Experimental Analysis of Faster RCNN and YOLO Network in Detection and Identification of Cotton Crop

Annapoorna B R*[1], Dr. S. Raviraja[2] and Dr. Ramesh Babu D R[3]
[*1]Assistant Professor, Department of Computer Science and Engineering, DSCE, India
br.anu243@gmail.com
[2]Professor of Medical Informatics, Sri Siddhartha Institute of Medical Sciences & Research Centre, Siddhartha Academy of Higher Education, Tumkur, Karnataka, India
drsraviraja@gmail.com
[3]Vice Principal, Professor & Head, Department of Computer Science and Engineering, DSCE, India
bobrammysore@gmail.com
[*]Corresponding author: Annapoorna B R: br.anu243@gmail.com

Abstract— *Agriculture is significant in human life. World population is directly reliant on cultivation. The agriculture industry has several challenges as out-migration of agriculture labors has profound effects on rural economic evolution, food security, nutrition, and poverty, affecting agricultural presentation, rural households and in general the rural economy, causing declination in the amount of agricultural labors and growing price of yield harvesting. Maintaining the labor and scrambling up in crop growing is essential in resolving these difficulties. In current scenario, the mechanization has been evolving for maintaining the labor and large-scale agriculture. There was never been encounter an experimental approach in detection of cotton crop. However, performance is being a greater opportunity to research. Therefore, in this research we propose an architectural frame work in detecting the cotton crop using faster R-CNN and Yolov3. The experiment carried out to identifying and detecting the cotton crop efficiency using faster R-CNN and YOLOv3. A dataset from the cotton crop field in Karnataka, south India was considered. The proposed framework performance is measured using mAP (mean Average Precision) and confidence-score. Experiment resulting 85% and 97.73% confidence-score for faster R-CNN and Yolov3 for 10000 iterations respectively. It is observed that Confidence score is sensitive to the number of iterations.*

Keywords— *Deep learning, Neural network, Faster R-CNN, YOLOv3, Cotton crop recognition.*

I. INTRODUCTION

Agriculture income plays a vital role in Indian Economic system. Contribution of cotton crop agriculture towards GDP of the country is considerably high. Cotton is a cash crop of India and plays a significant role in Industrial and agricultural sector. During harvesting time cotton crop requires excessive labor. In India cotton crop needs to be harvested 2 to 3 times in a year. In several media resources approximated that every minute, 25-30 Indian agriculture labors and young farmers are drifting to towns from rural areas in search of better source of revenue and lifestyle leaving agriculture. Assuming that pace of drifting from rural areas to towns continues, Indian urban residents is likely to reach 600 million by 2030 and no agriculture labors are left in rural. Out migration of agriculture labors has drastically affected the rural economy causing decreasing in a number of workers which in turn increases the cost of crop harvesting.

The development of modern robotic applications in agriculture field have been considerable alternative for migration of the agriculture labor workers, such as plant growth monitoring, crop harvesting, and yield estimation. All agricultural related work carried out manually by appointing labors but it would impose a high risk of back pain and requires more time for completion in work. Therein, for the estimation of the total amount of the yield, agriculturist have to sample the crops field by field. A traditional way of cotton crop detection and localization used to be taking pictures of cotton crop and then manually detecting and counting the total number of cottons in every image separately. However, this method is not an efficient for those who estimate the total yield for large-area farms.

Currently with modern agricultural robots, cotton crop images are composed distantly and appropriately. However, the composed images in the orchard environment might have features such that images may have insufficient or more illumination in natural lightning, Cotton crop images may be very small far away from camera taking pictures, Cotton crop may be occluded being hidden behind the leaf, bark, and so on. These characteristics cause challenges for detection and localization using a computer aided vision techniques.

