-CNN) and they are utilizing a region proposal network (RPN) such as the Faster R-CNN [10] to extract potential object regions. The extracted regions are then used to classify the objects of an image and execute the bounding box regression to mark the object. These algorithms are executing tasks with higher accuracy due to their capability to execute detection in two stages by enhancing object localization, but they are slow and not usually suitable for real-time applications. More variations of those methodologies have been developed through the years such as

Tenth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2024), edited by A. Christofe, S. Michaelides, D. Hadjimitsis, C. Danezis, K. Themistocleous, N. Kyriakides, G. Schreier, Proc. of SPIE Vol. 13212, 132121A · © 2024 SPIE · 0277-786X · doi: 10.1117/12.3037319

Fast R-CNN and Mask R-CNN [11], [12]. On the other hand, one-stage models are more efficient in the aspect of real-time processing as they are executing object classification and bounding box regression in the same neural network. Popular models are the different versions and variations of the You Only Look Once (YOLO) algorithm [13], [14].

According to a study, YOLOv5s achieved 1.84% higher precision on different varieties of grapes' real-time detection in complex agricultural environments compared to other YOLO variations [15]. A channel pruning is applied on YOLOv5 to develop a lighter architecture by reducing the floating-point operations and then used on real-time tracking of grapes on field images and reached a mean Average Precision (mAP) of 82.3% while also improved the detection performance on overlapping grapes [16]. An improved YOLOv5s using the method Soft Non-Maximum Suppression (Soft-NMS) is used for tomato detection in a working environment as a part of picking robots [17]. The model succeeds 92% and 82% of precision and recall, respectively. Another study shows the ability of YOLOv5 installed on picking robots for real-time apple detection [18]. A YOLOv5s enhanced with a spatial attention module and an adaptive context information fusion module is developed for pineapple buds' detection on UAV's in-situ captured images. The proposed methodology increased the mAP@0.5 (mean Average Precision at Intersection of Union equal to 0.5) by 7.4% and mAP@0.5:0.95 (mean Average Precision at Intersection of Union from 0.5 to 0.95) by 31% [19]. An image processing methodology is proposed for the detection of cotton bolls on images captured by a remote-controlled robot, using more traditional methodologies [20]. Later on, an improvement of this work was proposed by the same author, using a variation of Faster R-CNN, known as FrRCNN5-cls, which achieved an R-squared (R^2 or the coefficient of determination) of 0.88 and Root Mean Square Error (RMSE) of 0.79 [21]. Another model, which uses YOLOv5s as the base model with ShuffleNet-v2 as the backbone model, after fine tuning achieved an improvement of 3.5% at mAP on litchi fruit detection [22]. The YOLOv4 is evaluated on RGB chestnuts' images captured by a UAV and achieved an R^2 of 0.98 and RMSE of 6.3 [24]. Moreover, improved fundamental variations of YOLOv5s adapting attention mechanisms and robustness to fog are considered for real-time object detection tasks in uncertain agricultural landscapes [25], [26].In this study one of the light variations of YOLOv5s is adopted on two different benchmark datasets for counting mangoes and wheat heads. Mango fruit is one of the most demanded fruits around different geographical areas in the world such as Asia, Africa, and Latin America. On the other hand, wheat has a high importance for Cyprus with an annual value of ca. 16.4 million euros, covering an area of ca. 12.000 ha and a production of ca. 25.000 metric tons (t) per year [8]. Furthermore, this study aims to evaluate YOLOv5s on crop yield estimation tasks using in-situ captured images as a preliminary work before comparing it with other lightweight variations and examine the different on-board processing aspects.

The rest of the paper is structured as follows. Section 2 describes the methodology followed for the fine tuning of YOLOv5s, the experimental setup and the evaluation strategy used. Moreover, describes the two benchmark datasets used in the experimental study. Section 3 gives an overview of the experimental results concerning the performance and time consumption of the investigated model. Finally, Section 4 concludes this work.

