ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

On-line real-time fruit size estimation using a depth-camera sensor

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Mengoli, D., Bortolotti, G., Piani, M., Manfrini, L. (2022). On-line real-time fruit size estimation using a depth-camera sensor [10.1109/MetroAgriFor55389.2022.9964960].

Availability:
This version is available at: https://hdl.handle.net/11585/913120 since: 2023-01-31

Published:
DOI: http://doi.org/10.1109/MetroAgriFor55389.2022.9964960

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

25 June 2026

On-line real-time fruit size estimation using a depth-camera sensor

Dario Mengoli†, Gianmarco Bortolotti*, Mirko Piani*, Luigi Manfrini*
{*dario.mengoli2, gianmarco.bortolotti, mirko.piani2, luigi.manfrini*}*@unibo.it*
**DISTAL department - University of Bologna* - Bologna, Italy
†*DEI department G. Marconi - University of Bologna* - Bologna, Italy

***Abstract*—Fruit weight is one of the factors taken into account when performing yield estimations together with the trees density and orchard's area. Thus, having the possibility to collect data about the weight of a large number of fruits in the orchard gives the possibility to increase the reliability of the yield estimation. Over recent years, mathematical models able to convert the fruit size into fruit weight were evaluated as effective. Since then, manual data collection with calipers and automated/continuous fruit gauges were tested to collect fruit size data to perform yield predictions. Their main drawbacks are respectively the need for human-labour, repetitiveness, being time-requiring and the limited sample varying from 20 to 200 fruits per hectare. This research is trying to discover and deepen the use of AI in agriculture for doing a step further: sizing fruits after their detection with a YOLOv5 Neural network algorithm. To reach this goal, a system which takes as a input RGB-D depth-camera's color images and 16 bit depth maps was developed. After applying YOLOv5 detection, two different methodologies (by mean of squared bounding boxes and circular shapes) to extract from the depth map the distance data needed to size the target object were tested. Results from a preliminary data-set showed that the system could be a potential solution to increase the sample dimension and perform yield prediction. The main drawbacks of the developed vision-system are related to the errors in sizing the objects, which are ranging from an underestimation of about 9 mm to an overestimation of 24 mm. From the initial results was possible to identify the squared-bbox-mediated sizing process as a better pathway rather than the one performed with circular-bboxes, since the RMSE is always smaller with values of 7-9 mm.**

***Index Terms*—Yield prediction, fruit size estimation, online data collection, RGB-D, CNN**

I. Introduction and Problem Definition

Harvest estimation is of paramount importance to allow timely intervention during the growth season, thus increasing quality and quantity of the final harvest. Also, harvest labour, storage and transport logistics need to be prepared in advance to handle all required operations in the usually short harvesting time-span. The crop load estimation is particularly important for fruit crops because plant crop load generally affect ending fruit size and quality [1] and it is particularly important for those fruit crops presenting a short fruit ripening window and storage timings [2]. A reliable estimation may not only be used for marketing adjustments and side-channel acquisitions, but also help making an effective planning of the resources and thus creating efficiency both in terms of costs and resource usage [3]. The current practice yield estimation is made by manually counting and measuring fruits on sample parcels, then multiplying an average result by the total cultivated area. Manual counting is however a time-consuming task that require labour force that is therefore drained from actual management growth duties. Furthermore, repetitive, and long- lasting tasks are more prone to errors during execution. Lately, digital fruit "sizer" exploiting IoT configuration and wireless sensors networks are available, but they still rely on a parcel-based estimation due to their fixed position [4] [5].

Vision-based fruit counting has been investigated on different orchards, by using RGB cameras mounted on ground vehicles [6]. Fruit counting nowadays is typically performed by using Convolutional Neural-Networks (CNNs) approaches [7], by exploiting object detection with state-of-the art algorithms such as YOLOv5 [8]. Recently, the availability of low-cost depth camera sensors has opened new interesting scenarios both for counting and measuring single fruits simultaneously. Our solution aims to exploit the depth/distance information coming from the Intel Realsense D435i depth camera, together with a YOLOv5 fruit detection algorithm, to not only enable fruit counting but also allowing size estimation of the single fruits/objects. Preliminary results achieved are interesting also if further improvements are needed to obtain effective results.