In the recent days most of the approaches are moving towards building system by using deep CNN, the machine learning based on artificial neural networks have accomplished the highest performance in detection and localization of cotton crop. In identification and localization detectors built on deep neural networks will compute better results. Machine learning based on artificial neural networks contains procedures that projects advanced abstractions in data. Input data is received by an input layer, while succeeding layer gives a modified form of data. CNN based network entails a several sheets in between the input and output layer. CNN based networks are much harder to train. ConvNet which contains several convolutional layers and fully connected layers. ConvNet are persuaded by genetic processes and are deviations of multi-layer perceptron that incorporates the minor amount of data manipulation to enhance the quality. A ConvNet encompasses several sub-sampling and a set of filters, parameters and also intentionally fully connected layers. Convolutional layer receives an input image of the form $m*m*r$ image where the parameters 'm' represents image height and width, and the parameter 'r' represents number of channels. CNN contains several filters each of size $n*n*q$ where 'n' contains a dimension that is reduced in size with respect to original image and 'q' represents the number of channels. Parameter 'r' may consider lesser value or it may differ for each kernel. The size of the filters provides nearby associated structures that are covolued with the image generating 'k' feature-maps of dimension $mn+1$. CNN preceded with three concepts: shared weights, pooling and local receptive fields. Each neuron in the first hidden level connects itself with the smaller portion of the input neurons, called as local receptive field that is a minor box on the input pixels. For complete image slide the local receptive field. There exists different hidden neuron in the first hidden layer for every local receptive field.

The mapping of certain features in the image, maps the data from the initial level to the hidden level. Weights that describe the mapping of features are called shared weights and the weights associated with the feature-map is known as shared bias. Shared weights and bias defines the kernel. Final feature map is produced by the pooling layer which obtains output from individual feature map produced from a set of filters. The process makes the output data more planer. For pooling, Max-pooling concept is used. Researchers recommended a YOLOV3 model for attaining the said objectives. YOLOv3 architecture contains 106 NN-layers which considers characteristics like color which makes the model to study, identify, categorize the object of interest, and also another method called Faster-RCNN with classifier fusion model for detection of cotton crop from images captured. Authors have used multilevel features like shape, color and so on to make the model learn classifiers. Input to the system are images captured from videos and the output image contains cotton crop localized with the bounding boxes. At last the model combines all the possibilities of classifiers to produce a ultimate abjectness for proposal candidates.

II. LITERATURE SURVEY

In the following review section, authors had made an attempt to analyze the existing structure in practice and methodologies used to detect and distinguish the fruits or flowers or other crop having limitations defined.

SuchetBargoti[1] proposes a method anticipated in the identification of various fruits in the orchard. The methodologies used was Faster RCNN. In RCNN method the image data is disseminated through many convolutional layers. RCNN technique employs the VGG16 net which incorporates 13 layers of convolution, and the ZF network comprising 5 layers of convolution. During testing, over all 300 fruits were detected. Using tiled Faster-RCNN method, advanced level performance such as harvest mapping was assessed. Authors considered a total of 726 apples images, 385 almond and 1154 mango images as a dataset. ImageNet characteristics are sufficient for identification and classification in orchard environment. Over all through the Faster RCNN technique authors have achieved a precision and recall score of a value greater than 0.9 with respect to result. The drawback of RCNN technique was that, with varying light intensity the accuracy achieved in localizing the fruits in the images was not good enough.

Bresilla [2] recommends a you only look once (YOLO) which is a deep neural networks technique in which CNN is used for identification of pears and apples on the tree. Authors have used CNN called YOLO for localization and detection. In a single pass the classification, detection and region extraction are all accomplished. In order to overwhelmed the failure in recognition of smaller fruits, the entire image was divided into 26 X 26 and all other layers were excluded to upsurge the speed. For model training, authors have considered a dataset which comprises of apples (100 images), pears (50 images), and combination of both which consists of more than 5000 images were used. The advantage of YOLO methodology is that due to exclusion of some layers there is an improvement in processing speed while retaining the accuracy. The drawback of YOLO method which contains many layers of CNN network has a higher computational power.