2. MATERIALS AND METHODS

2.1 Benchmark datasets

In this study. two benchmark datasets are utilized for the preliminary evaluation of YOLOv5s on yield estimation from field images. Both datasets are acquired using the AgML Python library [27], which is an open-access package offering access to fundamental object detection models and benchmark datasets for agricultural tasks. The first dataset consists of 1730 images of mango trees during night-time. The images are captured using a ground-based RGB sensor from mango orchards in Australia [28]. On the other hand, the dataset for wheat heads is also developed from images acquired with a ground-based RGB sensor. The images are captured from plots cultivated with wheat from around the world and are 6512 in total [29]. Both datasets consist of single classes and are suitable for training fundamental models performing object detection tasks.

Figure 1. Example of the two benchmark datasets. On the left there is a representative example from the mango dataset whereas on the right a representative example from the wheat head dataset. Red boxes on both images indicate the annotation boxes.

2.2 You Only Look Once version 5

YOLOv5, or You Only Look Once version 5, is a fundamental object detection model with real-time capabilities. The architecture (figure 2) of YOLOv5 consists of a backbone network, neck and head (output). The primary function of the backbone, which is usually an altered CSPDarknet53, is to extract detailed, layered features from the input images. The neck, which consists of PANet (Path Aggregation Network), connects the various backbone stages to improve feature fusion and the model's capacity to detect objects at different scales. Ultimately, the head is made up of multiple convolutional layers that predict confidence scores, object classes, and bounding boxes. YOLOv5 is enhanced in four different aspects. The first advancement in the model is the operation of mosaic-data-augmentation which improves the training speed and network accuracy. YOLOv5, in order to enhance its ability of detecting targets of various sizes, is using the adaptive anchor computation and the adaptive picture scaling techniques. Furthermore, the model adopts fresh concepts in the backbone network and, to enhance the computation and improve the feature representation, adds the focus structure and the cross-stage partial (CSP) structure. Moreover, the Neck network is using a Feature Pyramid Network and a Pyramid Attention Network and adds the CSP2 architecture [30], [31]. Those newly added strategies enhanced the general model's performance. Still, YOLOv5 faces limitations in detecting tiny targets. YOLOv5s is the second smallest and fastest architecture in the family of YOLOv5 algorithms [32].

2.3 Evaluation metrics

The performance of YOLOv5 was evaluated using three different well known evaluation metrics for object detection tasks. The first metric given in eq.1 is the Precision, which measures the ability of a classification algorithm to identify only the relevant data points. Recall, which is given in eq.2, is the ability of a model to identify all the relevant cases within a data set.

$$Precision = \frac{true\ positives}{true\ positives + false\ positives}\ (1),$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives}\ (2),$$

where true positives is the number of model's correct predictions of the positive class, false positives is the number of model's incorrect predictions of the positive class and false negatives is the number of model's incorrect predictions of the negative class.

The mean Average Precision (mAP) given in eq.3 is widely used for object detection models to evaluate their performance based on a range of Intersection of Union (IoU).

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k\ (3),$$

where n is the number of classes and AP_k is the average precision of class k. In cases where mAP is evaluated on certain ranges of IoU, it is the average precision of class k at the specific IoU range.

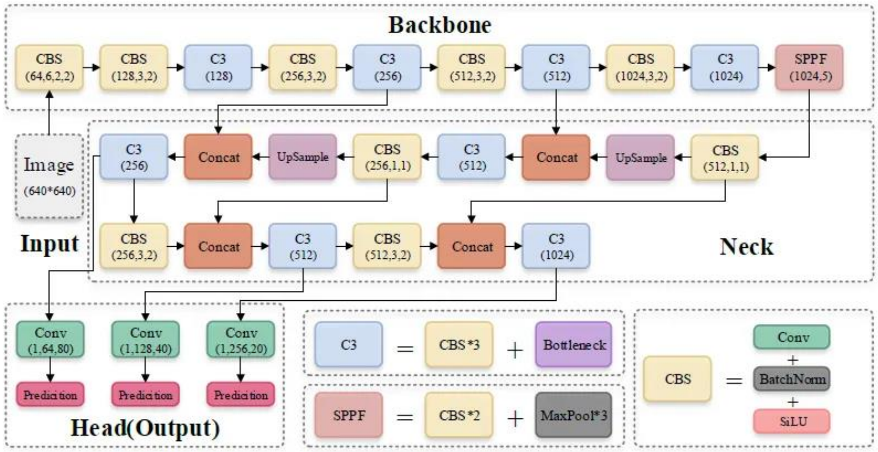

Figure 2. YOLOv5 architecture consists of the input layer, backbone, neck and head (output) components.