The paper unfolds as follows, in Section II describes in details the proposed solution, Section III reports the obtained results, and, finally, Section IV discusses what was achieved with related drawbacks and strength.

II. Materials and methods

The trial consisted in evaluating an adaptive computer-vision algorithm for sizing apple fruits in field and in real time. The utilized camera was an Intel RealSense D435i consumer-grade depth camera (RGB-D). The choice of this camera was made considering its specifications and because of preliminary field test results in which it performed better than other intel RealSense cameras (i.e., D455 and L515). To summarise the test, the last two cameras were unable to collect as much depth point as the D435i. The data collection of the presented study consisted in recording RGB-D videos (containing color and depth information) of a tree-row in an apple orchard in which a variable number of apples was tagged and measured for their maximum equatorial diameter by a digital caliper. Furthermore, tennis-balls were included in the scene as dimension reference to better evaluate the sizing capacity of the system (Fig.1). The apple orchard in which data collection was done, presented a robot-friendly architecture [9] that favors fruit visibility.

The RGB-D videos were recorded keeping the camera parallel to the orchard tree-row and at a fixed distance of 1 m approx. The recordings were performed at five different timings during the season (on 17/06, 07/07, 01/08, 30/08 and 12/10) so to evaluate the system on different fruit size and coloration. The camera was set at the best resolution suggested by the documentation, for both the streams: 1920x1080 pixels for RGB and 848x480 for depth; this to improve both fruit detection and depth/fruit size estimation. Recording frame rate was set at 30 frame per second (FPS), the maximum for the tested camera.

To identify fruits in the scene a YOLOv5 CNN algorithm was utilized. Considering both accuracy performances as well as computational needs the YOLOv5 “l” algorithm was chosen for this experiment. The model was trained on a data-set constituted of 123 apple trees images containing 5000 highly visible fruits (approx.) representing different fruits dimension and coloration. The data-set was split in train, validation, and test -set with a proportion of 70-20-10% respectively per each image set. Image augmentations was applied to the train-set, augmenting its image number by around 6 times; this creating images altered in color proprieties, rotation, applying blur etc. The model training was performed, for 500 epochs, on a dedicated Workstation equipped with 2 x Nvidia RTX2080super. The trained model showed an F1 score of 0.85 on the test set. Similarly, a model dedicated to tennis balls detection was trained obtaining an F1 score = 0.93. The trained YOLOv5 CNN algorithms were then applied on the RGB image to detect fruits (and tennis balls).

Following, for each detected fruit, a customized circle detection algorithm, based on OpenCV function [10] was applied to obtain the best fitting-to-the-detected-fruit circle (Fig.1: Detected circle). The customized circle detection algorithm utilized is based on OpenCV ”HoughCircles” function, which is applied to the grey-scale transformed RGB image and returns the center (pixel coordinates) and the radius (in pixels) of the detected circle.

To obtain the circle dimension in millimeters, two methodologies exploiting depth (z-dimension) information coming from the RGB-D camera were utilized. For both of them, before initialising fruit detection and sizing, the depth-map was resized and aligned to the color image. The first fruit sizing method (squared-bbox; Fig.1: Squared-bbox) consists in extracting the mean fruit distance from the camera, calculating it as the mean of all valid (i.e., non-zero) depth points enclosed in a square-shaped Region Of Interest (ROI), placed in the center of the bbox, and obtained by halving the bbox dimensions of the detected fruit. Then, the real scene dimension ($Scene_{dim}$ - mm^2) fitting in the camera frame at the extracted mean fruit distance ($mObj_{dist}$ - mm), was computed exploiting trigonometric relations between cameras intrinsic parameters (Field of View - FoV) and distance from the fruit, as reported in Eq.1.

$$Scene_{dim} = 2 * (mObj_{dist} * Tan(FoV/2)) \tag{1}$$

The real scene dimension was then divided by the RGB image resolution and the square root of that is considered as the real object (fruit) resolution (Obj_{res}) in mm/pixel.

To what concerns the second method (circle-bbox; Fig.1: Circle-bbox), the mean fruit distance is obtained as presented previously, but instead of using the square-shaped ROI, depth information are extracted from the projection of the area got by halving the radius of the circle returned by the OpenCv's ”HoughCircle” algorithm. This area is placed in the center of the detected circle. In this second method the real object resolution was computed exploiting fx, fy parameters, given by the camera intrinsic, those describe the focal length of the image as a multiple of pixel width and height respectively.