InkyuSa [3] recommends a deep CNN technique that identifies a several fruits. NIR and RGB images have been used by CNN method. Multi-model Faster RCNN method has been used by the authors to achieve higher accuracy. RCNN structure has been incorporated to identify several types of fruits or vegetables like apple, orange, strawberry, pepper, rock melon, avocado and mango. The structure incorporates a Deep CNN approach. The Faster RCNN methodology has 2 portions that is a region proposal and region classifier. These two portions incorporates RGB images, which helps in the detection of object of interest. In the region proposal step, the interested object exist in the boundary region. Region classifier will help in deciding whether the particular area is a class of interest or not. In order to fine tune faster RCNN PASCAL VOC dataset was used. The dataset comprises sweet pepper (100 images) and rock melon (109 images). Accuracy of 80% was achieved in identification of the fruit. The drawback of the proposed technique is that; accuracy was not up to the mark.

Maryam Rahnemoonfar[4] proposes a method in which they focused on harvesting. Accountability of fruits and flowers benefits the agriculturalists to upgrade their judgements on the farming practices. Crop estimation benefits agriculturalists in enlightening the yield excellence. Crop quality is accomplished by incorporating novel deep learning NN, it is reformed by comprising an improved form of Inception-ResNet layer. The CNN consist of several filters and sub-sampling layers. For the training purpose authors have used simulated data but for testing original data has been used. Further CNN based model is effective in the identification of occluded fruits with varying degrees of light. The dataset for the training purposes consists of 24000 images and for testing 2400 images were used. For the proposed application CNN based model offers higher performance. The precision of the incorporated methodology is 70-100% and for 100 images the average precision was approximately 91.03%. CNN based prototype is robust to several environmental circumstances. The drawback of proposed technique is that it failed in detecting green fruits, as it was not trained for green fruits.

Philipe A. Dias [5] recommends a new methodology in the identification of apple flower, which is built on a type of ML based on ANN. In contrast with existing approaches, the characteristics pulled out by our conventional NN combines both morphological data and color, which helps in achieving higher performances. The proposed methodology established for the identification of apple flower uses CNN and SVM algorithm. The best prototype extracts feature from CNN, diminishes feature dimensionality to 69, and SVM is used for classification purpose. Testing was accompanied on four different types of datasets. Proposed CNN-based prototype permits accurate detection of flower with ideal with F1score of 80% on testing images.

Jun Deng [6] reviewed a paper on object detection based on deep learning in which author analyzed various method for target detection. Object identification has become an important research topic in the past 20 years and has been widely used. It aims to quickly and accurately identify and locate a large number of objects of predefined categories in a given image. According to the model training method, the algorithms can be divided into two types: single-stage detection algorithm and two-stage detection algorithm. The algorithms that have been incorporated at each stage has been analyzed by author in detail. Then the open and special datasets commonly used in target detection are introduced, and various representative algorithms are analyzed and compared in this field. Finally, the possible challenges for target detection are mined.

Table 1: Analysis of existing techniques

Authors	Type of Crop	Method	Finding
SuchetBargoti	Almond and Mango	RCNN	Failed to provide good accuracy
KushtrimBresilla	Apple and Pears	YOLO	Higher computational power
Inkyu Sa	apple, orange, strawberry, pepper, rock melon, avocado and mango	DCNN	Failed to provide good accuracy
Maryam Rahnemoonfar	Tomato	CNN with Inception ResNet layer	The precision of the incorporated methodology is 70-100% and for 100 images the average precision was approximately 91.03%.
Philipe A. Dias	Apple flower	classification is based on SVM	Proposed CNN-based model permits accurate detection of flower with F1score of 80% on testing images.

III. METHODOLOGY

After reviewing many existing methods none of the researchers had made an attempt to automate the cotton crop. Authors in this paper has done an experimental analysis for automating the cotton crop. Following are the descriptions of the methodologies that have been incorporated.

