# Global Journal of Computer and Engineering Technology (GJCET)

**Vol. 2, No. 2, July, 2022** **ISSN: 2767-1933**

# FRUIT DETECTION FOR OPEN ORCHARD USING DEEP LEARNING APPROACH

Chitra Bhole, K. J. Somaiya Institute of Engineering and Information Technology, India (cbhole@somaiya.edu)
Chandani Joshi, Sir Padampat Singhania University, India (chandani.joshi@spsu.ac.in)

**ABSTRACT**

Agriculture plays an important role for people in both abstinence and social sectors. Fruit production is one of the basic requirements of all households. Nowadays, artificial intelligence plays an important technological tool that is used to achieve good results in today's society. Correct fruit detection is still a difficult task in vision based predictive analysis area. Deep Learning has challenging applications because of its capabilities to map patterns based on image inputs. Convolutional Neural Networks (CNNs) provide a deep learning approach for image classification. The objective of this paper is to build intelligent detection systems to improve tree fruit detection, fruit counting and yield estimation in agricultural applications, as well as to aid smart farming practices among fruit growers. The main goal is to implement a vision-based system which can detect and quantify fruits in open fields to predict yield in an open orchard.

**Keywords:** Deep Learning, ConvNets, Machine Learning, Fruits Classification, Fruits Detection, Yield Estimation.

## 1. INTRODUCTION

Agricultural work in the field of agriculture is one of the popular challenging factors. This is because of rising values of supplies like power, irrigation water and agro-chemicals. The challenge of meeting global population demand still faces food production, discussed by (Frilay et al., 2019). Harvesting with robots does bring reduction in labor costs and improve the quality of fruits significantly. Implementation of fruit detection system which is accurate is a key step towards fully automatic harvesting robots. When fruits are not detected or seen, the robot's sensors cannot pick them according to (Kondo et al., 2011) and (Bac et al., 2014). They found it difficult to tell if the fruit looked like the background. To achieve this goal, a generalized model is required that is robust to changes in the state of lights, viewpoints, which is highly discriminative. Fruit harvesting consists of two subsystems of a robotic vision system and a handling system. In this process, the vision system detects and locates the fruit before instructing the handler to separate it from the trees. However, due to the many variants, developing a seamless and effective fruit based detection system for orchards is a daunting task. Automation of yield mapping is a prerequisite for various types of precision agriculture and related work. In the event that automated operation of the system is not possible, manufacturers must rely on manual data collection to perform their tasks. In order to estimate yields, workers must randomly sample some trees and pool the results across the entire stand, which in turn provides inaccurate yield estimates. As a result, massive amounts of approaches have been experimented to automate the revenue mapping process. A typical yield mapping system collects photographs of fruit from one or both sides of a row of trees and then detects, counts and tracks them over many frames and then combines the fruit counts from the images. Fruit recognition algorithms have attracted academic interest in recent decades, and several different image processing strategies have been used. Most of these experiments have as their overarching goal robotic fruit picking, which was demonstrated by (Öztürk et al., 2016). However, picking fresh fruit is a delicate procedure. According to (Sarig, 2005), demanding harvesting is costly and time utilizing. Picking fruit by hand is also time-consuming. Human labor can be replaced by mechanical robots to solve these problems. Automated harvesting tasks save money as shown by (Thendral et al., 2014). In today's world, computer vision provides the fastest growing and most researched results. Computer vision refers to the technology and software used to capture and analyze images using a computer. With its speed, accuracy, reliability and objectivity, it has found a home on control lines. Computer vision is very good at detecting flaws in well-defined objects in systems such as inspection lines. Clear detection and counting of fruits in plants using images and computer vision is important not only to reduce the labor of manual phenotype data metrics, but also a crucial step in streamlining tasks such as harvesting. Deep learning techniques are very useful for gaining a complete understanding of an image and focusing on classifying different images, as well as trying to accurately prognosticate the localized objects for every image with their conceptualization. This process is called object detection. Deep learning leads to different sub-activities such as face detection, pedestrian detection and skeleton detection. A fundamental problem in computer vision is object detection, which provides valuable data for semantic understanding of videos and images. A few applications such as video image classification, human behavior analysis, face recognition and autonomous driving. In these conditions, it is very challenging to achieve perfect object detection using an additional object localization process. Deep learning supports state-of-the-art methodologies for many segmentation and classification tasks and is quite promising in challenging fields such as agriculture, which can cope better than conventional image processing methodologies with significant information diversity.

