On-Tree Mango Fruit Count Using Live Video-Split Image Dataset to Predict Better Yield at Pre-Harvesting Stage

Original Scientific Paper

Devender Nayak Nenavath
Vellore Institute of Technology, Vellore
School of Computer Science and Engineering-SCOPE
Vellore, Tamilnadu, India-632014
devendernayak.n@vit.ac.in

Boominathan Perumal*
Vellore Institute of Technology, Vellore
School of Computer Science and Engineering-SCOPE
Vellore, Tamilnadu, India-632014
boominathan.p@vit.ac.in

*Corresponding author

Abstract *– This study introduces a method for fruit counting in agricultural settings using video capture and the YOLOv7 object detection model. By splitting captured videos into frames and strategically selecting representative frames, the approach aims to accurately estimate fruit counts while minimizing the risk of double counting. YOLOv7, known for its efficiency and accuracy in object detection, is employed to analyze selected frames and detect fruits on trees. Demonstrated the method's effectiveness through its ability to provide farmers with precise yield estimations, optimize resource management, and facilitate early detection of orchard issues such as pest infestations or nutrient deficiencies. This technological integration reduces labor costs and supports sustainable agricultural practices by improving productivity and decision-making capabilities. The scalability of the approach makes it suitable for diverse orchard sizes and types, offering a promising tool for enhancing agricultural efficiency and profitability. The researcher compared YOLOv5n, YOLOv5s, YOLOv7, and YOLOv7-tiny with eight-sided imaging techniques around the tree. The experimental results of YOLOv7 with the eight-sided technique performed best and achieved a count accuracy of 97.7% on a single tree in just 17.112 ms of average inference time. On multiple trees, it is 95.48% in just 17 ms of average inference time, with the help of an eight-sided method on tree images.*

Keywords: *Computer Vision, Deep Learning, Image Processing, Agricultural Technology, Horticulture Fruit*

Received: May 10, 2024; Received in revised form: July 18, 2024; Accepted: August 16, 2024

1. INTRODUCTION

The famous Persian poet "Amir Khusrau named the mango Naghza Tarin Mewa Hindustan," which means the fairest fruit of Hindustan (India). Mango is a member of the Anacardiaceae family and includes several other species, notably cashews, sumac, and pistachio, which are traditionally grown in different climates [1]. Mango fruit is a seasonally available fruit, particularly in summer in India. Mango is rich in polyphenols, predominantly gallic acid, and has antioxidant, anticancer, and anti-inflammatory activities that improve chronic inflammation. Polyphenols and mango fiber may serve as prebiotics to increase probiotic bacteria in the intestines. Anti-inflammation prevents other symptoms, such as colon cancer, chronic intestinal diseases, and leaky intestines, and improves intestinal health [2]. Mango peel and bagasse are rich sources of dietary fiber, which is beneficial for cardiovascular diseases, type 2 diabetes, metabolic syndrome, and cancer. Mango seeds are rich in vegetable oils, proteins, and antioxidants with antibiotic potential [3].

Our focus is only on mangoes because of their shape, structure, and color. Compared to various fruits, man-

goes are different in shape, with more than 1000 varieties, a few varieties shown in (Fig 1). Although we have a model with a specific dataset of mango types for detection and counting, the same model cannot work for another type of mango. To make this work possible for all varieties, different varieties of images are required to create a dataset for training, testing, and validation to make an algorithm with reasonable accuracy, whereas, in other fruit scenarios such as apple, lemon, and sapodilla, this is not necessary for training, testing, and validation for all varieties.

Horticulture is a subdivision of agriculture that involves plants, flowers, turf, fruits, and nuts. Different storage techniques, transportation facilities, and marketing strategies are available for pulses and cereals, but these facilities are for something other than fruits. Different storage mechanisms are available, such as cold storage and traditional storage mechanisms constructed with the help of wheat straw, paddy straw, grass, bamboo, wood, bricks, mud, and cow dung for grains. However, cold storage can change the fruit's taste and natural fragrance within a few days [4].

Fig. 1. Various mango varieties. Each mango is different in shape, structure, color, and size

Detection is a technique related to computer vision, which locates and identifies objects within a given image or video. Various algorithms detect an object, such as R-CNN, Fast R-CNN, Faster R-CNN, and Histogram of Oriented Gradients (HOG), which was used in improved YOLOv2 to detect immature mango fruit [5], R-FCN, SSD, SPP-net, and YOLO. Counting harvestable and non-harvestable fruits, if the fruit count is well known before plucking from the tree or at maturity of the fruit, a farmer can predict their outcome based on the fruit count and can participate in the online sale or sell the fruits at a confident price based on the number of fruits. Using this procedure, the farmer can obtain the expected yield. The central research concept of counting tree fruit will help farmers obtain better yields [6].

Counting is a technique that will give the number of detected objects based on the detection technique. Counting the number of mangoes on trees and branches is the best method for obtaining production data. At the beginning of the harvest cycle, performing the count when measuring all fruits of the productive cycle on the tree will result in the number of fruits per tree [6]. The farmer will come to know the expected outcomes from an orchard. Only a few counting methods are available to count fruits; the primary and traditional method is manual counting, where humans count the fruit by their eye vision on the tree; it is a very high-cost and time-consuming technique [7]. Bounding box counting is the following counting method used in image processing. Where the total fruit count is the observed bounding box count. [8]. The third method uses Vertical and Horizontal line-based counting, which is performed based on a line. If an object passes through the line, counting starts and produces the total passed count [9]. The fourth method of counting uses SORT techniques. SORT is helpful in deep learning concepts, where the fruit is counted based on the SORT technique [10]. Another counting technique is ROS core counting, in which the fruits are counted based on a Robot Operating System [11]. Another counting method is the region-of-interest (ROI) and unique object identification (ID) methods [12].

The contribution of this procedure involves developing an innovative and efficient method to count fruits on a tree using video capture and object detection with YOLOv7. We strategically selected representative frames from the different segments by capturing a 360-degree video around the tree and splitting it into individual frames to avoid double counting. This approach ensures comprehensive coverage of the tree while minimizing redundancy. Applying the YOLOv7 model to these frames enabled accurate and realtime fruit detection. Our method optimizes computational resources and enhances the accuracy of fruit counting, thereby providing a practical solution for agricultural and horticultural applications. This contribution is significant for improving the monitoring and estimation of fruit production and supporting better yield management and resource allocation in the field.

