International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-10, Issue-7; Jul, 2023
Journal Home Page Available: https://ijaers.com/
Article DOI:https://dx.doi.org/10.22161/ijaers.107.8

Fruits and Vegetables Detection using YOLO Algorithm

S. Kanakaprabha[1], Dr. Gaddam Venu Gopal[2], D. Kaleeswaran[3], D. Hemamalini[3], Dr. G. Ganeshkumar[5]

[1]Department of Computer Engineering, Rathinam Technical Campus, Coimbatore, Tamilnadu, India
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram,Guntu.
[3]Department of Computer Engineering, Rathinam Technical Campus, Coimbatore, Tamilnadu, India
[4]Department of Computer Science and Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India
[5]Department of Information Technology, Hindusthan Institute of Technology, Coimbatore, Tamilnadu, India

Received: 08 Jun 2023,

Receive in revised form: 10 Jul 2023,

Accepted: 18 Jul 2023,

Available online: 31 Jul 2023

***Keywords*— Dark Flow, Fruit, OpenCV, Vegetable, YOLO**

***Abstract*—*The robotic harvesting platform's fruit and vegetable detection system is crucial. Due to uneven environmental factors such branch and leaf shifting sunshine, fruit and vegetable clusters, shadow, and so on, the fruit recognition has become more difficult in nowadays. The current method in this work is used to detect different types of fruits and vegetables in different size and shape. This method makes the use of OpenCV, Dark Flow, a TensorFlow variant of the YOLO technique. To train the necessary of network, a range of fruits and vegetable pictures were input into the network. The photos were pre-processed using OpenCV to create manual bounding boxes around the fruits and vegetables before into the training. YOLO detection algorithm is used. In, this method more accurately and rapidly recognizes of an item in an image. After the network has been trained, the test input is sent into the bounding boxes surrounding the recognized fruits and vegetables will be displayed as a consequence.*

I. INTRODUCTION

The main factor in the agriculture sector with the highest cost demands. This is brought on by rising supply costs for items like electricity, irrigation water, and agrochemicals, among the others. Because of this, the horticulture sector and farm enterprises are suffering from thin profit margins. Under these circumstances, food production will need to increase to meet the rising demands of a growing world population, which will be a major problem in this future. Due to its greater endurance and repeatability, robotic harvesting has the potential to save labor costs while simultaneously enhancing the fruit quality. These factors have led to a rise in interest in deploying agricultural robots to harvest fruits and vegetables during the past three decades. It takes a lot of challenging tasks, including choosing and manipulating, to build these platforms. Although it is the first perception of the system that comes under later manipulation and grasping system, building a reliable fruit identification is a crucial first step toward fully automated harvesting robots. Suppose the fruits cannot be detected or seen, it cannot be gathered. This level is challenging because of a number of factors, such as changing illumination, occlusions, and situations in which the fruit seems visually similar to the background. To deal with these problems, as well as we require a highly discriminative feature representations and a generalised model that is robust to changes in brightness and perspective. Fruits and vegetables are essential for human diet as well as animals and other living things. The requirement for food is two times more than it was previously due to the ever-increasing population of all living creatures. Farmers must work extremely hard and long hours to meet such a large demand, and farms must be monitored at all hours of the day and night. The product is affected by climate change in addition to the expanding population. Untimely rain and sweltering heat hinder the farmers' arduous task.

Convolutional neural networks (CNN), recurrent neural networks (R-NN), fast R-NN, YOLO, and other techniques are available in this field and may be used to identify and recognize fruits and vegetables. Based on training data given to the network, You Only Look Once (YOLO) is an efficient object identification method. Each frame updates into the input are examined, and the necessary items are frame into quickly. Various techniques have been to identify into fruits and vegetables. In techniques such as CNN, RNN, and Fast RNN, only certain sections of interest are applied to identify objects within an image. The networks mentioned above do not employ a holistic approach. These regions of interest are sent into the Needed model's network, and only those things are found there are trained. When compared to the region-based algorithms, YOLO is extremely different. One convolution network is sufficient to retrieve the YOLO class probability and bounding box information.