2.4 Experimental Setup

To evaluate the performance of the YOLOv5s object detection model on the two datasets, they were randomly split into training (70%), validation (20%) and testing (10%) subsets. YOLOv5s is a fundamental pre-trained model and thus hyperparameter tuning is not necessary to be performed in this case. Moreover, for each dataset, two independent instances of the model are fine-tuned with further training of 50 epochs. YOLOv5 is applying built-in augmentations (e.g., blur, median blur, gray, etc.) to the datasets.

3. EXPERIMENTAL RESULTS

3.1 Model's performance

The accuracy metrics are defined in Equations (1) to (3) and are used to evaluate the YOLOv5s. Table 1 shows the precision, recall, mAP@50 and mAP@50:95 achieved on both datasets. The investigated object detection model performed better on the mango dataset with respect to all the evaluation metrics, i.e., Precision = 0.960, Recall = 0.954, mAP@50 = 0.981, and mAP@50:95 = 0.734. On the wheat head dataset the model achieved relatively high accuracy with respect to three of the evaluation metrics, i.e., Precision = 0.910, Recall = 0.824, and mAP@50 = 0.896, but has low accuracy in mAP@50:95. This is because mango colour, shape, and size (especially when they are ripening) creates a contrast with the background of the image and, thus, it is easier for YOLOv5s to detect and generate a bounding box around the fruits. In contrast, wheat colour together with the density of the stems (during both of their states, green and dry) in an image imposes challenges for YOLOv5s to detect the wheat heads. Furthermore the limitations on detecting tiny objects probably influences the performance of the model, despite the large dataset. The model counted a total of 1151 mango instances on the 127 images of the testing set, and 4399 wheat heads instances on the 100 images of the testing set. Figures 3 and 4 present examples of the detected mangoes and wheat heads, respectively.

Table 1. Experimental results of object detection tasks using YOLOv5s on mango and wheat head benchmark datasets.

Dataset	Precision	Recall	mAP@50	mAP@50:95
Mango	0.960	0.954	0.981	0.734
Wheat Head	0.910	0.824	0.896	0.519