2. RELATED WORK

Counting the fruits and flowers was performed manually, which is very expensive; to overcome this expensive problem, they proposed a simulated deep convolutional neural network for yield estimation. In this study [7], they created a 26400 image dataset in which they used 24000 for training and the remaining 2400 for testing, and the error decreased by using the Adam optimizer. They used a modified version of the Inception-ResNet architecture to capture features at multiple scales. Finally, this network tested on authentic images. It achieved 91% accuracy. The advantage of this work is that it can be applied to other fruits because the dataset preparation involves filling the entire blank image with green and brown colored circles and simulating the background.

Ref. [9] proposed a lightweight YOLOV5-CS (Citrus Sort) object detection model with 3000 original images used to detect and count citrus fruits in the natural environment. First, to improve generalization image rotation, a convolutional layer with a block next to the

backbone and the subsequent detection layer was embedded for accuracy improvement. Both loss function full Intersection over Union and Cosine annealing applied for improved training. The developed model moves to implement an edge artificial intelligence system. For the counting scene segmentation method with the virtual region and the formed embedded system, mAP@.5 is 98.23%, and recall is 97.66% with a frame rate of 28 FPS.

In [13], the multi-scale multilayer perceptrons (MLP) and CNN were used to overcome the previous fruit segmentation performance of a benchmarked MLP network for fruit detection and counting in orchard image data. They incorporated metadata in these architectures to explicitly capture the relationships between meta-parameters and object classes. Watershed Segmentation and Circular Hough Transform algorithms were used to post-process pixel-wise image segmentation and achieved a computing and detection F1-score of 0.858. This model has the advantage of detecting partially circular regions, thereby enabling the merging of disjointed fruit regions into a single detection. In addition, it is not possible to visualize all fruits in the image data owing to occlusions and clustering.

To estimate the yield of citrus fruits under natural lighting conditions [14], a computer vision algorithm using a hybrid watershed transform was proposed to detect and count citrus trees and performed the image on 84 images from 21 trees. These images were noisy because they included some other tree parts. Therefore, some input images were subtracted from the background and resized to 1824:1028 to improve data processing speed. They converted these from RGB to HSV and evaluated the marker-controlled watershed and distance transform algorithms for automated watershed segmentation, obtaining an R^2 of 93%.

In [15], night-acquired images of 1515 trees across five orchards for single-stage architectures such as YOLOv3, YOLOv2, YOLOv2-tiny, SSD, and two-stage architectures such as Faster R-CNN with VGG, Faster R-CNN with ZF were train with an original resolution of 512 × 512 for a total of 11 models. Compared to a previous poor study on fruit, leaf color, shape, and texture, a hybrid model named MangoYOLO was developed. The MangoYOLO of 33 layers model was constructed based on the better features of YOLOv2-tiny's fewer layers and higher speed as advantages, as well as the multiple detection layers and high-speed features of YOLOv3. MangoYOLO achieved 0.97 of the F1-score for fruit detection in an image. They have proposed a new MangoYOLO model based on YOLOv3 and YOLOv2-tiny features and compared all models to obtain better results with the new MangoYOLO-512-pt model.

In [16] an R-CNN model performed training on 1160 unmanned aerial vehicles (UAV)- based data images of two years captured in different directions and at different distances from the ground level to detect and count the number of apple fruits on individual trees. The proposed model's results compare with the agro technician in situ apple counts; the acquired R-square value was 0.86, with a Mean Absolute Error of 10.35 and a Root Mean Square Error of 13.56. In the top-view images, the number of total images acquired R-square value was 0.80, with MAE: 128.56 and RMSE = 130.56. According to [16], using a colab is the main advantage.

In [17] a study based on a single-shot multi-box detector with MobileNet and a faster R-CNN with Inception V2 architectures for detection. Training and testing were performed on three different fruits, avocado, lemon, and apple, with two architectures, under different field conditions. For video-based fruit counting multi-object tracking with the Gaussian estimation algorithm, Faster R-CNN with Inception V2 achieved 93% of the result and 90% using SSD with MobileNet. A disadvantage of this study is that the results could be more conclusive for other fruits.

In recent years, deep learning has been widely applied in agricultural fields [18]. Using YOLO model techniques, detection was applied under various imaging and illumination situations to estimate the load of orange fruit in an orchard. They used 1115 trees for examination, conducted in three steps: creating an orange-tree dataset under different illumination conditions, evaluating the selected model on 100 sample trees, and finally extracting the yield based on detecting and counting the oranges of every image taken. Using this method, they observed some two-sided differences for thin canopy and four-sided differences for dense canopy imaging. With the help of the YOLOv4 model, the precision was 91.23%, the recall was 92.8%, F1-score was 92%, and mAP was 90.8%.

A novel methodology was developed for apple fruit detection and counting using deep learning with apple fruit trunk tracking. They [19] constructed their dataset using images and videos and divided the image data 800 into 80% and 20% ratios for training and testing. In early studies, these algorithms mismatched or lost their targets because of the large number of similar fruits. However, in this study of apple fruit trunk, which is usually more significant than the fruit in appearance, and YOLOv4-tiny with the channel spatial reliability-discriminative correlation filter (CSR-DCF) algorithm. The developed method was tested using the ID-switched number of fruits, MIDE, and RMSE to assess the performance of matching fruit in a video frame and observed an mAP of 99.35% for fruit trunk detection, 91.49% counting accuracy, and R-square of 0.9875. The advantage of the proposed method is that it provides the possibility of realtime yield estimation of the orchard using a CPU at 2–5 fps. They found some drawbacks with this procedure; this study considered only a single-sided row of the tree at the time of counting; if the practitioner uses both sides of the tree, it will provide a double count. Therefore, they suggested that some investigation is required to perform counting of both sides of the fruit.

For early crop load estimation of the apple fruit canopy, the [20] YOLOv4-based model used on 480 raw apple tree images split into three growing stages: early, mid, and harvest. Their previous research modified the YOLOv4 network architecture and fine-tuned it to adapt it for apple fruit detection. In this study, they designed the CA-YOLOv4 model with three significant improvements, which specifically addressed some major challenges: small fruit size, dense canopy conditions, and severe canopy occlusion. The first improvement is the convolutional block attention module (CBAM) mechanism, which learns to improve the detection accuracy based on target features and surpasses nontarget features. In the second stage, we added an adaptive layer and a large-scale map d together. The regression box loss function was optimized, and the last phase included the densely connected network structure. The results showed that CA-YOLOv4 had a lower final loss value, more excellent recall, f1-score, and precision. CA-YOLOv4 performed better than the Faster R-CNN and SSD. Finally, this proposed CA-YOLOv4 study performed a superior detector for fruit counting and can perform near realtime with an average detection time of 0.1 s per image with the described hardware.