Fig.1(a) *Fig.1(b)*

The photos used for YOLO identification are shown in Figures 1(a) and 1. (b). Yolo-based real-time application detection systems may be used to a variety of applications with good generalization. With a little amount of example picture data, it is easily adaptable to any sort of fruit and vegetable. methods that use a variety of data in both early and late fusion.

II. LITERATURE SURVEY

A non-destructive approach based on thermal imaging is suggested by S. Raka, et al. [1] for evaluating the interior and exterior quality of fruits. Analyzing the fruit surfaces of the heated properties allows for the precise determination of ripening conditions. The creation of an automated system can decrease the time-consuming human inspection tasks involved in fruit sorting.

According to Y. Yu *et al.* [2], the primary technical challenge to the implementation of robots for strawberry harvesting is the need for improved real-time performance in the localization algorithms. To locate the plucking point on the strawberry stem. By estimating the position of the fruit goal and this accuracy can be achieved by the fruit axes of orientation. In this survey, a novel strawberry result of robot for the ridge-planted berries is introduced, along with a rotating yolo r- yolo fruit pose estimator that increases the efficiency of picking point localization for the lightweight network.

Convolution neural network was replaced with mobilenet-v1 is backbone of the network for feature extraction. Alternative, rotation of the angle guideline was used to design the training set and establish the anchors, and then logistic regression and rotated anchors were used to predict the spin of the target fruits bounding boxes in a batch of 100 strawberry photos. This considerably boosted operating speed. The average identification rate and recall rate for the suggested model were 9443 percent and 9346 percent, respectively. The integrated controller of the robot processed eighteen frames per second, which showed strong real-time achievement in terms of actual identification and localization efficiency of choosing places. This study presents technical advice for enhancing the fixed controller of fruit picking robots target recognition, showing that the recommended design outperformed a variety of different target identification methodologies.

Osman, Y., *et al.* [3] A two-stage approach includes recognizing the fruits and then tracking them framework by framework. The principle of "You Only Look Once" is applied to identify threats (YOLO). Bounding boxes are collected from the finding and Non-Max Suppression (NMS) is utilized to produce the concluded detection. The tracking system is then supplied with the boxes. We use a Deep SORT algorithm that was especially developed to deal with fruits for tracking. Using box coordinates, the original image is cropped to eliminate each recognized object. ResNet, a convolutional neural network (CNN), then extracts features from the cropped image to build the feature map. By comparing the attributes of new and old detections using a distance metric, which links the two things with the smallest distance, new detections are connected to previous detections.

Input items with no associations are studied as branded different objects to be monitored. We maintain path of the fruits through-out the video frames to assure that we are conditional accurate they are initially observed. We determine the method using videos taken in an apple garden to demonstrate this approach's very effectiveness in the natural light. The decision show that fruit counting on real-time video grain can be performed with great precision. The new method works with all types of fruits and vegetables and doesn't require any modifications to the algorithms.

In this study, a prototype of an autonomous fruit harvesting robot built around a mobile chassis and a robotic arm is proposed by S. M. Mangaonkar et al. Our suggested architecture can recognize fruits using an object identification method and an image pre-processing module (YOLO v3). This study proposes a prototype of an autonomous fruit harvesting robot based on a robotic arm mounted on a mobile chassis, developed by S. M. Mangaonkar et al. [4]. Our suggested architecture is capable of recognizing fruits thanks to an image pre-processing module and an object detection algorithm.

K. R. B. Legaspi and colleagues [5] Whiteflies and fruit flies were identified and classified using YOLOV3. The analyst used a Raspberry Pi camera to acquire images, and also set up both desktop and online applications for viewing the images captured by the Raspberry Pi camera. The confusion matrix showed that the miniature had the overall accuracy of 83.07 percent in recognizing and recognizing fruit flies and whiteflies.

According to S. K et al. [6], The Regional Built Convolutional Neural Network (RCNN), Fast RCNN, and Faster RCNN are examples of pre-trained Deep Neural (DNN) representations. To detect fruits in an input image, the You Only Look Once (YOLO) V3 and the Single Shot Multibox Detector (SSD) were implemented on the RISC-V architecture. COCO datasets are used for pre-training to ensure uniformity across all DNN models. In terms of accuracy and inference efficiency, experimental results demonstrate that YOLO and SSD-Mobile Net outperform all existing DNN models for object recognition on the RISC-V architecture.