The actual fruit diameter (F_{size}) was then estimated, for both the methods, multiplying the detected-circle diameter in pixels (d_{px}), by the object resolution (Obj_{res} - mm/pixel), obtained in each methods, as reported in Eq. 2.

$$F_{size} = Fd_{px} * Obj_{res} \tag{2}$$

The same process was applied also to the tennis-balls included in the images so to obtain a clearer interpretation of the system performances since these objects present a size variability lower than 1 mm. The vision-system estimated diameter was then compared with the manually measured one and absolute (RMSE) and relative (RMSE %) root-mean-squared-error were computed for both apples and tennis-balls.

III. RESULTS

Results here presented refer to a preliminary data-set made up of 156 apple fruits and 73 tennis balls, taken by using the Unmanned Ground Vehicle described in [11] inside the experimental planar apple orchard in Cadriano described also in [12] . The entire data-set collected is still under evaluation due to the time needed for matching YOLOv5 detected apple with manually tagged ones. From these initial data, it appears that the developed detection and sizing processes (previously referred to as method 1 and method 2) resulted in being promising at determining fruits and tennis balls dimension. Concerning the obtained results (Fig. 2), it came into view that applying a squared-bbox (method 1) the rightness of the estimated object size increases since the RMSE for both apples (7.2 mm) and tennis balls (7.9 mm) is lower than the one obtained by running the circular- bbox method (i.e., 8.6 mm and 9.4 mm, respectively for apples and tennis-balls). Both methods seem to perform better in estimating fruit

