Survey of Object Detection Methods in Fruit Detection and Counting

Amaan Sharif Nirban
Computer Engineering
M.H. Saboo Siddik College of Engineering
Mumbai, India
amaan.3119032.co@mhssce.ac.in

Maryam Ansari
Computer Engineering
M.H. Saboo Siddik College of Engineering
Mumbai, India
maryam.3119006.co@mhssce.ac.in

Abdullah Ansari
Computer Engineering
M.H. Saboo Siddik College of Engineering
Mumbai, India
abdullah.3119003.co@mhssce.ac.in

Zafar Shaikh
Computer Engineering
M.H. Saboo Siddik College of Engineering
Mumbai, India
zafar.3119040.co@mhssce.ac.in

Saiqa Khan
Computer Engineering
M.H. Saboo Siddik College of Engineering
Mumbai, India
saiqa.khan@mhssce.ac.in

***Abstract*— Computer vision is a field of artificial intelligence that deals with extracting useful information from images, videos and performing tasks using that information. Computer vision is used extensively in almost every field. With computer vision techniques, one can easily automate tedious tasks that would take forever when done by humans. Object detection is a computer vision technique that is used to locate and identify objects in an image or a video. This object detection and localization can further be used to count specific objects present in the scene. An application of this technique is counting on-tree fruits from images or videos of the trees in the field. Many researchers have proposed their systems to detect and count the number of fruits for yield estimation. This paper is a review of some studies performed on fruit detection and counting. We address different approaches that use Support Vector Machine (SVM), Convolutional Neural Network (CNN), Faster Region Based CNN (Faster-RCNN), Mask RCNN, GrabCut Model, You Only Look Once (YOLO), Simple Linear Iterative Clustering (SLIC) and Single-Shot Detector (SSD) methods along with their results and drawbacks. In addition to it we mention the research gaps while summarizing these studies. The studies have an average precision of 87.88%. Further, we propose a system using superpixels image segmentation and CNN to fill the research gaps and overcome the drawbacks of the previous works mentioned.**

Keywords—Computer Vision, Artificial Intelligence, Object detection.

I. INTRODUCTION

Computer vision can be termed the eye of computers. It enables them to understand the world through image and video inputs. Computer vision consists of methods for acquiring, processing, analyzing, and comprehending images. This field is attracting a lot of attention because it is viewed as one of the most useful instruments for reducing human workload. In the agriculture sector, computer vision has been extensively studied in various aspects of precision agriculture, including autonomous harvesting robots, crop yield estimation, plant phenotyping, animal welfare assessment, plant pest, and disease detection, and so on. Fruits and vegetables are detected and their three-dimensional positions are located using computer vision systems. But, due to varying illumination conditions and severe occlusions, and the lack of publicly available image and video datasets, this field remains challenging. Fruit recognition, despite receiving less attention, has gained traction in recent years as a result of its applications in the agricultural and culinary industries.

Fruit recognition seeks to identify fruits based on their type, maturity, or both by analyzing and processing photographs of fruits. These techniques are simple for a human to perform unless he has no prior knowledge of the fruit. Computers, on the other hand, have failed miserably at these tasks. The process of recognizing fruits can be broken down into three key steps: 1) Image acquisition, which is accomplished by using image capture devices to acquire images of the fruit samples. 2) Fruit picture samples are restored, smoothed, or enhanced during pre-processing.
According to some sources, pre-processing also include transforming raw photos to a predefined state (i.e. Grayscale or different color spaces). 3) Image Analysis examines the output of the preprocessing step in order to identify the fruit. Increasing agricultural demands and product development have created new chances for fruit recognition to be used to obtain better products at reduced costs. The current fruit recognition technology is limited to a single type of fruit in a single location. It is not yet robust enough to be utilized to recognize fruit in extreme conditions, environments, or in situations where multiple varieties of fruits must be distinguished.

Once the fruits are recognized and located from the visual inputs, count if on-tree fruits can be obtained which helps in yield estimation. Yield estimation is crucial when it comes to improving field management and getting the outcome for the season's harvest. Also, farmers can plan their next plantation strategy based on the previous yields. Currently, yield estimation is done mainly by manual counting which is time-consuming. Computer vision may help in improving the efficiency of yield estimation.

The rest of the paper is structured as follows: Section II is the summary of different studies performed on fruit detection and counting. In Section III we discuss the research gaps and propose a system to bridge the research gaps. Finally, the conclusion is given in Section IV.