Fruit counting is essential in orchard management and plantation science. Ref. [21] early studies showed a need for robust and accurate fruit-counting methods in complex orchards, such as covering, shadows, clustering images, and complete fruit counting on whole trees. This study proposed and validated a panoramic method based on deep learning object detection for complete yield estimation for holy fruit. This method used a holly fruit dataset of 640×640 samples divided into 75% and 25% ratios for training and testing purposes. To form a complete panoramic unfolding map of the fruit tree surface, the images surrounding the fruit trees were captured using a UAV, and SIFT-based image matching was performed. Tested the accuracy and effectiveness of this method at different scales and scenarios and observed that high-quality built panoramic images for an accurate fruit count. The statistical rate between the detected and actual number is more than 96% when the ring shot parameter of the holly tree is less than or equal to 1.2 m; when the shot ring parameter is less than or equal to 1.6 m, then the statistical rate is 95%. The detection rate between the detected and captured numbers in the panorama image is over 99% when R≤1.2 m and over 97% when R≤2.0 m. Even though the model has a high detection rate, the current confidence threshold is still missing. These missed fruits are difficult to identify because of incomplete fruit contours, fake pixel values, insufficient pixels, and mutual interference between highly similar targets.

An automatic apple counting system for modern orchards was developed by [22], where they acquired ten sets of original videos and 1600 images with 720×1280 pixels of two consequent harvesting seasons. These 1600 images were divided randomly into 1280 images and 320 images for training and testing, in which they were labeled manually to the fruits and trunks. In the third year, labele 93050 samples with two classes: fruit and trunk. They then performed a regular detection-matched fruit counting system (NDMFCS) test on ten sets of original videos. In NDMFCS, based on YOLOv4-tiny performed object detection, abnormal fruit detection was abatement based on a threshold, and fruit counting was performed based on trunk tracking and identity document (ID) assignment. Finally, the results indicated that the average fruit detection precision was improved from 89.1% to 93.3% based on ten sets of original videos. Implementing CPU at 3-5 FPS is the advantage of this model. However, it was developed based on Intel RealSense D435 camera-based videos, which are challenging to use widely because they require computing equipment.

Ref. [23] picking the litchi fruit failed but was located successfully due to random obstruction in early studies. In this study, with the help of 1000 training sets and 100 test sets with the YOLOv8-seg model on litchi and its branches classified images with binocular vision technology picked points, they proposed a picking point framework for the robot system. This procedure achieved 88.1% precision for segmenting litchi fruit branches with an 88% picking-point success rate, and the overall success rate was 81.3%, with an average error of 2.8511 mm. The advantage of this study is that it is quick and accurate in identifying target points, which is a realtime operation.

In this [24] article, they investigated fruit detection methods, including traditional and deep learning methods. However, they focused on deep learning and optimization strategies for fruit detection in two ways: optimization strategies for fruit detection on pre-image sampling and optimization strategies after image collection to overcome the unstructured background challenge in the orchard field environment. They studied complex background factors and adverse effects, lighting conditions, occluded fruits, fruits with different degrees of maturity, and complex backgrounds in outdoor orchard environments and suggested future work.

We created a dataset of 1021 pictures as every 20 frames of video data from the UAV obtained 304 images. The remaining 717 images were captured using a mobile phone, scaled to 640 × 640, and labeled using the LabelImg tool, divided into a ratio of 80:10:10. Existing works fail to balance speed and accuracy and perform well when features are distinct and occlusion minimized. Therefore, they [25] introduced a novel lightweight network architecture based on the YOLOv5 foundation. First, they pruned YOLOv5 using filter pruning and then introduced an adaptive BN layer to identify the best-pruned subnet based on the score. Finally, an ECA module is appended to the optimal network to form and fine-tune a new one. They observed that the proposed YOLOv5_E has 24.2% parameters as 26.2% of YOLOv5 size; it runs at 178 FPS, with only a 0.9% loss in accuracy. This model pruning is advantageous for efficiency, incurring a subtle decrease in accuracy.

Ref. [26] the citrus detection and dynamic counting method was proposed based on the lightweight target detection network YOLOv7-tiny, Kalman filter tracking, and the Hungarian algorithm. This work uses the YOLOv7-tiny algorithm for predictive tracking of discovered fruits utilizing the Kalman filter to recognize citrus fruit in a video. Added the Euclidean distance, overlap matching to the Hungarian technique, and a two-stage life filter. Finally, drawing line counting was proposed. The average detection accuracy of YOLOv7-tiny was 97.23%, dynamic detection was 95.12%, multi-target tracking accuracy was 67.14%, and the improved dynamic counting algorithm was 67.14%.

This study [27] introduces a novel blueberry ripeness and count detection methodology that integrates an attention mechanism with a bidirectional feature pyramid network (BiFPN) within the YOLOv5 framework. In 192 images of the 2515 blueberries, 1612 immature and 903 mature blueberries were divided 192 images into 80% for training, 10% for testing, and 10% for validation in this study. Their proposed YOLOv5-CA model achieved an mAP at an IoU threshold of 0.5, recall of 88.2%, precision of 88.8%, and culminating mAP of 91.1%. As the model was YOLOv5-SE+BiFPN, the mAP was 90.5%, the recall was 88.5%, and the precision was 88.4%.

Previous studies treated the detection of clusters and berries as separate tasks to count the grape, owing to the clustered nature of the grape. In this study, they [28] proposed a probability map-based grape detection and counting framework, where first detects two intermediate maps through a neural network and uses three stages to finish the three grape detection and counting subtasks; for this study, they used the WGISD dataset; Chengdu dataset, and BpGC dataset three different types of datasets. They used the WGISD dataset and combined 100 more datasets from the Chengdu dataset and tested the proposed framework, achieving a localization performance of AP of 0.851, counting performance of MAE of 1.845, RMSE of 2.142 for grape clusters, 23.414 for MAE and 31.391 for RMSE of counting performance for grape berries, and MRD of 0.142, 1-FVU 0.865 of counting performance for berries per grape cluster. However, this study also has certain limitations. First, our method detects only visible grape clusters and grape berries, while occluded, invisible grape berries still need to be discussed. Second, our research only focuses on grape detection and counting, upstream tasks in digital viticulture. However, we do not apply the grape detection results to downstream tasks such as predicting grape picking points and actual grape productions, which are more relevant to practical production activities.