## 2. LITERATURE REVIEW

Accurate fruit detection and yield estimation is a challenging topic that seems to attain the inclination in apparent times in yield estimation of specialty crops. Complete process pipelines for single component ideas like fruit detection, counting are all included in the various methodologies that have been proposed. Precision farming and phenotype operations are preceded by automated yield mapping. Without an automated method, growers must collect data manually. Several trees must be randomly selected and counted, leading to error infused predicates for yield. For this reason, a significant amount of research has been devoted to standardizing yield mapping. In such systems, images are generally collected from the same or opposite side of a row of trees, and then the number of fruits in multiple images is tracked using fruit detection and counting. According to (Roy et al., 2018) do not report any tree rows where two different fruit counts can be obtained by merging counts from two sides of the row without using external navigation sensors. Instead of trying to build a model from the ground up, here they first generated a semantically rich representation of tree ranks and then combined their counts with a merged reconstruction. This process achieved yield estimation accuracies between 91.98% and 94.81%, respectively.

(Häni et al., 2020) research proposed a new method that detects and counts fruits, called as semantic segmentation-based approach. It shows extensive comparisons with other SOTA techniques. In this paper, they performed a direct comparison of detection and counting methods for multiple fruit species using identical datasets. Fruit detection results show that in most datasets, gaussian mixture models do outperform the deep learning-based models which does highlight the semi-supervised method. However, considering the counting of fruits, the deep learning approach outperforms all other approaches across all datasets. And finally, they achieved a yield estimation accuracy of 95.56% to 97.83% by combining the two methods.

A computer vision facet that maps the apple orchard yield using single camera images. The derived mechanism agnostic and never requires any special illumination according to (Roy et al., 2019). The fruit detection method has an accuracy range of 89 to 98 percent. The yield estimation method has an overall accuracy of 91.98% to 94.81%. The detection and counting of apples is done using the video taken from the apple orchard and the video is processed in stand-alone mode for further analysis.

Using a single shot detection mechanism, YOLOV3, which represents a convNet based model for predicting pineapple fruit and flowers in the accurate manner. Real farm images are collected and manually labeled with ground truth binding boxes to create the dataset. The network was fine-tuned after being trained on the COCO dataset. Fruits had 64% accuracy while flowers had 69% accuracy demonstrated by (Wang et al., 2020).

The work by (Gongal et al., 2016) had a 79.8% accuracy rate in terms of identification. This process used the positioning of the apples in 3D space so that the apples were not counted twice because the cameras saw them from both sides of the tree canopy. When identifying duplicate apples, the error rate was 21.1 percent. With two-sided imaging, the method was able to estimate crop load on trees with 82 percent accuracy, compared to 58 percent with one-sided imaging. Apple recognition techniques used, in which the features of the apple are considered to be features of the color and shape of the apple. Later, Hue, Saturation and Intensity (HSI) color space and equalization for histogram were applied to the color images, highlighting the color variations between the apples (foreground) and the leaves and branches (background). An Wiener filter that is adaptive with a neighborhood size of 5x5 was used to remove noises that were amplified during equalization of the histogram. In a study by (Sa et al., 2016), the VGG16 model with transfer learning tuning was used for the task of fruit detection using images acquired with two techniques, the first is color (RGB) and the second is Near-Infrared (NIR). A real-time fruit detection system based on Deep Convolutional Neural Networks (DCNN) was implemented and yield prediction accuracy for rockmelon with F1score

0.848, strawberry F1score 0.948 and apple F1score was found to be 0.948. The framework of this study includes an automatic label generation module and a fruit detector based on deep learning "LedNet". This experiment includes the comparison and relationship of LedNet, YOLOV3, YOLO-V3(tiny), as well as Faster-RCNN. The accuracy of LedNet found in this study ranges from 82.1 percent to 85.3 percent for fruit detection. Experimental results discussed by[Kang and Chen, 2020) with the LedNet model effectively detect apples in orchards in real time.