The team of Yogesh [7] The fruit quality detection technique described in this study was built on the basis of the form, size, and colour of the fruits' external features. Manual fruit monitoring is ineffective in the agricultural industry due to growing demand. Therefore, the agriculture sector needs a capable approach to support it in meeting customer demand. The recommended method makes advantage of a sturdy feature that is speeded up. The approach discusses object detection by eliminating the local feature from the segmented picture. Creating a flaw detection method that can be utilized to quickly extract features and descriptions is the goal.

A fruit identification technique is suggested by Z. S. Pothen et al. [8] that makes use of the fruit's surface's slow change in intensity and gradient orientation. For potential fruit sites, gradient orientation profiles and monotonically falling intensity profiles are both examined also named as means by either "seed spots" To categorized into potential fruit spots that pass the first filter, altered histogram of directed gradient is to combined with a pair of the depth comparison of texture caption with a random forest classifier. The effectiveness of the fruit recognition algorithms on the fruit's datasets using the human-labeled images on the ground truth. This methodology is to identify the size invariant, resistant to partial occlusions, to be precise than existing method for identifying potential fruit locations.

In order to address issues with human health, K. Roy, et al. [9] offer a method for segmenting rotting vegetables. Edge Detection, Color Based Segmentation, and Marker Based Segmentation were three segmentation techniques that delivered effective and beneficial outcomes. The segmentation techniques outlined above successfully distinguish between rotting and healthy parts of a vegetable, allowing the diseased veggies to be distinguished from the healthy ones. Using an automated system to sort vegetables can save money on labor and increase accuracy for any company that manufactures food goods. On numerous levels, the ways to spot rotting veggies are examined.

An image-based technique is to identify the grade fruit size is presented by H. Dang and colleagues [10]. Following the acquisition of the fruits image, of several fruit characteristics are extracted into detection techniques. These characteristics are used to grade students. This integrated into grading system has to the benefits of high grading getting better accuracy, quick speed, and low cost, according to experiments. It is likely to be applied to yield-related detection and grading.

According to colour and form data, T. Gayathri Devi *et al.* [11] provide an image processing technique for completely independent separation and production forecast of fruits. The pre-processing procedure is started using the supplied fruits images. The picture is then determined to transform from RGB to HSV Color information to analyze the berry from the roots. The required colors may be hidden using colour edge detection. To diminish noise, of Gaussian filter is used. The picture outline is measured. The photographs are then processed using an image analysis technique. Fruit counting based on colour and shape is displayed in the result. The fruit and vegetables in the image are automatically segmented and counted using feature extraction and a circular fitting approach. Various fruits such as (orange/tangerine, pomegranate, apple, lemon, mango, and cherry) are used for automated conditional. Using the Open CV, the necessary image processing operations are completed.

Orange fruit pictures taken in natural illumination were segmented using edge-based and color-based detection techniques done by R. Thendral et al. [12]. The objective of this study was to locate and identify an orange in each of the twenty digitized fruit images that were casually selected from the Net. Edge-based segmentation is consistently outperformed by color-based segmentation. The computation is carried out using the MATLAB image-processing toolbox, and the computed outcomes are exposed in the segmented image results.

III. METHODOLOGY

Fruits and vegetables must be divided in order to be seen clearly against a backdrop of leaves and stems. Due to the variations in color and lighting, significant quantities of the occlusion, and other considerations, this test is difficult. Yolo is a real-time object tracking system that is offered as a technique. Yolo's primary premise is that you only look once when configuring a model for training. This approach then requires that you test the model with the necessary versions since the model versions change. Yolo has overtaken the market leader, CNN, in terms of popularity. Yolo and CNN are equivalent, although CNN does real-time object tracking less well. Both boundary boxes with a different CNN are analyzed by YOLO. YOLO is favored due of its speed.