Developing an efficient control method for each basic module and constructing its internal conditions is vital to transitioning a harvesting robot from a functional prototype to a practical machine. Therefore, this study [29] tackles efficient locomotion, picking, and seamless integration. They built a system and proposed a set of algorithms for locomotion-destination estimation, realtime self-positioning, and dynamic harvesting. They have established a solid coordination mechanism for continuous locomotion and picking behavior. As a result, the success rate of positioning was 95.8%; at 17 destinations of dragon fruit in an orchard, the robot carried out 24 positioning operations and obtained 14 successful movements, where the time consumption was 7.71 s. whereas the fig orchard had a 76.9% picking success rate, where at 11 destinations, seven successful movements, one collided with the branch, and three lost visual tracking. Each method offers distinct advantages such as improved accuracy, adaptability to varying conditions, and enhanced picking efficiency to operate a robot autonomously and continuously. This method is limited because the module works better in daytime conditions than in night vision.

3. MATERIALS AND METHODS

Counting fruit on trees where it is hidden under a leaf or branch and occluded with another fruit avoids double counting [30]. To solve this research problem, we used multiclass classification with YOLOv5 and YOLOv7 versions and compared each. We will obtain better detection and counting accuracy with a clear solution using a better YOLO version.

3.1. ACQUISITION OF DATA

The images in this study are of Vikarabad District, Telangana State, India. The variety of mangoes is 'Banganapalle,' also known as 'Benishan' [31]. The video captured the mango fruit's maturity stage between 06:00 AM and 08:00 AM, which is a perfect time for capturing images under natural lighting conditions. They used an iQOO Z3 5G mobile phone to capture the video, with a duration of 45 s to 75 s around the tree in a clockwise direction, which covered 360 °of the tree. The data storage was from 70MB to 160 MB with 1080 pixels ×1920 pixels of resolution in a portrait capturing way and saved as mp4 video. The approximate speed of the captured video was 30 fps. The video was acquired under natural daylight, while the outdoor environment was warm. (Fig 2) shows a few images of the data frames. Split the mp4 video into frames, and a video of 152 MB of storage data splits into 2228 image frames of on-tree mango fruit at approximately 30 fps. After splitting the 152 MB video into 2228 frames, the size of these 2228 frames was 2.03 GB of storage.

3.2. ANNOTATION

To create the images as a dataset, we must annotate every image into a .jpg file and a .txt file. Among these 360 image frames, 288 were for training, and 72 were for testing and validation purposes, with an 80:20 percent ratio. An open-source tool, LabelImg [32], was used to annotate the images in this research. It is straightforward to use and has better options in labeling formats.

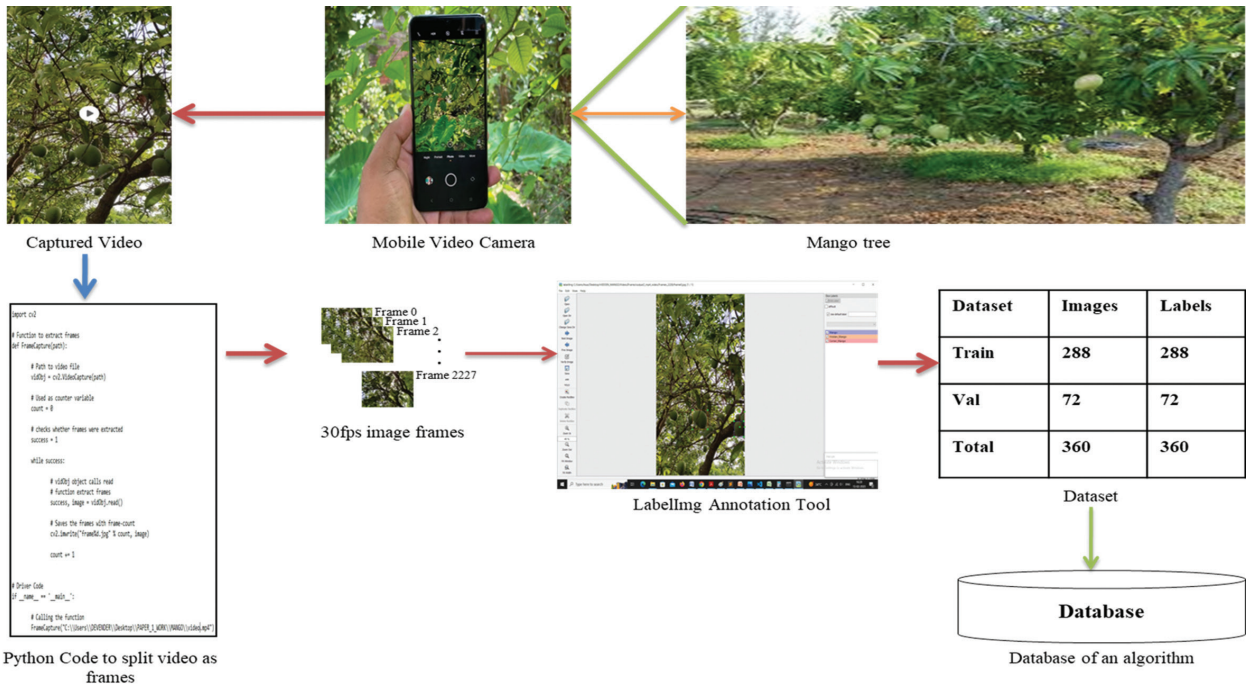

Fig. 2. Create a dataset and store it in YOLO format data to make predictions

3.2.1. LabelImg: LabelImg is written in Python with Qt as its graphical interface and is a graphical image annotation tool. Tzutalin created the popular image annotation tool LabelImg with some contributors, and now it is a developed tool. It is a part of the Label Studio Community [32]. Annotations of image data saved in XML files in the PASCAL VOC format (the tool LabelImg-1.8.6, released on October 10, 2021) also support the YOLO and CreateML formats. The working procedure in Fig. 2 shows that the image will be selected and asked for its format to store for future research. After proper format allocation, the user must draw a label for the object to annotate. Therefore, we used YOLO format with the text file content as "class; x-center; y-center; width; height." This text file will saved in the system, including images, and the user will use it to train the data using pictures.

3.3. METHODS

The flow diagram contained the input, algorithm, database, detection, counting, and output steps. The first step in Fig. 3 is the input part, which obtains the image as input for the algorithm, starts the process on the given input image, and works for detection with the help of a trained dataset from the database created by the researcher. After training the dataset with the help of the YOLOv7 algorithm, mango fruits were detected on the input image and counted using the DeepSORT-count algorithm.

Step 1: Input image.

Step 2: Use the given image for testing and validation based on the dataset already generated by the user and stored in the database.