The paper by (Gené-Mola et al., 2019) describes a new approach for detection and locating apples from Fuji that use a scanner that is mobile ground-based laser. An MTLS consisting of a Velodyne VLP-16 LiDAR sensor coordinated with an RTK-GNSS receiver for satellite navigation was used to create a 3 dimensional point of the scene. This work developed a fruit detection algorithm and achieved a total number of fruits with a localization success rate of 87.5% which is a successful identification rate of 82.4%, and an F-score of 85.8% seamlessly. The values of detection rates are comparable to systems that are RGB-based, with the added advantage of receiving straight 3 dimensional fruit position data that is unaffected by conditions such as changes in sunlight.

Improvised variant of a deeply trained neural net, DaSNet-v2, designed by (Kang and Chen, 2020). The experimental results give DaSNet-v2 results with resnet-101 achieving 0.868, 0.88 and 0.873 in recall and detection accuracy and instance segmentation accuracy on fruit and 0.794 in branch segmentation accuracy. In terms of recall and instance detection accuracy and fruit segmentation accuracy, DaSNet-v2 with resnet-18 lightweight backbone scores 0.85, 0.87, and 0.866 and 0.775 respectively for branch segmentation accuracy.

(Yu et al., 2019) paper provides a fruit based detecting system for a berries robot operating for harvesting in an unstructured environment, which derives on a Mask-RCNN network. As a backbone network, Resnet50 was used in accordance for a Feature Pyramid Network (FPN) design to extract features. The detection of berry fruits with fluctuating illuminating frequencies in photos, the coercion of several fruits to each other, overlapping, occlusion and varied stages that involve growth, were all made possible by the trained model, which was particularly effective in this case. In the study by[Yu et al., 2019), a new type of neural network called Mask Region CNN was used. The backbone network was the Resnet50 network and the data was processed using the Feature Pyramid Network (FPN) architecture. The trained model excelled in excelled in berry detection with different illumination frequencies, several fruit adhesion, over-lapping, occlusion, and other difficult growth stages, as well as recognizing strawberry fruits with different light intensities in photographs. The accuracy found using this method was 95.41 percent.

Probabilistic partial shape matching proposed by (Lin et al., 2020) is detecting the fruit class sub-fragments and Hough transform is used to detect fruit. The SVM based fruit classification is used respectively. The accuracy of the proposed algorithm was found to be 78.3, 84.8, 74.5, 76.2, 80.7 and 91.9 percent for fruit identification and categorization in citric fruits and vegetable categorization respectively.

Research by (Roy et al., 2019) uses Principle Component Analysis (PCA) with Gaussian mixture models to detect and estimate fruit yield. Simulated and real datasets were compared using 3D reconstruction. The results showed an accuracy of 91.98% and 94.81% on these data sets.A technique for measuring mango fruit yield was proposed by[Kestur et al., 2019). In this article, the authors collected 40 photos with a resolution of 4000 3000 pixels, and it is difficult to analyze the same images in their original resolution. Multispectral dynamic imaging was used in the study of[Feng et al., 2019) to develop a strategy for detecting apples. The photographs were taken at a significantly high degree of contrast between the background and the fruit as a result of this, and the recognition accuracy improved to 92 percent.