II. REVIEW OF FRUIT DETECTION AND COUNTING

Lanhui Fu et al. performed banana detection based on color and texture features in the natural environment in the year 2019 [1]. They performed detection of banana fruit from the images of banana clusters by using a red-green-blue camera. The background from the images was first removed in HSV color space by analyzing the relationship between the S color component and V color component which saved their detection time and improved detection efficiency. The banana area was then found adapting a support vector machine. They used the local binary pattern features and histogram of oriented gradient features of the banana for SVM classification. The dataset used was a custom developed of 4400 image samples. The results show that the system developed can be applied to detect bananas under different illumination and occlusion conditions. However, counting of the fruit has not been addressed in the study. Manya Afonso et al. proposed a system to detect and count tomatoes using the MaskRCNN algorithm [2]. The system managed to get an average precision of 87.5%. However, the dataset developed contained only 123 samples.

J.P. Vasconez et al. devised a system using two approaches, one being Faster RCNN along with Inception V2 and the other one with SSD and Mobilenet [3]. The system was prepared to detect and count avocados, apples, and lemons. The models were trained, tested, and validated over a custom-developed dataset consisting of 2858 samples. Nicolai Hani et al., adapted CNN to count the number of apples by training the model on 64000 image samples collected from multiple orchids [4].

Fangfang Gao et al. used apples as the sample fruit for testing [5]. They used 12,800 custom-developed image samples. They used the Faster R-CNN-based VGG16 model to detect the fruits hanging on the tree [11]. Their precision for non-occluded, branch occluded, leaf occluded, fruit occluded fruits were 90.9%, 85.8%, 89.9%, 84.8% respectively. The detection speed was nearly 0.241 (s/image). The VGG network used is extremely slow to train and the size of the weights is quite large. Sashuang Sun et al. proposed a method for detecting green apples consisting of fruit region extraction, segmentation, and recognition [6]. For the first section, a modified GrabCut algorithm was developed for the preliminary extraction of fruit target regions in the natural environment. The Ncut segmentation was the second part, and it was designed to handle the challenge of overlapping fruits in targets. In the final step, the three random point reconstruction method was used to generate circle fitting for each apple, based on the recovered contour information. But the methodology was tested on only 200 images of the sample, which was insufficient to justify the occluded fruit scenario.

Addie Ira Borja Parico et al., selected sample fruit for testing as pear [7]. Dataset was custom-developed with 1337 sample images. They used three approaches YOLOv4-tiny, YOLOv4, and YOLOv4-CSP for fruit detection and counting [12]. Their models YOLOv4-tiny, YOLOv4, and YOLOv4-CSP reached average precision of 93.93%, 95.72%, and 97.74% respectively. Maryam Rahnemoonfar et al., developed a simulation-based learning method, using deep learning architecture for counting fruits based on CNN and a modified version of Inception-ResNet [8]. The network consists of several convolution and pooling layers in addition to modified Inception-ResNet which helps to capture features in multiple scales. It is trained on simulated data of 24000 sample images which were generated with some degree of overlap along with variation in size, scales, and illumination. But when tested, it proved effective for synthetic samples with an accuracy of 93.01%, while not so effective for random real images which resulted in average accuracy above 70%.

Xiaoyang Liu et al. used SLIC and an SVM classifier for the detection of apple fruits based on color and shape features [9]. They gathered 1844 image samples from two different orchards and took 444 samples from the cifar-100 dataset. Their precision for different illumination i.e. front light, backlight, sidelight, and artificial light were 95.87%, 90.70%, 94.52%, and 100.00% respectively. And the fruit detection time was noted to be 1.92s. Juan Feng et al. used a thermal camera to acquire an image for apple fruit recognition [10]. They used SVM Classifier on custom-developed 846 image samples. their system acquired an average processing time of 740.42s, and 91.62% of fruit recognition accuracy.

Table I summarises the fruit detected, the dataset used, models used, results, and research gaps along with the drawbacks.