Furthermore, unlike CNN's moving window and area proposed bill algorithms, it generates predictions while maintaining a global perspective. The secondary cause is YOLO's fast learning of generalizable representations of objects. One of the distinctive characteristics that the network discovers for each border box is the size of the boundary box and the many class choices. Only item classification with a quality greater than the edge is utilized to identify the images inside the box when a threshold has been specified. It is crucial to consider the output encryption technique YOLO employs. On the basis of the supplied picture segments, a N x N matrix is created. Even when there are numerous images are just one square of the grid, cell in the object's centre aids in predicting its existence. Each cell is surrounded by five bounding boxes, each of which has five distinct characteristics denoted by a letter (x, y, w, h, c).

The coordinates of the box's core cell are: (x, y). the bounding box's dimensions are (w, h). The confidence score is the last element that determines whether or not an item is in the box (c). If this is the case, the item is not included within the box, the score will be 0. Ideally, the element should be zero, but if it is present, it should be one. The formula used to determine the confidence factor favor's the intersection of the box and the accuracy over the union of the prediction box. Additionally, YOLO determines the probability for each category. Class probability refers to the possibility of each class that the object. The class possibility is the likelihood that the images in the case that fits to session. As a consequence, N x N x C possibilities, where C is the number of classes, are generated, with each cell forecasting one class probability.

Pre-processing is done on the pictures to get rid of noise and outliers, improve contrast enhancement, and speed up the algorithm. Although additional pre-processing techniques may be used in this approach, Non-Max Suppression is the major emphasis. The network topology resembles that of a typical CNN with 24 convolutional layers and two completely connected layers at the end. The Google Net idea is used to construct the YOLO network architecture. Fast YOLO is a quicker variation of YOLO that uses nine convolutional layers slightly than 24 and maintains all other limitations constant with exception of the system size.

The spatial arrangement of the grid cells that go into creating the bounding box makes YOLO less successful at recognizing little objects in big groupings. Since YOLO learns mostly from data, the system cannot recognize in advanced or changed shape of aspect ratios.

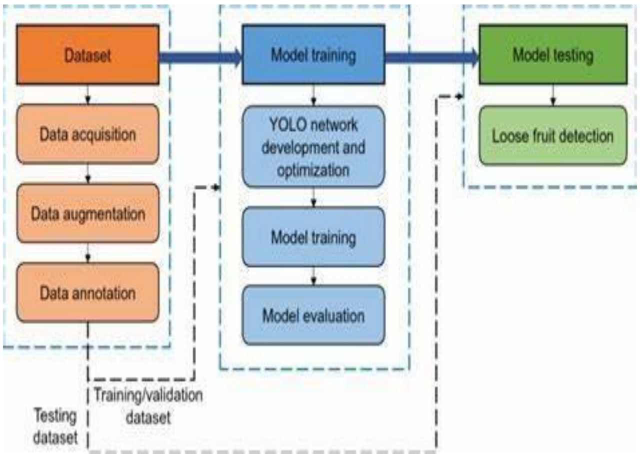

Fig.2. Yolo Architecture for Detecting Fruit and Vegetables

There are several methods to implement the YOLO approach, and our system uses Darknet, an open-source neural network framework. An open-source real-time computer vision library is called OpenCV. Since its creation in C++, it has been translated into a number of programming languages. We will use the OpenCV cv2 Python library. Drawing a bounding box that represents the upper left edge and lowest of the correct coordinates is one of cv2's four main features. Rectangles are drawn using the provided coordinates using this function. The expected class labels and confidence ratings are combined on each bounding box using a second algorithm. The final task is to scan the picture and look for class labels. The last approach is employed to read a movie from regional cache and give it class names. Additionally, it may be used to access the real-time video from a webcam or new computer hardware.

Tensor flow operates on both the CPU and the GPU equally; however, the Yolo GPU version runs quicker than the Yolo CPU version. According to GPU specs, Yolo operating on a GPU can analyze video at a rate of 40100 frames per second, whereas Yolo consecutively on a CPU can only manage 38 frames per second.

A. Training Image

The first phase in training is to get relevant photographs from the web, which are mostly pictures of distinct kinds of cucumber, apples, and capsicum. For a quick and precise categorization, train as many images as you can. From various web sources, a total of 100 photographs of each vegetable were collected, with 60 images being utilized for train and the 40 for testing. The network's distinctive qualities are determined by a number of variables in the YOLO configuration file; this variable quantity must be altered to match our contributions and productivities.