Authors examined AlexNet (Krizhevsky et al., 2012), VGG16 (Simonyan and Zisserman, 2014), and GoogLeNet (LeCun et al., 1998) as CNN architectures for fruit identification, categorization, sorting, and quality assurance according to (Naranjo-Torres et al., 2020). They determined that to optimise feature extraction, the number of kernels and layers should be raised as the task complexity increases. (Wang et al., 2020) advocated using deep CNN to categorise fruits. Their 8-layer CNN architecture delivered total accuracy of 95.67 percent. The authors utilised an activation function, Rectified Linear Unit (Agarap, 2018) (ReLU) non-linearity function instead of a plain ReLU, and added a dropout layer before each fully linked layer. The Deep learning architecture is a data demanding architecture; therefore, performance depends on a lot of training data. But obtaining and identifying sufficient training data images is laborious. The features of all things must be identified accurately in order for achieving representations which are accurate of the images. A number of features that hand engineered, including Speeded-Up Robust Features in the paper by Woods N. C(2019), Histograms of Oriented Gradient by (Tan et al., 2018) and Scale Invariant Feature

Transform by (Tu et al., 2020), were employed to extract appropriate features from a given image in this context. It is necessary, however, to identify relevant characteristics in these methods, which is a time-consuming procedure. The outcomes of object recognition are strongly dependent on the characteristics that have been retrieved; otherwise, the system would be unable to provide reliable results. Obtaining complicated data features while dealing with a big amount of data is difficult using the approaches described above.

According to the study of (Liu et al., 2018), a novel strategy to segmenting fruits from video frames was developed, which makes use of semantic segmentation FCN based systems. Using the segmentation results, algorithm which is Hungarian with a loss function defined using Kalman kernel, was used to perform tracking on the data. To evaluate the suggested method, two fruit datasets (oranges and apples) were employed, each having a unique set of characteristics such as differences in depth and illumination, the likeness of color between foliage and fruit, and occlusion. Using the orange and apple datasets, we found that the error mean for each was 0.20 and 3.30 percent, the standard deviation for each was 7.80 and 4.10 percent, and the L1 error for each was 203 and 322 percent.

Wan and Goudos, 2020 suggested an approach for multi-class fruit detection based on a faster R-CNN architecture that is more efficient. Apple, mango, and orange were the fruits employed by the authors in their experiment. Each image in the collection has a 100 x 100-pixel dimension, and there are 820 apple photographs, 822 mango images, and 799 orange images in total. Two factors were considered: the penalty factor and the number of iterations. These were 200 and 5,000 respectively. Precision values of 92.51 percent for apples, 88.94 percent for mangoes, and 90.73 percent for oranges were obtained using the upgraded quicker R-CNN technique. A total of 50 microseconds were spent processing each image. While various Deep learning-based semantic segmentation designs, such as the YOLO, fast R-CNN, faster R-CNN, YOLOv2, and YOLOv3, were evaluated, it was found that our suggested method beat the other structures in terms of precision and processing speed.

Ganesh et al., 2019 proposed a subtle method that detects the oranges in orchard through segmentation. The precision, sensitivity, and F1 score were used to evaluate the suggested technique. Image data from RGB, HSV and RGB/HSV have been used as multi-modal data input. The combined RGB/HSV colour space had the highest F1 score (0.88), precision (0.97), and sensitivity (0.60). Many false positives were found using the HSV colour scheme.

A comparison study conducted by (Häni et al., 2020) revealed counting of images per fruit on different images can be done using different architectures. Their architectures consisted of three types of networks, notably U-Net[Ronneberger et al., 2015), the quicker R-CNN, and the GMM. The U-Net requires less than 100 milliseconds per picture patch for the input image patch of 224x224 pixels that is provided to it. The original input image with a 1920x1080 pixel resolution is processed in less than 4.5 seconds per frame. A quicker R-CNN, on the other hand, takes 120 milliseconds every patch for image 500x500 and 46 milliseconds for every frame with 1920x1080 resolution. A GMM operates at a frame rate of 5 frames per second. These compute times were achieved using the NVIDIA Quadro M1000 graphics processing unit, which was utilised in the proposed method. The video frames were captured at a rate of 30 frames per second and moved at a pace of 2 metres per second.

The dropout (Srivastava et al., 2014) layer based technique is for the regularization during the training phase that focuses on nullifying the additional activation neurons that are arbitrarily selected. This emphasizes the learning parameters to consider a different training approach and not just depend solely on the approach that was considered earlier. The concept of spatial dropout is completely focused on the bigger picture that contains the feature maps apart from the the individual based neurons.