Step 3: Check whether the given input image class is trained perfectly.

Step 4: Use the trained dataset and the input image with YOLOv5 and YOLOv7 to detect and count the classes available in the input image.

Step 5: If there are more than two classes, the output is given in a multiclass classification of the detected object and count.

Step 6: Output is the Number of Classes with the number of fruits, as shown in (Fig. 4) in a single frame.

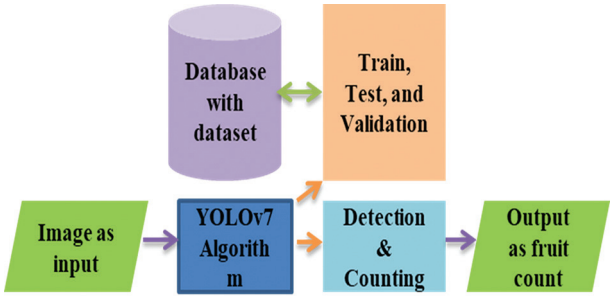

Fig. 3. Counting fruits using YOLOv7: an image as input, and the number of fruits is output

3.3.1. Experiment Platform: Windows 10 Pro 64-Bit, Core 19-9900KF CPU @ 3.60GHz, 32 GB RAM, Dedicated 8 GB Memory, NVIDIA GeForce RTX 2080 SUPER GPU. The model framework was PyTorch with CUDA 11.6, cudnn 11.6, and Python 3.9.0. The original YOLOv7 model used YOLOv7.pt and YOLOv7-tiny.pt for pretraining and retraining based on the pretraining results. The parameters for training were as follows: image input size 640×640, epoch 500, initial learning rate 0.01, and batch size 8.

Fig. 4. Hidden-Mango=20; Mango=4; and Corner-Mango=4; Total=28 Mangoes on tree.

3.3.2. Evaluation and performance of the model: Several evaluation indicators are available to help ascertain and validate the model's functionality, including the confusion matrix, precision, recall, mAP, tradeoff, F1-score, mean, median, mode, variance, standard deviation, and root mean square error. Accuracy is applicable only for binary classes, not for multiclass classification, and the confusion matrix will take care of multiclass classification.

3.3.3. Confusion Matrix: A confusion matrix defines a classification algorithm's performance. It visualizes and summarizes its performance. *TP*: True Positive, *FP*: False Positive, *FN*: False Negative, *TN*: True Negative.

Accuracy: Accuracy is the ratio of the exact classified on-tree mango samples to the number of actual on-tree mango samples in the image for binary class classification.

$$Accuracy = \frac{(TN + TP)}{(TN + FP + FN + TP)} \quad (1)$$

Precision is the ratio of correctly predicted positive on-tree mango classes for all items to be positive.

$$\text{Pr}\ ecision\ = \frac{(TP\ )}{(TP\ +\ FP\ )} \quad (2)$$

*F*1-*Score*: The *F*1-*Score* is a performance metric considering precision and recall values. It will calculated using the two metrics' harmonic mean.

$$F1 - Score = \frac{(2 * \text{Pr}ecision * \text{Re}call)}{(\text{Pr}ecision + \text{Re}call)} \quad (3)$$

3.3.4. SORT—Simple Online and Realtime Tracking: SORT stands for simple online and realtime tracking, an approach for tracking multiple objects with the help of any deep learning algorithm. To obtain the count of the on-tree fruit, we used YOLOv5 and YOLOv7 with SORT. SORT can track an object for extended periods to determine its occlusions [10].

3.3.5. YOLOv5n Algorithm for on-tree fruit counting: After a few days of YOLOv4, YOLOv5, a PyTorch-based approach, was released on May 27, 2020. In YOLOv5, some sub-variants based on 640 image size YOLOv5n (Nano), YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), and YOLOv5x (extra-large); based on an image size of 1280 are YOLOv5n6, YOLOv5s6, YOLOv5m6, YOLOv5l6, and YOLOv5×6 [33]. All YOLOv5 versions of the first two sub-variant models, called YOLOv5n and YOLOv5s, are used in this study. They then worked on both models and found that YOLOv5n works better for their self-prepared dataset with better accuracy and inference time than YOLOv5s.

Divide the YOLOv5n model into four regions: the input, backbone, neck, and head regions. Here, in the input region, the model takes an image of 1×3×640×640 and calculates the best-fit anchor box value according to the custom dataset. The convolutional layers and spatial pyramid pooling fast [34] were the backbone of this model. The combination of Feature Pyramid Networks (FPN) [35] and Path Aggregation Network (PAN) network layers [36] acts as the neck region in this model. The three detection heads with 1×3×80×80×8, 1×3×40×40×8, and 1×3×20×20×8 scale integration will give the predicted bounding box information to the final output [37].

3.3.6. YOLOv7 Algorithm for counting: YOLOv7 works excellently as an object detector with a high speed from 5 to 160FPS and has the highest accuracy of 56.8% AP using the MS-COCO dataset with 30FPS or higher on a GPU machine. YOLOv7 is a very balanced object detector compared with all known object detectors in speed and accuracy. YOLOv7 architecture [38], the pipeline has three significant parts: backbone,

encoder, and decoder. Again, the architecture of YOLOv7 consists of three parts: the input, backbone, and head. The 640×640 image was used in the input part, whereas, in backbone feature extraction, the image at the head strengthened the feature extraction network and made it ready for prediction.

3.4. COUNTING MANGO ON THE TREE

Fig. 5 shows a combination of several procedures: the input, output, annotation, database, and processing components, respectively.

Input: Eight image frames of each tree are considered input images; these images are 1080 pixels wide by 1920 pixels high and between 900 KB and 1100 KB in size.

Annotation: Based on chapter (3.2), the researcher labeled fruit annotations.

Database: Based on the above annotation procedure, the created dataset was stored in this database for future reference and use during training, testing, and validation. If the dataset images are not maintained and the labels should be the same, then algorithm failure is possible.

- **Processing:** The backbone and neck network-related procedures will processed as a processing model, which is the immediate step to the input part.
- **Output:** The output part is immediately adjacent to the neck region of the YOLO version. In these models, the three head-bounding boxes provide output predictions.