An xml document is created in which the top-left and bottom-right coordinates of all picture in the train dataset are listed. After that, a rectangle selector runs a Python script to complete this task. Moreover, the data is supplied to the system, a pre-trained dataset like yolo-tiny should be amount. The epochs are set to 300, overwriting the default of 1000, and the learning rate is set to 0.001. Each training phase will run the full batch of 16 photos done the unseen layers and adjust the masses accordingly.

The complete dataset has been divided into batches. All the epoch consists of 11 batches and 11 step due to the 180 pictures in the dataset. We have the same number of batches as there are epochs in our system. The mean error would have been calculated after each step, the weights would have been back-propagated, and all of the pictures in that all the batch subjected to all hidden layers.

As a consequence, all of the pictures have disappeared through the hidden layer once at the end of each epoch, allowing us to calculate the mean error. The regular error does not alteration training the times, the train is still or the learning rate is change. The average error is discovered to be between 4 and 6 after 137 training iterations and did not drop any other. At the conclusion of every 125 steps, Yolo will save the masses file in the resident manual. These mass files test our miniature utilizing it.

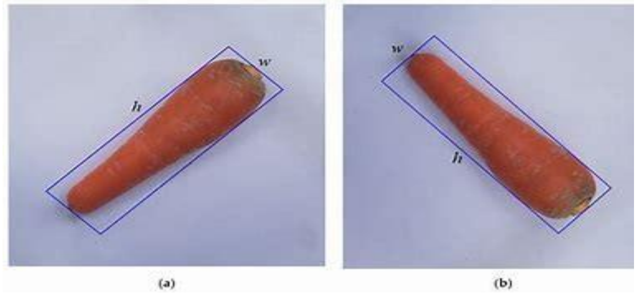

Fig.3(a) and Fig.3(b) boundary box for the practice pictures.

B. Testing Images

To accurately identify and categorize in presence of vegetable images, and that can be loads into the trained YOLO model and then lots into the dataset that has to be classified. The location and class of the veggie are also determined by the writing. A bounding box with the class label and confidence score is produced using OpenCV. The classical only identifies bounding boxes with a confidence score of 0.15 or above because a confidence threshold of 0.15 has been established. The same method is demonstrated in a movie to identify fruits and vegetables.

Every frame of the film is taken out, and using a threshold of 0.15, each picture is categorized like every other image. A lot of computing power is required for movie processing; else, object detection and classification on a film would be quite sluggish. In this algorithm is classified into the 3 modules into actual period, a script is printed into load the classical was trained into capture the video. The bounding boxes for all picture in individually edge by passing the edges to the algorithm. Here, too, the 0.15 cutoff is employed. For a fast comparison, the microchip type of Tensor Flow processes webcam footage at 4-6 frames per second.

IV. RESULT AND ANALYSIS

The dataset utilized comprises of 180 test photos of all three vegetables in different arrangements, such as horizontally, vertically, or in challenging lighting as a group or as a backdrop. Decreasing the learning rate to 0.001 will accelerate the training process. The program was set to run for 300 epochs, with 11 stages per epoch, as the batch size was set to 16. Following 100 epochs, the average error rate is 4-6. When the dataset was further trained, into 137th epoch, the average error did not significantly vary across successive stages.

The training remained halted and the weights that were obtained were saved in order to evaluate their accuracy. Then, this classical is tested into a range of sample photos. 70% of the veggies were correctly identified and classified more often than 70% of the time when a video was used as an input to the algorithm. Vegetables in a range of situations and orientations, including vertical, in bunches, and against complex backgrounds, could all be successfully identified and classified by the model. The model also produces a confidence score for every prediction, which was more than 50% for almost all of the photos.