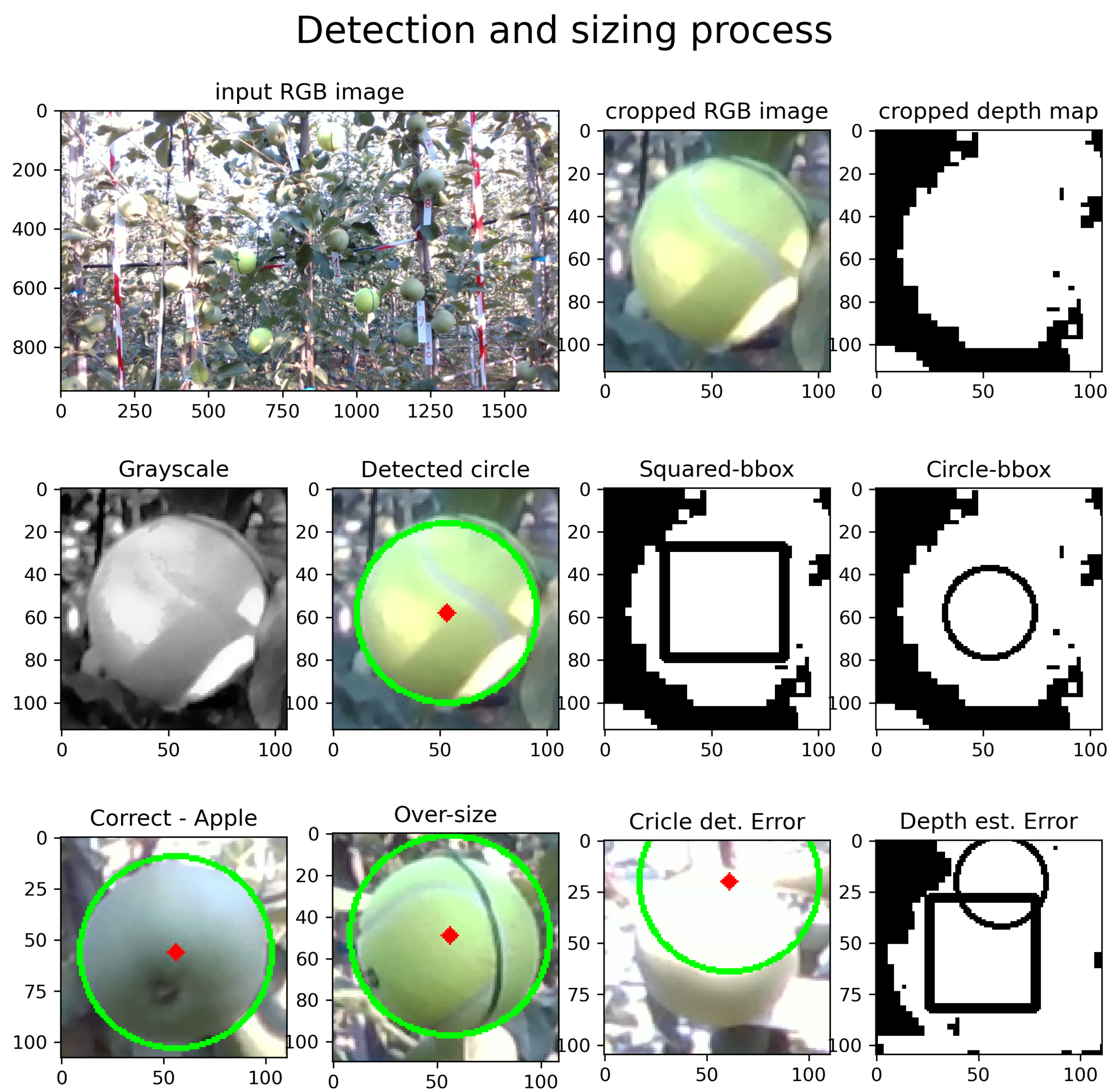

Fig. 1: In the first line, the image of the scene and cropped RGB and depth after CNN object detection application. In the second line the circle detection and sizing process, In the last line, examples of correct circle detection, oversized circle detection, erroneous circle detection due to excessive light: Squared- and Circle- bbox methods are affected differently in depth estimation.

size compared to tennis balls size; this is somehow counter-intuitive considering the reduced size variability and the more regular shape of tennis balls. If considering the RMSE % values, the system seems to perform equally on both types of objects (apple, tennis ball) presenting values of 12% and 14% respectively for method 1 (squared-bbox) and 2 (Circular bbox).

Concerning the field performances of the YOLOv5 algorithm, no results can be presented since not all the object in each scene was manually label to test this, but of the 156 actually tagged fruits only one was not detected, while ll tennis balls were detected. Regarding the efficacy of the customized circle detection algorithm, it was able to identify “best-fitting circles” on 141 apples out of the whole 155 apples (i.e., 90%); for tennis balls, in each of the 73 tennis balls, the algorithm was able to extract a “best-fitting circle” on which to estimate size.

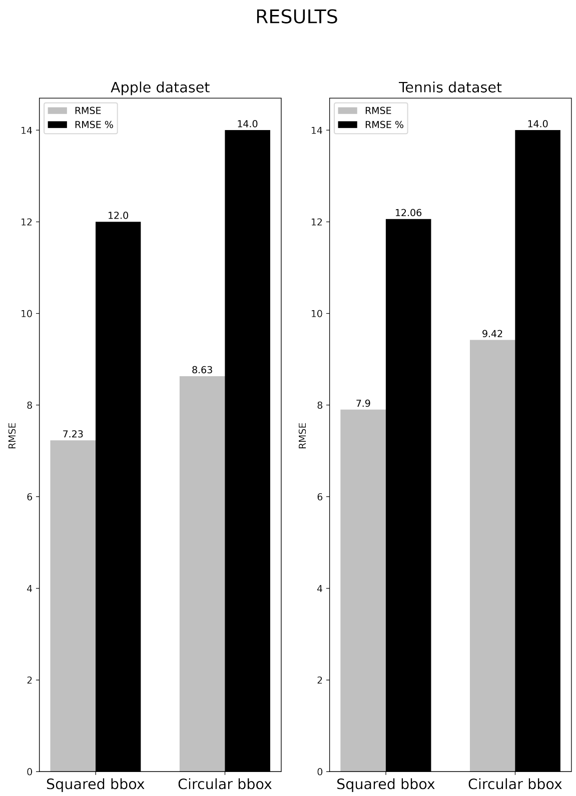

Fig. 2: The bar-plot is showing the RMSE and RMSE% for both squared-bbox and circle-bbox methods utilized for sizing the apples, on both tennis-balls and apple data-sets.

Regarding the real time application of the system, the whole process from loads the images to save the results presented a mean processing time of 0.55s (0.34 s for the YOLOv5 model application; 0.21 s for the fruit size estimation with both the methods) per image, on a laptop equipped with a Intel i7-9750H CPU and 16 GB of RAM.