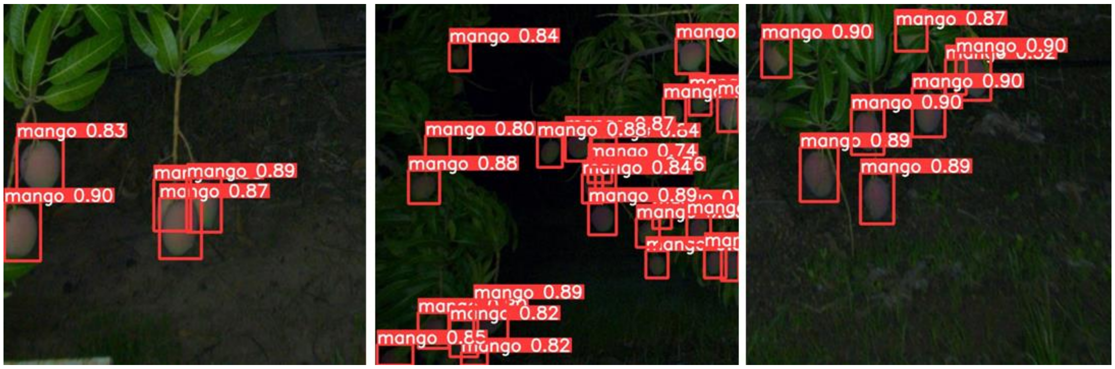

Figure 3. Indicative examples of detected mangoes

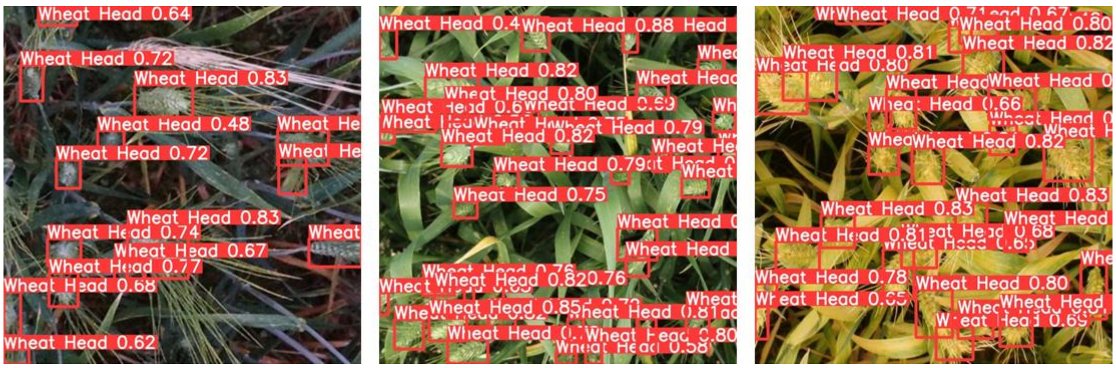

Figure 4. Indicative examples of detected wheat heads

3.2 Time consumption of YOLOv5s

YOLOv5s needed almost the same time for the 50 epochs of training on both mango and wheat head datasets with 1753.2 and 1828.8 seconds, respectively. YOLOv5s is faster on the wheat head dataset regarding pre-process and inference with 0.0004 and 0.0116 seconds, respectively. The model performed non-maximum suppression in 0.0141 seconds for the mango dataset, in contrast to the 0.0190 seconds needed for the wheat head dataset. All the performance metrics are shown in Table 2. Further than training, the model applies other functionalities that are pre-process, which is the step where the YOLOv5 prepares the image by applying resizing and normalization on pixels' values, inference, which is the step where the model is making the predictions, and Non-Maximum suppression, which is the step where the model is keeping the best bounding box for each object by comparing to others and suppressing low-confidence and overlapping boxes. The experiments are conducted using a Virtual Machine from Google Colab equipped with 52GB of RAM and an A100GPU. Future experiments will include the use of specific on-board hardware.

Table 2. Cost effective metrics.

Dataset	Training (s)	Pre-process (s)	Inference (s)	Non-maximum suppression (s)
Mango	1753.2	0.0083	0.0142	0.0141
Wheat Head	1828.8	0.0004	0.0116	0.0190

4. CONCLUSION

In this work, YOLOv5s, a light variation of YOLOv5, is investigated for its performance on two benchmark datasets for yield estimation using object detection. Based on the model's performance on the experiments, the following concluding remarks can be drawn. YOLOv5s performed better on the mango dataset, mainly due to the colour, shape and size of the fruits. However, the shape, colour and density of wheat heads generates multiple objects overlaps in the images resulting in lower accuracy for the YOLOv5s. The fine tuning of the model requires approximately 1800 seconds. The inference time of 0.0142 and 0.0116 enables the further examination of on-board yield estimation using ground-based cameras or UAVs. Indeed, the use of AI on-board requires evaluating specific aspects, such as the size of the model and the use of computational resources, which must be limited compared to a ground-based scenario. Therefore, the use of specific on-board hardware will be considered for the future development.

ACKNOWLEDGEMENTS

This work was partially supported by the European Union's HORIZON Research and Innovation Programme under grant agreement No 101120657, project ENFIELD (European Lighthouse to Manifest Trustworthy and Green AI), the AI-OBSERVER project funded from the European Union's Horizon Europe Framework Programme HORIZON WIDERA-2021-ACCESS-03 (Twinning) under the Grant Agreement No 101079468, and the 'EXCELSIOR': ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment H2020 Widespread Teaming project (www.excelsior2020.eu). The 'EXCELSIOR' project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857510, from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development and the Cyprus University of Technology.