TABLE I. SUMMARY OF STUDIES PERFORMED

Work cited	**Study**	**Fruit Detected**	**Dataset**	**Model**	**Results**	**Drawbacks and Research gaps**
[1]	Lanhui Fu et al.	Banana	Custom Developed (4400 Samples)	Support Vector Machine, Otsu's threshold	Single-scale detection: Average accuracy = 89.63% Average detection time = 1.325s Multi-scale detection: Average accuracy = 92.55% Average detection time = 10.31s	Multi-scale detection is time consuming. The count of fruits present in the cluster after detection is not addressed.
[2]	Manya Afonso et al.	Tomato	Custom Developed (123 Samples)	MaskRCNN with R50, R101 and X101	Precision = 81.57% Recall = 82.09%	Less amount of image sample were collected.
[3]	J.P. Vasconez et al.	Avocado, Apple, Lemon	Custom Developed (2858 Samples)	Faster RCNN + Inception V2, SSD + Mobilenet	Faster RCNN + Inception V2:	Not all fruits were detected from the

					Average Precision = 72% SSD + Mobilenet: Average Precision = 53%	image by the algorithms. The occlusion of fruits by leaves is not addressed.
[4]	Nicolai Hani et al.	Apple	Custom Developed (64000 Samples)	CNN	Average accuracy (yield estimation) = 96.85% Average accuracy (patch counting) = 88.48% w/o pre-or-post processing accuracy = 80% and 94%	Occlusion of fruits by leaves or fruits and fruits on the ground are often ignored. Incorrect selection of the regions during detection. Most of the features are treated as image noise.
[5]	Fangfang Gao et al.	Apple	Custom Developed (12,800 Samples)	Faster R-CNN on VGG16 network	Non-occluded = 90.9% Branch/wire occluded = 85.8% Leaf-occluded = 89.9% Fruit-occluded = 84.8% Detection speed = 0.241(s/image)	The VGG network used is extremely slow to train and the size of weights is quite large.
[6]	Sashuang Sun et al.	Green Apple	Custom Developed (200 Samples)	GBVS-based GrabCut model + Ncut Segmentation Algorithm	Precision = 93.92% Recall Rate = 90.84%	Sparse dataset, due to which insignificant results are obtained for occluded clusters.
[7]	Addie Ira Borja Parico et al.	Pear	Custom Developed (1337 Samples)	YOLOv4, YOLOv4-tiny, YOLOv4-CSP	Average Precision YOLOv4-tiny = 93.93% YOLOv4 = 95.72% YOLOv4-CSP = 97.74%	High computational cost.
[8]	Maryam Rahnemoonfar et al.	Tomatoes	Custom Developed (24,000 synthetic sample + 100 real random samples)	CNN (modified version of Inception-ResNet)	Average accuracy (above 70%) for real sample, Average accuracy (93.01%) for synthetic sample	Partially ripped tomatoes are not recognized. Direct illumination and color saturation in synthetic samples leads to misleading results in real samples.
[9]	Xiaoyang Liu et al.	Apple	Custom Developed (1844 samples), Cifar-100 dataset (444 samples)	SLIC and SVM Classifier	Precision = 95.12% Detection time = 1.92s	Pixel-wise segmentation would give out more precise results than the used detection box method.
[10]	Juan Feng et al.	Apple	Custom Developed (846 Samples)	SVM Classifier	Average processing time = 740.04s, Fruit recognition accuracy = 91.62%	The large average relative error in recognition of fruits from incomplete fruit regions.

III. RESEARCH GAPS AND PROPOSED SYSTEM

A. *Research Gaps*

The system discussed in the earlier section proved to be fruitful in their respective requirements. However, there are research gaps that we discuss in this section.

The lack of a publicly available dataset for the detection of fruits from images lead the researchers to develop their custom datasets. Some of the studies discussed above could manage to collect only a hundred image samples which limit the efficiency and reliability of the system created. Most of the papers focus on the detection of fruit clusters or localization of fruit. Only a few researchers address the fruit counting problem which is crucial in yield estimation.

Despite attaining high average precision, not all fruits were detected. Occlusion of fruits by leaves or adjacent fruits and fruits on the ground are often ignored. The regions that do not contain any fruit clusters are identified as fruit regions due to the same color features. Whereas, many of the features from the images are treated as background image noise. Furthermore, incomplete fruit regions and scenarios where fruits were partially ripped were not recognized. While the systems that detected incomplete fruit regions had a large relative error in detection.

The illumination factor has a huge impact while detecting fruits. Fruits in images with direct illumination were not detected. The datasets that were developed do not cover images from every angle. Besides, illumination conditions were controlled while the acquisition of images. Not all illumination conditions were considered which leads to inefficiency of the systems while detecting fruits from images that are brighter, darker, or have different illuminations all over.