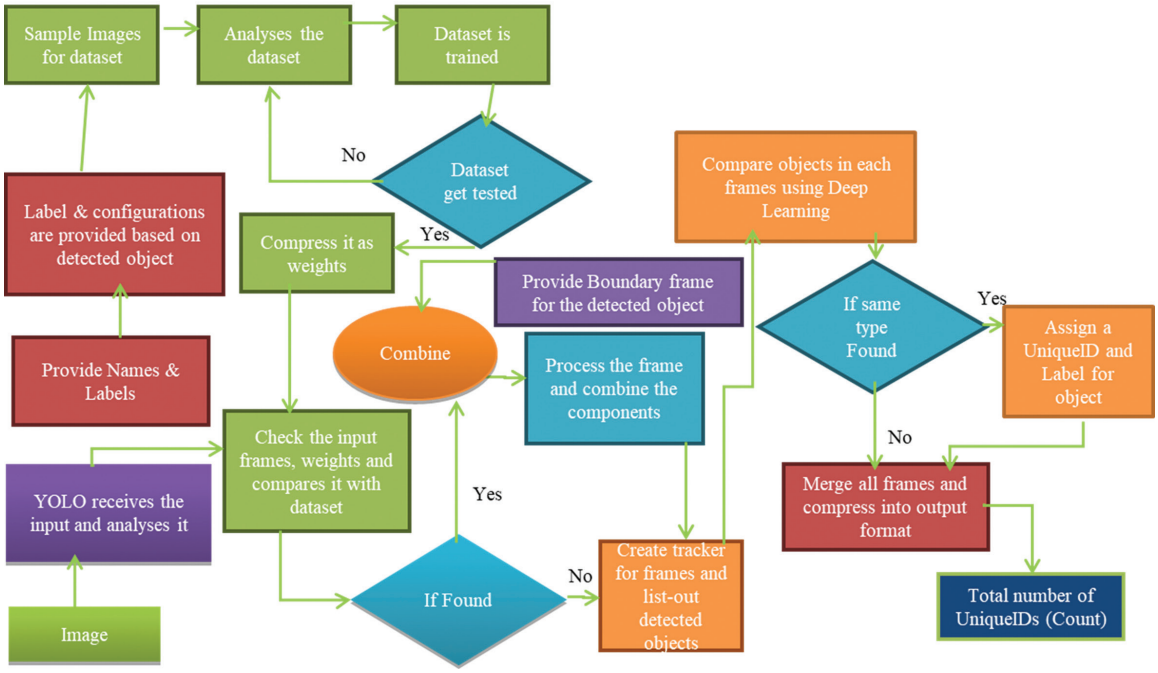

Fig. 5. Fruit counting with YOLOv7 algorithm using an image as input and number of fruits as output based on stored database data

4. RESULTS

4.1. ONE TREE FRUIT COUNT IN ORCHARD

One tree was selected, and 130 fruits were counted manually on May 30, 2022. Based on (Chapter 3.3.1), we used multiclass classification, a 4-side image model on YOLOv5n, YOLOv5s, YOLOv7, and YOLOv7-tiny and worked; the results were stored and performed the same with 8-side model got the results (Eq. (5)) and finally evaluated the data for a more suitable model. Among the four YOLO sub-versions, YOLOv7 is very close to the reality of 97.7% accuracy and achieves a lower average inference of just 17.112 ms of time. Counting the fruit using the 4-side image of the tree procedure is too far from the actual result, but the 8-side image of the tree is too close to the reality of the tree fruit count. Consequently, employing the 8-side model is preferable to the 4-side model when counting the fruit on the tree. Selected One hundred seventy-nine images for training from 224 images; the remaining 20% were for testing and validating the results in Table 1.

Table 1. On-tree fruit counting with YOLOv5 and YOLOv7 models of a single tree.

Models	Manual Count	4-Side Count	Inference (s)	8-Side Count	Inference (s)
YOLOv5n	130	66	0.0384	135	0.03927
YOLOv5s	130	71	0.07155	145	0.07292
YOLOv7	130	60	0.01575	133	0.01711
YOLOv7-tiny	130	71	0.01075	141	0.01200

$$4 - Side\_Count = \sum_{i=1}^{4} (F_{TM} - F_{CM}) \quad (4)$$

$$8 - Side\_Count = \sum_{i=1}^{8} (F_{TM} - F_{CM}) \quad (5)$$

F_{TM} is the frame's total mango count, and F_{CM} is the corner mango of the frame.

4.2. MULTIPLE TREE FRUITS COUNT IN AN ORCHARD

4.2.1. Manual Counting of On-Tree: Some trees in an orchard were counted based on our eight-sided model. Then, the number of fruits available on each tree and the resulting count can be validated. They picked a few fruits off the tree, and some were in different structures, so there were ups and downs in the graph in the manual count flow, as shown in (Fig. 6).

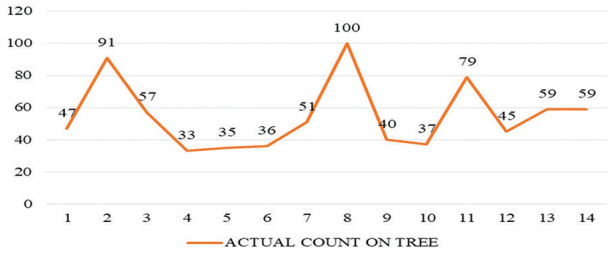

Fig. 6. On trees manually, there is a fruit count of fourteen trees

4.2.2. Avoiding Double Count using the Eight-Sided Model: If images are taken precisely on the eight sides of the tree, then the procedure below will work. Because eight exact images of the tree based on eight directions, such as east, south, west, north, south-east, south-west, north-east, and north-west, will give only eight frames. At the corner, mangoes will counted twice; to avoid this problem, we used an eight-sided model (Eq. (5)). However, in this research, we split a video into frames to obtain eight-sided images; thus, the total frames were divided by eight, and then eight were obtained. Then, the frame in the east is the first frame, and the last in the northeast frame is the eighth frame used as the eight-direction images.

- YOLOv7: Considered 14 trees of an orchard and performed the task using the YOLOv7 model, and observed as in Fig. 7, which has a significant impact on the count of the tree using the YOLO model, where the predicted count is very close to the actual count of fruits on the tree with a good accuracy of 95.48% with 17 ms of inference time.
- YOLOv7-tiny: YOLOv7-tiny also used the same dataset and performed the same task as in the YOLOv7 model, as shown in Fig. 7, where YOLOv7-tiny also performed the same as YOLOv7, with some differences. The detection and counting accuracy was only 94.1%, with an average inference time of 16 ms.
- YOLOv5n: YOLOv5n also used the same dataset and performed the same task as in the YOLOv7 model, as shown in Fig. 7, where YOLOv5n also performed well, with an average inference time of 94.1% with 104.5 milliseconds of average inference time.
- YOLOv5s: With the help of the same dataset, performed the same task as in the YOLOv7 model and observed as in Fig. 7, where YOLOv5s also performed well, with an average inference time of 97.2% with 85.69 milliseconds of average inference time.