Batch normalizing is a regularization approach (Ioffe and Szegedy, 2015) that normalizes the set of activations in a layer, and it is similar to the previous technique. In order for normalization to work, the batch mean must be subtracted from each activation and then divided by the batch standard deviation. When it comes to the preprocessing of pixel values, this normalization technique, together with standardization, is considered standard practice.

Data Augmentation is technique for creating new images from existing ones by overlaying a random image from the training set. But several specific training tricks limit its real-world use. This is done by teaching the neural network to create fresh samples while minimising the network's error according to (Perez and Wang, 2017) and (Frid-Adar et al., 2018). New research directions for data augmentation have emerged using Generative Adversarial Networks (GANs). With GAN-generated adversarial samples, training deep CNNs can increase generalisation and overcome activation function deficiencies, as presented in (Wang et al., 2018). It takes a long time to train and converge GANs in practise.

When applying sampling techniques for fruit yield estimation, it is important to consider a number of factors according to (Sharma, 2017). including the work's goal, the level of precision required, the amount of heterogeneity present in the population, and the size of the entire population. In theory, fruit yield estimation can be accurate if factors have been accurately defined and determined via the sampling technique chosen. Because of the high demand for fruits around the world, fruit growing is widely practiced in most countries. It is necessary to make more accurate estimations of fruit output and size in vast orchards. Harvesting and counting by hand (by agricultural specialists) has traditionally been the method of choice for estimating fruit production. This method has low precision, high costs, and requires lengthy estimation periods. Recently, artificial intelligence-based intelligent systems, Machine learning and Deep learning approaches are utilized for calculating fruit yield by automatic detecting fruits in the orchard have demonstrated promising results, eliminating the problems associated with traditional yield prediction.

According to the review of different research, it is found that machine learning algorithms perform well in most fruit detection tasks but suffer in large-area yield estimation. Techniques are unable to fit the model due to a lack of generalization capabilities. It's a new hierarchical neural network technology that shows potential in practically every agricultural industry. By calculating intelligent fruit yields with deep learning (DL) technology, this application of Precision farming minimizes human effort and enhances product (i.e., fruit) management. Deep learning algorithms are best suitable in this estimation with some modification in the pre-processing or combining other special techniques for feature extraction.

FCN was the first deep CNN-based semantic segmentation system. No fully connected layers here; merely convolutional layers. The fully connected layers of SOTA networks such as AlexNet (Krizhevsky et al., 2012), VGG16[Simonyan and Zisserman, 2014), and GoogLeNet (LeCun et al., 1998 ) were fine-tuned into a novel semantic segmentation architecture. Many studies employed R-CNN, where computation time is a barrier. Thus, researchers investigated faster R-CNN detection algorithms.

## 3. RESEARCH METHODOLOGY

Due to the high demand for fruits around the world, fruit growing is widely practiced in most countries. It is necessary to make more accurate estimations of fruit output and size in vast orchards. Harvesting and counting by hand (by agricultural specialists) has traditionally been the method of choice for estimating fruit production. This method has low precision, high costs, and requires lengthy estimation periods. Recently, artificial intelligence-based intelligent systems, Machine learning and Deep learning approaches are utilized for calculating fruit yield by automatic detecting fruits in the orchard have demonstrated optimistic results, avoiding the hurdles associated with traditional yield prediction. The proposed model for fruit detection and yield estimation is as follows:-