Table 1. Fruits and vegetables' effectiveness

The quantity of pictures examined in fruits	50
Average degree of assurance	67.6%
Number of vegetables photographs that underwent testing	50
Average degree of assurance	67.6%
Number of pictures where different fruits may be seen	75%
Number of pictures in which different vegetables are visible	75%
Percentage of photos with false positives	50%
The proportion of photos with a confidence rating of at least 50%.	65%
Number of photos with a greater than 80% degree of confidence	20%

The YOLO algorithm works by dividing the image into a grid and predicting bounding boxes and class probabilities for each grid cell. This approach allows YOLO to detect multiple objects simultaneously and with high speed and accuracy. Keep in mind that the quality and accuracy of YOLO's detections depend on various factors, including the quality and diversity of the training data it was exposed to during its training phase.

As a result, the model was able to correctly identify and classify the the fruits and vegetables. With a high confidence score of over 50%, the model accurately detects cucumber, and it also correctly detects numerous cucumbers. When there are a lot of cucumbers, the model sometimes provides false positives by identifying half of a cucumber as a full cucumber. There are a few instances where the model fails to detect a vegetable due to the unknown orientation of a vegetable.

Fig.4. Images of several fruits that YOLO detected

Green mango was likewise confidently and effectively identified and identified. In the bare minimum number of photos, the model accurately classifies fruits and vegetables with a 65% accuracy rate. The model has a high degree of confidence in its ability to identify several green apples in a picture. The model correctly identifies green capsicum and categorises it, as well as various other capsicums in complicated backdrops.

V. CONCLUSION

A model for identifying fruits and vegetables has been developed, along with the recommended approach, which has been built, trained, and tested. Our algorithm can identify and classify 60–70% of the crop and can identify different vegetables in a single picture under a variety of limitations. 70 of the photographs were accurately categorized when the threshold was set to 015 since the bulk of the images were downloaded from the internet. The more effective training set is greater than the accuracy in categorizing each of the shots if we had utilized off-field images as our training examples. For this model to discriminate between foreground and background produce in an automated harvesting system, depth information is crucial. This may be done by utilizing 3D photos and altering the system so that it no longer uses 2D images for training characteristics like size, colour and texture.