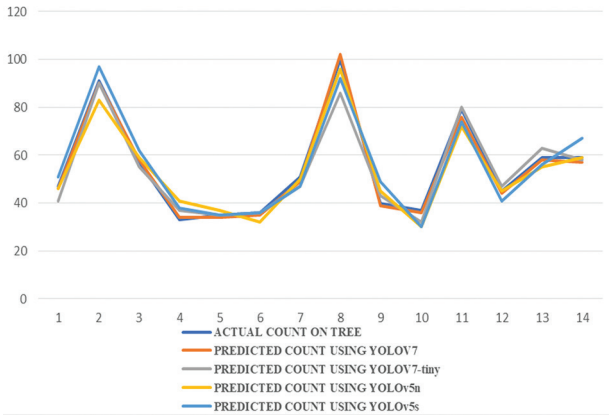

Fig. 7. Comparison of four models and validation of fourteen trees with actual count using eight-sided models

5. DISCUSSION

The researchers used four models and compared them in different ways. With the help of 360 image frames and their annotated labels, only the model trained at different epochs with batch sizes of 8 and 16. Then, 288 image frames were used only for training purposes, and the remaining for testing and validating the YOLOv7 and YOLOv7-tiny models. Trained these models, compared them at 100 and 500 epochs, and observed that YOLOv7 is the best model with good accuracy and a meager inference time for tree fruit count prediction.

Three procedures determine the best time to capture an image under natural lighting conditions. In these three procedures, we used 96 images of datasets; however, based on their lighting conditions, the storage size differed automatically, as shown in Table 2.

Table 2. Lighting conditions validation accuracy in a day to capture images

Time	Dataset Size	Storage size	Detection Accuracy
Morning	96 Images	363MB	84%
Day	96 Images	375MB	42%
Evening	96 Images	359MB	57%

We considered 06:00 AM to 09:00 AM as the morning time, 10:00 AM to 03:00 PM as the daytime, and 04:00 PM to 06:00 PM as the evening time. We captured images, prepared a new dataset for each time of the day, and trained and evaluated them.

Natural lighting was impossible at night, except during the full moon day. It is necessary to use artificial light for capture, which is compatible with the device's camera.

Consider three different time-captured images and the number of fruits in these images counted manually. Images were compared with the best training result of each timing separately, and then a table was created for these results. We then compared the images and the timing of the picture capture, as shown in Table 3, and observed tremendous results with the help of YOLOv7. Based on the results of Tables 2 and 3, we propose that morning image capturing is the best time to capture images of trees in an orchard.

Table. 3. Comparison of best fruit image capturing time of on-tree

Time	Morning	Day	Evening	Actual Count
Morning Images	21	9	11	25
Day Images	8	6	5	14
Evening Images	16	8	11	19

6. CONCLUSION

Implementing video capture and YOLOv7 for fruit counting offers significant advancements in agricultural technologies. Employing a strategic frame selection process ensures accurate and efficient fruit counts while minimizing redundancy and double counting. This innovation enhances yield estimation and optimizes resource allocation and early issue detection, reducing labor costs and increasing overall productivity. Its scalability and adaptability make it suitable for various orchards and commercial agricultural operations. Ultimately, this approach empowers farmers with precise, realtime data, enabling informed decision-making and contributing to improved profitability and sustainable farming practices. Owing to the rapid explosion of data in agriculture and horticulture sciences, a new trending computer science area, deep learning technology, has become a hot research focus for a new era in artificial intelligence. To determine the actual count of on-tree mango fruits, the researcher performed experiments with four algorithms, YOLOv5n, YOLOv5s, YOLOv7, and YOLOv7-tiny, using an eight-sided imaging technique around the tree, which showed that YOLOv7 performed the best about accuracy and inference. In this study, a deep learning framework was compared and applied to a computer vision algorithm for fruit detection and counting of trees using videos and images. It also presented the most suitable time to capture the images for better detection in the morning, daytime, and evening. It proposed that capturing morning-time images under natural conditions is the best time for on-tree fruits. In this on-the-spot situation, the video split image-converted dataset model helps count the object using the most miniature image training with better accuracy. Even though this method counts ideally, it also has some limitations; the video should be captured only in the forward direction, not in slow motion, and the reverse direction while capturing the video.

7. REFERENCES

[1] P. Wang, Y. Luo, J. Huang, S. Gao, G. Zhu, Z. Dang, J. Gai, M. Yang, M. Zhu, H. Zhang, X. Ye, "The genome evolution and domestication of tropical fruit mango", Genome Biology, Vol. 21, 2020, pp. 1-17.

[2] H. Kim, M. J. Castellon-Chicas, S. Arbizu, S. T. Talcott, N. L. Drury, S. Smith, S. U. Mertens-Talcott, "Mango (Mangifera indica L.) polyphenols: Anti-inflammatory intestinal microbial health benefits, and associated mechanisms of actions", Molecules, Vol. 26, No. 9, 2021, p. 2732.

[3] A. Wall-Medrano, F. J. Olivas-Aguirre, J. F. Ayala-Zavala, J. A. Domínguez-Avila, G. A. Gonzalez-Aguilar, L. A. Herrera-Cazares, M. Gaytan-Martinez, "Health benefits of mango by-products", Food wastes and by-products: Nutraceutical and Health Potential, 2020, pp. 159-191.

[4] A. J. Mobolade, N. Bunindro, D. Sahoo, Y. Rajashekar, "Traditional methods of food grains preservation and storage in Nigeria and India", Annals of Agricultural Sciences, Vol. 64, No. 2, 2019, pp. 196-205.

[5] Y. Xue, N. Huang, S. Tu, L. Mao, A. Yang, X. Zhu, X. Yang, P. Chen, "Immature mango detection based on improved YOLOv2", Transactions of the Chinese Society of Agricultural Engineering, Vol.34, No. 7, 2018, pp. 173-179.

[6] P. J. Ramos, F. A. Prieto, E. C. Montoya, C. E. Oliveros, "Automatic fruit count on coffee branches using computer vision", Computers and Electronics in Agriculture, Vol. 137, 2017, pp. 9-22.

[7] M. Rahnemoonfar, C. Sheppard, "Deep count: fruit counting based on deep simulated learning", Sensors, Vol. 17, No. 4, 2017, p. 905.

[8] S. Kamkar, R. Safabakhsh, "Vehicle detection, counting and classification in various conditions", IET Intelligent Transport Systems, Vol. 10, No. 6, 2016, pp. 406-413.