REFERENCES

[1] M. Eliades *et al.*, 'Earth Observation in the EMMENA Region: Scoping Review of Current Applications and Knowledge Gaps', *Remote Sensing*, vol. 15, no. 17, p. 4202, Aug. 2023, doi: 10.3390/rs15174202.
[2] A. Nyéki and M. Neményi, 'Crop Yield Prediction in Precision Agriculture', *Agronomy*, vol. 12, no. 10, p. 2460, Oct. 2022, doi: 10.3390/agronomy12102460.
[3] M. T. Linaza *et al.*, 'Data-Driven Artificial Intelligence Applications for Sustainable Precision Agriculture', *Agronomy*, vol. 11, no. 6, p. 1227, Jun. 2021, doi: 10.3390/agronomy11061227.
[4] M. Weiss, F. Jacob, and G. Duveiller, 'Remote sensing for agricultural applications: A meta-review', *Remote Sensing of Environment*, vol. 236, p. 111402, Jan. 2020, doi: 10.1016/j.rse.2019.111402.
[5] L. Deng *et al.*, 'Lightweight aerial image object detection algorithm based on improved YOLOv5s', *Sci Rep*, vol. 13, no. 1, p. 7817, May 2023, doi: 10.1038/s41598-023-34892-4.
[6] A. Pretto *et al.*, 'Building an Aerial–Ground Robotics System for Precision Farming: An Adaptable Solution', *IEEE Robot. Automat. Mag.*, vol. 28, no. 3, pp. 29–49, Sep. 2021, doi: 10.1109/MRA.2020.3012492.
[7] B. Siciliano and O. Khatib, Eds., *Springer Handbook of Robotics*. in Springer Handbooks. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-32552-1.
[8] 'Agriculture, Livestock, Fishing - Predefined Tables'. Accessed: May 30, 2024. [Online]. Available: https://www.cystat.gov.cy/en/KeyFiguresList?s=28
[9] E. Benami *et al.*, 'Uniting remote sensing, crop modelling and economics for agricultural risk management', *Nat Rev Earth Environ*, vol. 2, no. 2, pp. 140–159, Jan. 2021, doi: 10.1038/s43017-020-00122-y.
[10] B. Cheng, Y. Wei, H. Shi, R. Feris, J. Xiong, and T. Huang, 'Revisiting RCNN: On Awakening the Classification Power of Faster RCNN', in *Computer Vision – ECCV 2018*, vol. 11219, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11219. , Cham: Springer International Publishing, 2018, pp. 473–490. doi: 10.1007/978-3-030-01267-0_28.
[11] A. K. Das, J. Nayak, B. Naik, S. K. Pati, and D. Pelusi, Eds., *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, vol. 999. in Advances in Intelligent Systems and Computing, vol. 999. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-13-9042-5.
[12] Ross Girshick, 'Fast R-CNN', in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.
[13] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi, and A. K. Bhoi, 'Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection', *Int J Comput Intell Syst*, vol. 16, no. 1, p. 126, Aug. 2023, doi: 10.1007/s44196-023-00302-w.
[14] T. Diwan, G. Anirudh, and J. V. Tembhurne, 'Object detection using YOLO: challenges, architectural successors, datasets and applications', *Multimed Tools Appl*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/s11042-022-13644-y.