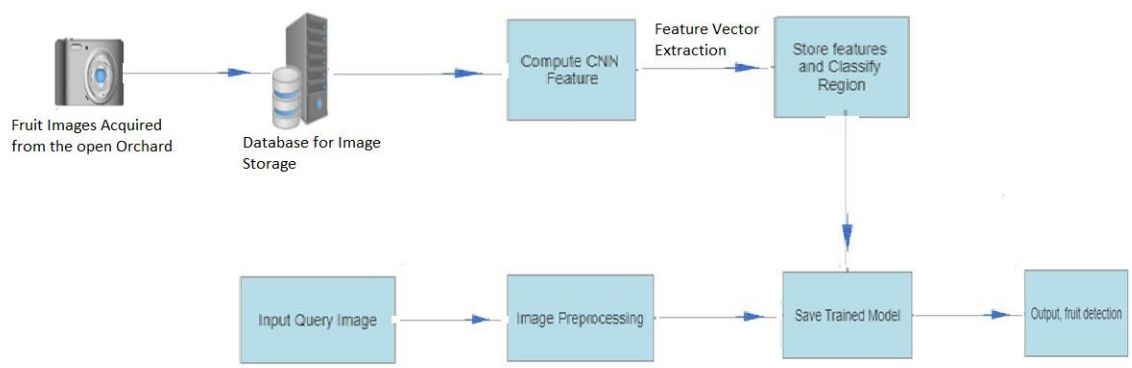


**Figure 1:** Proposed Model

### 3.1 Sampling of Tree for data collection

Initial sampling of trees decides how many representative trees will be picked from the population. This way, the entire orchard's production may be predicted perfectly. The two basic survey theory sampling methods are design-based sampling and model-based sampling. For diverse populations, design-based sampling works well, while model-based sampling works well for geographically defined groupings.

### 3.2 Use of Diverse Sensing Technologies to capture data

The sensory system is the most important system in the identification of fruits. It should be able to capture focused photos in field conditions while dealing with difficult challenges such as changeable lighting conditions and resolution. There are a variety of camera types available in the market today, each with a unique set of characteristics (i.e., black and white, IR, RGB, thermal, etc.).

### 3.3 Data Collection and Labeling

Approximately 80% of the time required on an AI project is spent obtaining, organizing, and labelling data. The process of labelling data is referred to as data annotation in general terms. In many cases, the terms data annotation and data labelling are used interchangeably, though both are being used in a variety of ways depending on the application. Labeled data draws attention to data variables - such as traits, characteristics, or classifications - that can be studied for patterns that can aid in the prediction of the goal variable or outcome.

### 3.4 Data Augmentation

For data-intensive tasks like deep learning, data augmentation can be quite valuable. It is difficult to collect and annotate such a vast dataset for training. Due to the short dataset, the system will perform well in the training phase but will not perform well in the testing phase. To improve the available dataset, different data augmentation techniques will be done, such as translation, rotation, adding noise, cropping, and flipping. It will help the network learn the data's deep properties.

### 3.5 Model Selection

Using neural networks model to classify images is called image classification. Image classification accuracy has grown rapidly in recent years. Convolutional neural networks (CNN) are common deep learning algorithms that can classify images. With directly learnt features, CNN models can offer cutting-edge classifier predictions.

### 3.6 Fruit Detection and Counting

Illumination, occlusion, and overlapping are issues that must be considered while designing an intelligent fruit detecting system in fruit recognition and localization. Using overcast or night imaging reduced shadows and increased detection. Detecting occluded fruit in a picture can increase localization accuracy using advanced DL architectures. Even in large orchards, deep learning can detect fruit well.

### 3.7 Yield Prediction

Incorrect forecasts may result false prediction. Thus, enhanced object detection requires a trustworthy system with human-like cognition. As a self-learning architecture, Deep Learning-based systems are chosen for near optimum prediction. The proposed idea can be implemented with YOLO versions like YOLOV3 and YOLOV4 versions, FRCNN and Mask-RCNN for detection of fruits from open orchard. Further counting and yield estimation can be done using instance segmentation.