[9] 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, 2022, p. 576.

[10] N. Wojke, A. Bewley, D. Paulus, "Simple online and realtime tracking with a deep association metric," Proceedings of the IEEE International Conference

on Image Processing, Beijing, China, 17-20 September 2017, pp. 3645-3649.

[11] Y. Qiao, Y. Hu, Z. Zheng, H. Yang, K. Zhang, J. Hou, J. Guo, "A counting method of red jujube based on improved YOLOv5s", Agriculture, Vol. 12, No. 12, 2022, p. 2071.

[12] A. I. B. Parico, T. Ahamed, "Real time pear fruit detection and counting using YOLOv4 models and deep SORT", Sensors, Vol. 21, No. 14, 2021, p. 4803.

[13] S. Bargoti, J. P. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards", Journal of Field Robotics, Vol. 34, No. 6, 2017, pp. 1039-1060.

[14] U. O. Dorj, M. Lee, S. S. Yun, "An yield estimation in citrus orchards via fruit detection and counting using image processing", Computers and Electronics in Agriculture, Vol. 140, 2017, pp. 103-112.

[15] A. Koirala, K. B. Walsh, Z. Wang, C. McCarthy, "Deep learning for realtime fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO", Precision Agriculture, Vol. 20, No. 6, 2019, pp. 1107-1135.

[16] O. E. Apolo-Apolo, M. Pérez-Ruiz, J. Martinez-Guanter, J. Valente, "A cloud-based environment for generating yield estimation maps from apple orchards using UAV imagery and a deep learning technique", Frontiers in Plant Science, Vol. 11, 2020, p. 1086.

[17] J. P. Vasconez, J. Delpiano, S. Vougioukas, F. A. Cheein, "Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation", Computers and Electronics in Agriculture, Vol. 173, 2020, p. 105348.

[18] H. Mirhaji, M. Soleymani, A. Asakereh, S. A. Mehdizadeh, "Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions", Computers and Electronics in Agriculture, Vol. 191, 2021, p. 106533.

19] F. Gao, W. Fang, X. Sun, Z. Wu, G. Zhao, G. Li, R. Li, L. Fu, Q. Zhang, "A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard", Computers and Electronics in Agriculture, Vol. 197, 2022, p. 107000.

[20] S. Lu, W. Chen, X. Zhang, M. Karkee, "Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation", Computers and Electronics in Agriculture, Vol. 193, 2022, p. 106696.

[21] Y. Zhang, W. Zhang, J. Yu, L. He, J. Chen, Y. He, "Complete and accurate holly fruits counting using YOLOX object detection", Computers and Electronics in Agriculture, Vol. 198, 2022, p. 107062.

[22] Z. Wu, X. Sun, H. Jiang, W. Mao, R. Li, N. Andriyanov, L. Fu, "NDMFCS: An automatic fruit counting system in modern apple orchard using abatement of abnormal fruit detection", Computers and Electronics in Agriculture, Vol. 211, 2023, p. 108036.

[23] C. Wang, C. Li, Q. Han, F. Wu, X. Zou, "A performance analysis of a litchi picking robot system for actively removing obstructions, using an artificial intelligence algorithm", Agronomy, Vol. 13, No. 11, 2023, p. 2795.

[24] Y. Tang, J. Qiu, Y. Zhang, D. Wu, Y. Cao, K. Zhao, L. Zhu, "Optimization strategies of fruit detection to overcome the challenge of unstructured background in field orchard environment: A review", Precision Agriculture, Vol. 24, No. 4, 2023, pp. 1183-1219.

[25] G. Yu, R. Cai, Y. Luo, M. Hou, R. Deng, "A-pruning: a lightweight pineapple flower counting network based on filter pruning", Complex & Intelligent Systems, Vol. 12, No. 2, 2024, pp. 2047-2066.

[26] Y. Feng, W. Ma, Y. Tan, H. Yan, J. Qian, Z. Tian, A. Gao, "Approach of Dynamic Tracking and Counting for Obscured Citrus in Smart Orchard Based on Machine Vision", Applied Sciences, Vol. 14, No. 3, 2024, p. 1136.

[27] X. Zhai, Z. Zong, K. Xuan, R. Zhang, W. Shi, H. Liu, Z. Han, T. Luan, "Detection of maturity and counting of blueberry fruits based on attention mechanism and bi-directional feature pyramid network", Journal of Food Measurement and Characterization, 2024, pp. 1-16.

[28] C. Yang, T. Geng, J. Peng, Z. Song, "Probability map-based grape detection and counting", Computers and Electronics in Agriculture, Vol. 224, 2024, p. 109175.

[29] M. Chen, Z. Chen, L. Luo, Y. Tang, J. Cheng, H. Wei, J. Wang, "Dynamic visual servo control methods for continuous operation of a fruit harvesting robot working throughout an orchard", Computers and Electronics in Agriculture, Vol. 219, 2024, p. 108774.

[30] M. Stein, S. Bargoti, J. Underwood, "Image based mango fruit detection, localisation and yield estimation using multiple view geometry", Sensors, Vol. 16, No. 11, 2016, p. 1915.

[31] C. Zheng, P. Chen, J. Pang, X. Yang, C. Chen, S. Tu, Y. Xue, "A mango-picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard", Biosystems Engineering, Vol. 206, 2021, pp. 32-54.

[32] L. Tzutalin, "LabelImg", https://github. com/tzutalin/labelImg (accessed: 2020)

[33] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, J. Fang, K. Michael, D. Montes, J. Nadar, P. Skalski, Z. Wang, "ultralytics/yolov5: v6. 1-tensorrt, tensorflow edge tpu and openvino export and inference", Zenodo, 2022.

[34] D. Xu, H. Zhao, O.M. Lawal, X. Lu, R. Ren, S. Zhang, "An automatic jujube fruit detection and ripeness inspection method in the natural environment", Agronomy, Vol. 13, No. 2, 2023, p. 451.

[35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21-26 July 2017, pp. 936-944.

[36] W. Wang, E. Xie, X. Song, Y. Zang, W. Wang, T. Lu, G. Yu, C. Shen, "Efficient and Accurate Arbitrary-Shaped Text Detection With Pixel Aggregation Network", Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October - 2 November 2019, pp. 8439-8448.

[37] W. Yang, X. Ma, W. Hu, P. Tang, "Lightweight blueberry fruit recognition based on multi-scale and attention fusion NCBAM", Agronomy, Vol. 12, Vo. 10, 2022, p. 2354.

[38] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17-24 June 2023, pp. 7464-7475.