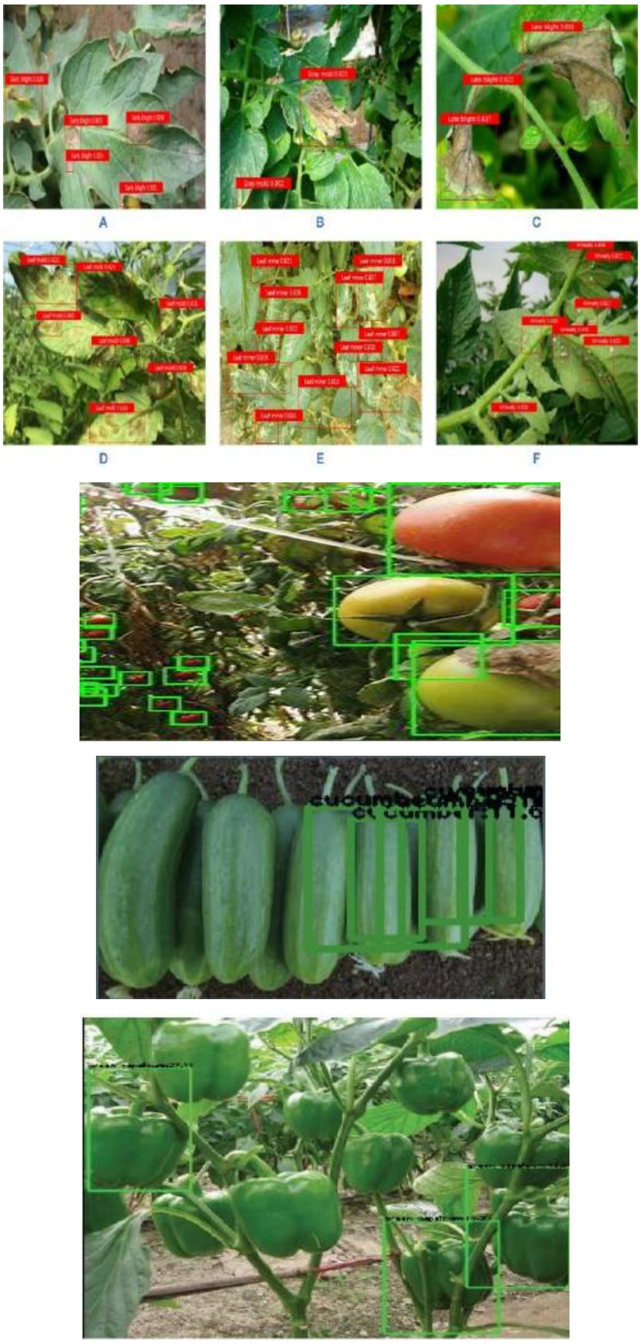

Fig.5. YOLO detection of various vegetable images

REFERENCES

[1] S. Raka, A. Kamat, S. Chavan, A. Tyagi and P. Soygaonkar, "Taste- wise fruit sorting system using thermal image processing," 2019 IEEE Pune Section International Conference (PuneCon), 2019, pp.1-4,doi: 10.1109/PuneCon46936.2019.9105726.

[2] Y. Yu, K. Zhang, H. Liu, L. Yang and D. Zhang, "Real-Time Visual Localization of the Picking Points for a Ridge-Planting Strawberry Harvesting Robot," in IEEE Access, vol. 8, pp. 116556-116568,2020,doi:10.1109/ACCESS.2020.3003034.

[3] Y. Osman, R. Dennis and K. Elgazzar, "Yield Estimation using Deep Learning for Precision Agriculture," 2021 IEEE 7th World Forum on Internet of Things (WF-IoT), 2021, pp. 542-550, doi: 10.1109/WF-IoT51360.2021.9595143.

[4] S. M. Mangaonkar, R. Khandelwal, S. Shaikh, S. Chandaliya and S. Ganguli, "Fruit Harvesting Robot Using Computer Vision," 2022 International Conference for Advancement in Technology (ICONAT),2022,pp.1-6,doi: 10.1109/ICONAT53423.2022.9726126.

[5] K. R. B. Legaspi, N. W. S. Sison and J. F. Villaverde, "Detection and Classification of Whiteflies and Fruit Flies Using YOLO," 2021 13th International Conference on Computer and Automation Engineering (ICCAE), 2021, pp. 1-4, doi: 10.1109/ICCAE51876.2021.9426129.

[6] S. K, A. K. Srivastava, S. Allam and D. Lilaramani, "Comparative analysis on Deep Convolution Neural Network models using Pytorch and OpenCV DNN frameworks for identifying optimum fruit detection solution on RISC-V architecture," 2021 IEEE Mysore Sub Section International Conference (MysuruCon), 2021, pp. 738-743, doi: 10.1109/MysuruCon52639.2021.9641594.

[7] Yogesh and A. K. Dubey, "Fruit defect detection based on speeded up robust feature technique," 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2016, pp. 590-594, doi: 10.1109/ICRITO.2016.7785023

[8] Z. S. Pothen and S. Nuske, "Texture-based fruit detection via images using the smooth patterns on the fruit," 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016, pp. 5171-5176, doi: 10.1109/ICRA.2016.7487722.

[9] K. Roy, S. S. Chaudhuri, S. Bhattacharjee, S. Manna and T. Chakraborty, "Segmentation Techniques for Rotten Fruit detection," 2019 International Conference on Opto-Electronics and Applied Optics (Optronix),2019,pp.1-4,doi: 10.1109/OPTRONIX.2019.8862367.

[10] H. Dang, J. Song and Q. Guo, "A Fruit Size Detecting and Grading System Based on Image Processing," 2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics, 2010, pp. 83-86, doi: 10.1109/IHMSC.2010.120

[11] T. Gayathri Devi, P. Neelamegam and S. Sudha, "Image Processing System for Automatic Segmentation and Yield Prediction of Fruits using Open CV," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication

(CTCEEC), 2017, pp.758-762, doi: 10.1109/CTCEEC.2017.8455137.

[12] R. Thendral, A. Suhasini and N. Senthil, "A comparative analysis of edge and color based segmentation for orange fruit recognition," 2014 International Conference on Communication and Signal Processing, 2014, pp. 463-466, doi: 10.1109/ICCSP.2014.6949884.