[15] C. Zhang, H. Ding, Q. Shi, and Y. Wang, ‘Grape Cluster Real-Time Detection in Complex Natural Scenes Based on YOLOv5s Deep Learning Network’, *Agriculture*, vol. 12, no. 8, p. 1242, Aug. 2022, doi: 10.3390/agriculture12081242.
[16] L. Shen *et al.*, ‘Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLOv5s’, *Computers and Electronics in Agriculture*, vol. 206, p. 107662, Mar. 2023, doi: 10.1016/j.compag.2023.107662.
[17] G. Gao, C. Shuai, S. Wang, and T. Ding, ‘Using improved YOLO V5s to recognize tomatoes in a continuous working environment’, *SIViP*, vol. 18, no. 5, pp. 4019–4028, Jul. 2024, doi: 10.1007/s11760-024-03010-w.
[18] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, ‘A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5’, *Remote Sensing*, vol. 13, no. 9, p. 1619, Apr. 2021, doi: 10.3390/rs13091619.
[19] G. Yu, T. Wang, G. Guo, and H. Liu, ‘SFHG-YOLO: A Simple Real-Time Small-Object-Detection Method for Estimating Pineapple Yield from Unmanned Aerial Vehicles’, *Sensors*, vol. 23, no. 22, p. 9242, Nov. 2023, doi: 10.3390/s23229242.
[20] S. Sun, C. Li, A. H. Paterson, P. W. Chee, and J. S. Robertson, ‘Image processing algorithms for infield single cotton boll counting and yield prediction’, *Computers and Electronics in Agriculture*, vol. 166, p. 104976, Nov. 2019, doi: 10.1016/j.compag.2019.104976.
[21] Y. Jiang, C. Li, R. Xu, S. Sun, J. S. Robertson, and A. H. Paterson, ‘DeepFlower: a deep learning-based approach to characterize flowering patterns of cotton plants in the field’, *Plant Methods*, vol. 16, no. 1, p. 156, Dec. 2020, doi: 10.1186/s13007-020-00698-y.
[22] L. Wang, Y. Zhao, Z. Xiong, S. Wang, Y. Li, and Y. Lan, ‘Fast and precise detection of litchi fruits for yield estimation based on the improved YOLOv5 model’, *Front. Plant Sci.*, vol. 13, p. 965425, Aug. 2022, doi: 10.3389/fpls.2022.965425.
[23] R. Sanya, A. L. Nabiryo, J. F. Tusubira, S. Murindanyi, A. Katumba, and J. Nakatumba-Nabende, ‘Coffee and cashew nut dataset: A dataset for detection, classification, and yield estimation for machine learning applications’, *Data in Brief*, vol. 52, p. 109952, Feb. 2024, doi: 10.1016/j.dib.2023.109952.
[24] T. Arakawa, T. S. T. Tanaka, and S. Kamio, ‘Detection of on-tree chestnut fruits using deep learning and RGB unmanned aerial vehicle imagery for estimation of yield and fruit load’, *Agronomy Journal*, vol. 116, no. 3, pp. 973–981, May 2024, doi: 10.1002/agj2.21330.
[25] T. Jiang, C. Li, M. Yang, and Z. Wang, ‘An Improved YOLOv5s Algorithm for Object Detection with an Attention Mechanism’, *Electronics*, vol. 11, no. 16, p. 2494, Aug. 2022, doi: 10.3390/electronics11162494.
[26] X. Meng, Y. Liu, L. Fan, and J. Fan, ‘YOLOv5s-Fog: An Improved Model Based on YOLOv5s for Object Detection in Foggy Weather Scenarios’, *Sensors*, vol. 23, no. 11, p. 5321, Jun. 2023, doi: 10.3390/s23115321.
[27] A. Joshi, D. Guevara, and M. Earles, ‘Standardizing and Centralizing Datasets for Efficient Training of Agricultural Deep Learning Models’, *Plant Phenomics*, vol. 5, p. 0084, Jan. 2023, doi: 10.34133/plantphenomics.0084.
[28] Anand Koirala, C McCarthy, Kerry Walsh, and Z Wang, ‘MangoYOLO data set’. Central Queensland University, 2021. Accessed: May 23, 2024. [Online]. Available: http://hdl.handle.net/10018/1261224, https://researchdata.edu.au/mangoyolo-set
[29] Etienne David *et al.*, ‘Global Wheat Head Detection (GWHD) dataset: a large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods’. Zenodo, 2020.
[30] G. Jocher and others, ‘YOLOv5 by Ultralytics (Version 7.0)[Computer software]’. 2020.
[31] G. Jocher and others, ‘YOLOv5 by Ultralytics. 2020’, *URL https://github. com/ultralytics/yolov5*, 2023.
[32] Y. Zhao, Y. Shi, and Z. Wang, ‘The Improved YOLOV5 Algorithm and Its Application in Small Target Detection’, in *Intelligent Robotics and Applications*, H. Liu, Z. Yin, L. Liu, L. Jiang, G. Gu, X. Wu, and W. Ren, Eds., Cham: Springer International Publishing, 2022, pp. 679–688.