## 4. RESULTAND DISCUSSION

The dataset of fruits like oranges and apples here collected from the internet and COCO dataset and by using YOLO pretrained versions we have achieved the result as follows:
For Orange the total 4 versions of yolo like yolov3,yolov3-tiny,yolov4-p5 and yolov4-tiny are used and got the result summarized in the following table1.

| **Table 1**: Result of oranges using YOLO pretrained models | | |
|---|---|---|
| **YOLO Versions** | **Accuracy** | **Time Period** |
| YOLOV3 | 80 to 95% | 0.43 sec |
| YOLOV3-TINY | 50 to 82% | 0.0692 sec |
| YOLOV4-p5 | 50 to 91% | 0.65 sec |
| YOLOV4-TINY | 80 to 95% | 0.06 sec |

Fruit identification in orchards with the development of a model based on pretrained models are tested. Here YOLO different versions like YOLOV3, YOLOV3-TINY, YOLOV4, YOLOV4-TINY, YOLOV4-CSP-SWISH on fruits like apple and orange with threshold value 0.5 and confidence threshold 0.2 shows the accuracies above 90%. The results are as follows:-

**The results are shown as in fig a b ,c, d and e respectively.**

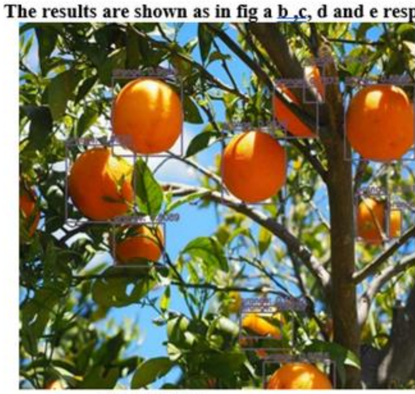

**Fig a.Yolov3**

**Fig b. Yolov3-tiny**

**Fig c Yolov4-csp-swish**

**Fig d. Yolov4-P5**

**Fig e.Yolov4-tiny**

**Figure 2:** Image Results for Oranges on different models

For apples the total 4 versions of yolo like yolov3,yolov3-tiny,yolov4-p5 and yolov4-tiny are used and got the result summarized in the following table2

**Table 2**: Result of Apple using YOLO pretrained models

| YOLO Versions | Accuracy | Time Period |
|---|---|---|
| YOLOV3 | 23% to 92% | 0.73sec |
| YOLOV3-TINY | 20% | 0.07sec |
| YOLOV4-p5 | 50% to 82% | 0.77sec |
| YOLOV4-TINY | 40%to 50% | 0.07sec |

**Yolo v3**

**Yolov3-tiny**

**Yolov4-p5**

**Yolov4-tiny**

**Figure 3:** Image Results for Apples on different models

## 5. CONCLUSION

Implementation of an exact and reliable fruit detection system is a challenging task. The different conditions include orchard environments like change of illumination, appearance variation, and occlusion. The objective of the proposed system mentioned in the paper will find out the solution for the different conditions. The results of pretrained model using COCO dataset for apple and orange gives the accuracy but it can be increased using the customized models.

## REFERENCES


Agarap, A. F. (2018). Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375.

Das, and Kumar V. (2018). Robust fruit counting: Combining deep learning, tracking, and structure from motion. In 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS), pages 1045–1052. IEEE.

Feng, J., Zeng, L. and He L. (2019). Apple fruit recognition algorithm based on multi-spectral dynamic image analysis. *Sensors (Basel, Switzerland)*, 19.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H. (2018). Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321, 321–331.

Frilay J., Aruni Weragoda, and Dale Ashton. Australian vegetable-growing farms: an economic survey 2017-18 and 2018-19. 2019.

Ganesh, P., Volle, K., Burks,T. and Mehta, S.(2019). Deep orange: Mask r-cnn based orange detection and segmentation. *IFAC-Papers On Line*.

Gené-Mola, J., Gregorio, E., Guevara, J., Alfredo A.F., Sanz-Cortiella, R., Escolà, A., Llorens, J., Morros J.R., Ruiz-Hidalgo J., Vilaplana, V. and Rosell-Polo, J. (2019). Fruit detection in an apple orchard using a mobile terrestrial laser scanner. *Biosystems Engineering*.

Gongal, A., Silwal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2015). Apple crop-load estimation with over-the-row machine vision system. *Comput. Electron. Agric*., 120(C):26–35, jan 2016. ISSN 0168-1699. doi:10.1016/j.compag.2015.10.022. URL https://doi.org/10.1016/j.compag..10.022.

Häni, N., Roy, P. and Isler, V.(2020). A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics*. 37(2), 263–282.