Along with these research gaps, the models and architecture used such as VGG and YOLO, have large weights and have high computational costs. A solution that requires comparatively less computation power would be much more appreciated as it can be then implemented on an end-user basis without much hassle.

B. *Methodology*

Images will be collected from different sources and will then be pre-processed using superpixels segmentation. Superpixels are the product of perceptual pixel grouping, or, to put it another way, the effect of picture oversegmentation. Superpixels contain more information than pixels and match with picture borders better than rectangular image patches. The datasets will be divided into training and validation sets. The proposed solution will be tested on the validation sets and on real-world input images.

C. *Proposed System*

Before the CNN application, we are proposing to implement a segmentation of the input image using superpixels because superpixels are generally used to divide the input image based upon textures. They are larger than pixels and try to combine different parts of the image-based upon texture. This will help us in removing background information from the image and we will be getting only foreground information. When this region of interest would be given as input to the CNN, it would be focusing more only on the foreground and there would not be any disturbance of the background which would affect the performance of the CNN. This CNN method would be more accurate compared to the past methods where researchers have used either CNN or image processing. But we intend to combine image segmentation using superpixels and CNN. Superpixel-based segmentation is more accurate compared to k-means, OTSU, or the various other state-of-the-art techniques. Fig 1. shows a block diagram of the proposed system.

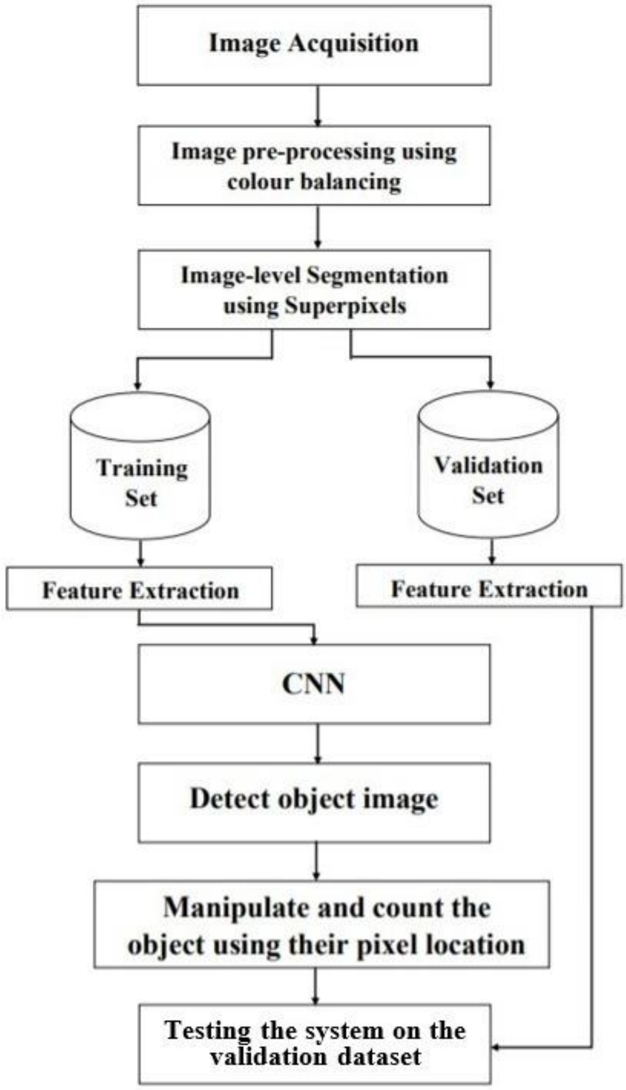

Fig. 1: Block diagram for the proposed system

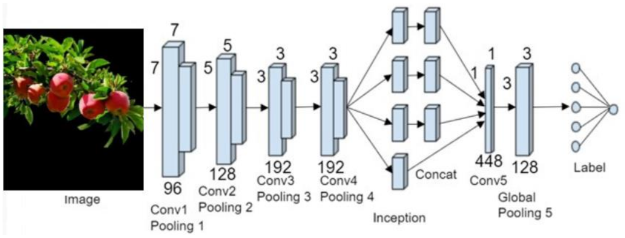

Fig 2. Proposed system using CNN

The input images will be classified according to the fruits detected i.e. a separate class for every unique fruit will be created and hence the output layer of CNN will contain names of the fruit detected as output classes. The CNN will be given an RGB image as an input processed by superpixels in the form of a three-dimensional matrix reshaped into a single column. Suppose the dimension of the input image is $28 \times 28 = 784$, it will be reshaped into 784×1. Output from CNN will contain the label of the class detected. Fig 2. shows the proposed system using CNN. Fig 3a. and Fig 3b. show the output from the CNN detecting the fruits.