IV. Discussion

The described trial allowed to highlight the importance of object detection step in the process since errors shown to be sometimes related to a wrong-sized object detection bbox, in which the circle detection algorithm did not work properly (or at all - missing detection). Considering regular shape object like tennis balls, it can be assumed that, the system tend to generally oversize the dimension estimation, since the smallest RMSE and mean error(2.4 mm - +1.3mm) were found, substituting the detected-circle diameter ,in pixel, with the minimum bbox dimension (i.e. the smallest between x and y). Supporting that, is the fact that, for tennis balls, all the mean errors computed in respect of the actual tennis ball dimension are greater than 0. For fruits, the non-regular shape led to a negative mean error when substituting minimum bbox dimension to detected circle diameter (i.e., -1.3 mm; RMSE = 4.7mm -the smallest-); so for fruits the system still tends to generally overestimate size, but with a higher variability related to single case. The maximum error in fruit size estimation were -9.2 mm and - 10.4 mm in under estimating, while +24.4 mm and +26.9 mm in over estimating, respectively for method 1 and 2 in both the cases. What just exposed suggests that ”HoughCircle” algorithm tend to detect oversized circles, so it was tried to tried to reduce the size of the bounding box (after the detectition), but this resulted in a negligible reduction of the sizing error. So, results seem to indicate that over-sizing is mainly related to the customized ”HoughCircle” algorithm that sometimes detect oversized wrong circles (Fig. 1: Over-size) affecting the overall dimension estimation.

Additionally, some technological issues related to the RGB-D camera sensors were found: the optical sensor saturates with levels of incident radiation, making an object so bright that the CNN is not able to detect it; while, the infrared sensor used to detect the distance data is affected as well by high levels of radiation because of the infrared waveband which alters the depth estimation [13]. High illuminance can also affect circle detection algorithm inducing important errors (Fig. 1: Circle and Depth errors). For this reasons, we have identified a workflow which consists in acquiring the images needed for the detection and sizing process, during the morning or late afternoon when the direct light interception and high light intensity are not an issue. Obviously, even acting on the orchard environment by mean of intensive shading helps in reducing those technological problems. Knowing those issues is useful to improve the efficiency of the vision - system by acting at the right time or by calibrating the infrared/ RGB sensor’s sensibility. An adding advice would be to take recordings during night time, with artificial lighting, this also to standardize image/video collection conditions, favouring the overall system performances.

According to results, the squared-bbox method performed always better than the circular-bbox based one (-2% of RMSE both for apples and tennis-balls) and this is mainly related to the different object resolution computed by the two methods. Method 2 resulted more erroneous due to the not precise circle detection: if the circle detected was not placed in the center of the object to be sized, then the distance estimation of the object itself resulted less correct.

Despite interesting results, the system is currently not in line with suitable error range for fruit size estimation (i.e. 1-2mm), but large scale data collection together with system improvements could lead to average sizing error falling in 1-2mm range.

Regarding the real-time application of the system, in the results are presented computational time needs for CPU based processing, and those report a rate of 2 image analyzed per second. The system is currently under improvement so to be run on high-power processing units, dedicated to image analysis (i.e., GPU/TPU), known to highly reduce the processing time. Preliminary results obtained from on a Nvidia RTX-A2000 shown a mean processing time of 20 ms per frame, resulting in a frame rate of application > 30 FPS (i.e., 50 FPS, approx) and therefore showing the possibility to obtain online real-time sizing information of detected object.

V. Conclusions

In this paper we presented a promising vision-system for detecting and sizing apple fruits directly in field and in real-time with the potentiality to improve yield prediction and safeguard fruit growers during the marketing procedures. The system performed mainly by overestimating the actual object size both with squared-bbox and circular-bbox method utilized, with an RMSE of 7.9 mm and 8.6 mm for apples dimension, respectively.

Introducing the tennis balls in the evaluation process allowed to understand that the fruit size variability is a parameter which is increasing the difficulties in getting the correct object dimension, considering the prospective with which the camera frame the fruit. The system currently present an excessive sizing error for the application in field, but further improvements will be investigated thanks to the results of this trial. We are confident that this system could support growers and fruit chain in getting large scale information regarding fruit size attribute.