Häni, N., Roy, P. and Isler, V. (2020). A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics*. 37(2), 263–282. doi:https://doi.org/10.1002/rob.21902. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21902.

Kang, H. and Chen, C. (2020). Fast implementation of real-time fruit detection in apple orchards using deep learning. Comput. Electron. Agric., 168.

Ioffe S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning, pages 448–456. PMLR.

Kestur, R., Meduri, A. and Narasipura, O. (2019). Mangonet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. *Eng. Appl. Artif. Intell.*, 77:59–69.

Krizhevsky, A., Sutskever I., & Geoffrey E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Kondo, N., Monta, M. and Noguchi, N. (2011). Agricultural robots: mechanisms and practice. *Apollo Books*.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.

Lin, G., Tang, Y., Zou X., Cheng J., & Xiong J. (2020). Fruit detection in natural environment using partial shape matching and probabilistic hough transform. Precision Agriculture, 21(1), 160–177.

Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R.J., Fredes, C. and Valenzuela A. (2020). A review of convolutional neural network applied to fruit image processing. *Applied Sciences*.

Öztürk B., Kirci M., and Günes¸E. (2016). Detection of green and orange color fruits in outdoor conditions for robotic applications. *In 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–5, 2016. doi:10.1109/Agro-Geoinformatics..7577641.

Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv*:1712.04621.

Roy, P., Kislay, A., Plonski, P., Luby, J. and Isler, V. (2019). Vision-based preharvest yield mapping for apple orchards. Computers and Electronics in Agriculture, 164:104897.

Roy, P., Dong, W. and Isler, V. (2018). Registering reconstructions of the two sides of fruit tree rows. *In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1–9. doi:10.1109/IROS.2018.8594167.

Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. *Springer*.

Sa, I., Ge, Z., Dayoub, F., Upcroft, G., Perez, T. and McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. *Sensors*, 16(8). ISSN 1424-8220. doi:10.3390/s16081222. URL https://www.mdpi.com/1424-8220/16/8/1222.

Sarig, Y. (2005). Mechanized fruit harvesting - site specific solutions.

Sharma, G. (2017). Pros and cons of different sampling techniques. International journal of applied research, 3, 749–752,.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint ar*Xiv:1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56),1929–1958. URL http://jmlr.org/papers/v15/srivastava14a.html.

Tan, K., Lee, W., Gan, H. and Wang S. (2018). Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes. *Biosystems Engineering*.

Thendral, R., Suhasini, A. and Senthil, N. (2014). A comparative analysis of edge and color based segmentation for orange fruit recognition. *In 2014 International Conference on Communication and Signal Processing*, 463–466.

Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., Wan, H. and Xue, Y. (2020). Passion fruit detection and counting based on multiple scale faster r-cnn using rgb-d images. *Precision Agriculture*, 1–20. Springer-Verlag. ISBN 978-3-030-59829-7. doi:10.1007/978-3-030-59830-3_19. URL https://doi.org/10. 1007/978-3-030-59830-3_19.

Wan, S. and Goudos, S.K. (2020). Faster r-cnn for multi-class fruit detection using a robotic vision system. *Comput. Networks*, 168.

Wang C., Zhou J,, Xu C. & Xiao B. (2020). A deep object detection method for pineapple fruit and flower recognition in cluttered background. In Pattern Recognition and Artificial Intelligence: *International Conference, ICPRAI 2020, Zhongshan*, China, October 19–23, 2020, Proceedings, page 218–227, Berlin, Heidelberg.

Wang G., Li W., Ourselin S. & Vercauteren T. (2018). Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation. In International MICCAI Brainlesion Workshop, pages 61–72. Springer, 2018.

Wouter B., Eldert J., Hemming J. & Edan Y. (2014). Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. Journal of Field Robotics, 31(6):888–911.

Yu, Y., Zhang, K., Li Yang, and Dongxing Zhang (2019). Fruit detection for strawberry harvesting robot in nonstructural environment based on mask-rcnn. Comput. Electron. Agric., 163.