Fig. 3a: Output from the CNN

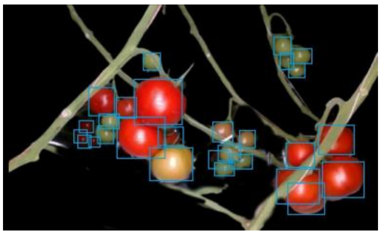

Fig. 3b: Output from the CNN

IV. CONCLUSION

We discussed work done in fruit detection and counting. All the above-mentioned studies were successful in the detection of respective fruits in certain conditions. Most of the studies discussed covered fruit detection under various illumination and occlusion conditions except for a few. Some studies detected and counted fruits with high precision and covered all the challenging aspects while some studies only performed detection and some of them did not address occlusion conditions along with failing to detect or count all of the fruits present in the input image or video. We mentioned the drawbacks of these studies and the research gaps as well. Thus, researchers in this field can benefit from this study and try to fill the research gaps with their proposed systems in the future. We suggest the researchers to consider video-feed as input and carry out real-time detection including on-ground fruits and also detection from varying distances. Finally, we proposed a system using superpixels-based image segmentation combined with CNN to fill the research gaps mentioned in the near future.

REFERENCES

[1] L. Fu et al., "Banana detection based on color and texture features in the natural environment", *Computers and Electronics in Agriculture*, vol. 167, p. 105057, 2019. Available: 10.1016/j.compag.2019.105057.

[2] M. Afonso et al., "Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning", *Frontiers in Plant Science*, vol. 11, 2020. Available: 10.3389/fpls.2020.571299.

[3] J. Vasconez, J. Delpiano, S. Vougioukas and F. Auat Cheein, "Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation", *Computers and Electronics in Agriculture*, vol. 173, p. 105348, 2020. Available: 10.1016/j.compag.2020.105348.

[4] N. Häni, P. Roy and V. Isler, "Apple Counting using Convolutional Neural Networks," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 2559-2565, doi: 10.1109/IROS.2018.8594304.

[5] F. Gao et al., "Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN", *Computers and Electronics in Agriculture*, vol. 176, p. 105634, 2020. Available: 10.1016/j.compag.2020.105634.

[6] S. Sun, M. Jiang, D. He, Y. Long and H. Song, "Recognition of green apples in an orchard environment by combining the GrabCut model and Ncut algorithm", *Biosystems Engineering*, vol. 187, pp. 201-213, 2019. Available: 10.1016/j.biosystemseng.2019.09.006.

[7] A. Parico and T. Ahamed, "Real Time Pear Fruit Detection and Counting Using YOLOv4 Models and Deep SORT", *Sensors*, vol. 21, no. 14, p. 4803, 2021. Available: 10.3390/s21144803.

[8] M. Rahnemoonfar and C. Sheppard, "Deep Count: Fruit Counting Based on Deep Simulated Learning", *Sensors*, vol. 17, no. 4, p. 905, 2017. Available: 10.3390/s17040905.

[9] X. Liu, D. Zhao, W. Jia, W. Ji and Y. Sun, "A Detection Method for Apple Fruits Based on Color and Shape Features", *IEEE Access*, vol. 7, pp. 67923-67933, 2019. Available: 10.1109/access.2019.2918313.

[10] J. Feng, L. Zeng and L. He, "Apple Fruit Recognition Algorithm Based on Multi-Spectral Dynamic Image Analysis", *Sensors*, vol. 19, no. 4, p. 949, 2019. Available: 10.3390/s19040949.

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *Export.arxiv.org*, 2021. [Online]. Available: https://export.arxiv.org/abs/1409.1556. [Accessed: 18- Sep- 2021].

[12] J. Redmon, "YOLO: Real-Time Object Detection", *Pjreddie.com*, 2021. [Online]. Available: http://pjreddie.com/yolo/. [Accessed: 18- Sep- 2021].