References

[1] S. Serra, R. Leisso, L. Giordani, L. Kalcsits, and S. Musacchi, "Crop load influences fruit quality, nutritional balance, and return bloom in 'honeycrisp' apple," *HortScience*, vol. 51, no. 3, p. 236 – 244, 2016, cited by: 48. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84963997706&doi=10.21273%2fhortsci.51.3.236&partnerID=40&md5=85645e9be4837304e01a94698bd603b5

[2] W. Qureshi, A. Payne, K. Walsh, R. Linker, O. Cohen, and M. Dailey, "Machine vision for counting fruit on mango tree canopies," *Precision Agriculture*, vol. 18, no. 2, pp. 224–244, 2017.

[3] G. Moreda, J. Ortiz-Cañavate, F. García-Ramos, and M. Ruiz-Altisent, "Non-destructive technologies for fruit and vegetable size determination – a review," *Journal of Food Engineering*, vol. 92, no. 2, pp. 119–136, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0260877408005451

[4] B. Morandi, L. Manfrini, M. Zibordi, M. Noferini, G. Fiori, and L. C. Grappadelli, "A low-cost device for accurate and continuous measurements of fruit diameter," *HortScience*, vol. 42, no. 6, p. 1380 – 1382, 2007, cited by: 61. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-34648816416&doi=10.21273%2fhortsci.42.6.1380&partnerID=40&md5=35e328e399f74e6b9cabd15888947203

[5] L. M. Peppi, M. Zauli, L. Manfrini, P. A. Traverso, L. Corelli Grappadelli, and L. D. Marchi, "A low-cost and high-accuracy non-invasive system for the monitoring of fruit growth," 2020, Conference paper, p. 18 – 23, cited by: 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099011516&doi=10.1109%2fMetroAgriFor50201.2020.9277571&partnerID=40&md5=a6186f79e638c65c762059927a6aab19

[6] A. Syal, D. Garg, and S. Sharma, "A survey of computer vision methods for counting fruits and yield prediction," *International Journal of Computer Science Engineering*, vol. 2, no. 6, pp. 346–350, 2013.

[7] J. Gené-Mola, R. Sanz, J. Rosell-Polo, A. Escolà, and E. Gregorio Lopez, "Pfuji-size dataset: A collection of images and photogrammetry-derived 3d point clouds with ground truth annotations for fuji apple detection and size estimation in field conditions," *Data in Brief*, vol. 39, p. 107629, 11 2021.

[8] G. Jocher, A. Stoken, A. Chaurasia, J. Borovec, NanoCode012, TaoXie, Y. Kwon, K. Michael, L. Changyu, J. Fang, A. V, Laughing, tkianai, yxNONG, P. Skalski, A. Hogan, J. Nadar, imyhxy, L. Mammana, AlexWang1900, C. Fati, D. Montes, J. Hajek, L. Diaconu, M. T. Minh, Marc, albinxavi, fatih, oleg, and wanghaoyang0106, "ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support," Oct. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5563715

[9] D. Tustin, B. van Hooijdonk, and K. Breen, "The planar cordon – new planting systems concepts to improve light utilisation and physiological function to increase apple orchard yield potential," *Acta Horticulturae*, vol. 1228, p. 1 – 11, 2018, cited by: 12. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059975078&doi=10.17660%2fActaHortic.2018.1228.1&partnerID=40&md5=a5d9dfcdb207bb6ea7295aa709bd3295

[10] G. Bradski, "The opencv library." *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

[11] D. Mengoli, R. Tazzari, and L. Marconi, "Autonomous robotic platform for precision orchard management: Architecture and software perspective," in *2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, 2020, pp. 303–308.

[12] G. Bortolotti, K. Bresilla, M. Piani, L. C. Grappadelli, and L. Manfrini, "2d tree crops training system improve computer vision application in field: a case study," in *2021 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, 2021, pp. 120–124.

[13] J. Gené-Mola, J. Llorens, J. R. Rosell-Polo, E. Gregorio, J. Arnó, F. Solanelles, J. A. Martínez-Casasnovas, and A. Escolà, "Assessing the performance of rgb-d sensors for 3d fruit crop canopy characterization under different operating and lighting conditions," *Sensors*, vol. 20, no. 24, p. 7072, 2020.