A. Structure Outline:

Deep learning method along with YOLOv3 has been used for identification and localization of cotton crop. The main feature to carry out the detection problem is the capability to overcome the issue like obstruction. In order to overcome the problem of occlusion authors have used approaches based on Deep Learning. Fig.1 represents the proposed system block diagram. As an input, the system proceeds with the image and permits it to trained layers of a ConvNet. Sooner the image processing, ConvNet identifies the object of interest in the image as cotton crop. The output of prototype results in detecting cotton crop image with bounding boxes and provides tag for each bounding box.

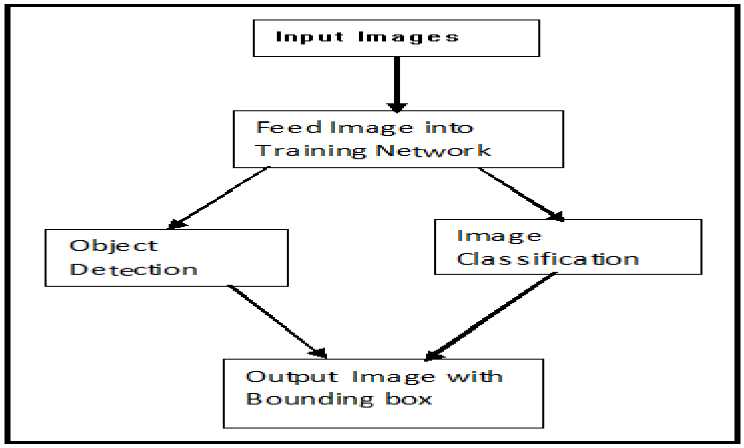

Figure. 1. Block diagram for proposed system

Training phase and testing phase are two phases in this projected system as shown in Fig 2a and Fig 2b respectively.

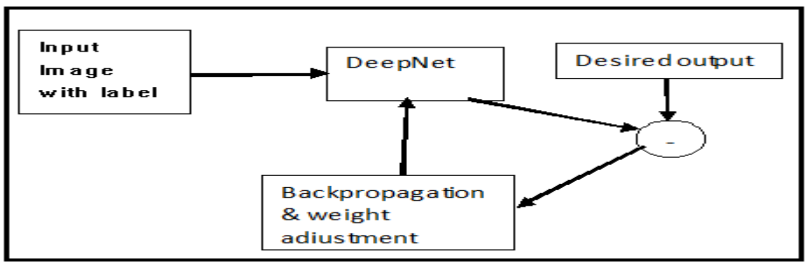

Figure. 2a. Training Phase

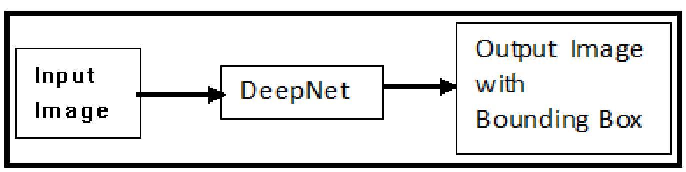

Figure .2b. Testing phase

B. YOLO and Faster RCNN Architecture:

YOLO "You Only Look Once", the algorithm that incorporates ConvNet for object recognition. YOLOv3 procedure makes use of a single NN to the whole image, this system initially splits the input image into various sections and then forecasts the bounding boxes and provides the assurance for individual section. YOLOv3 uses layers of convolutional and makes it a fully convolutional network (FCN). The procedure comprises anchors which is of three in number, which helps in forecasts 3 bounding boxes for each cell. Features are extracted from these 3 bounding boxes which are used in training the system.

Fig. 3 signifies the Yolo architecture. The proposed method contains three phases, RB, DL, and UL.

- RB has stack of layers, in which the output of one layer is added to another layer deeper in the block. It will overcome performance degradation problem associated with deep neural architectures.
- DL predict across 3 different scales. DL aids in recognition at feature-maps of 3 different sizes, having 3 different strides 32, 16, 8 respectively.

- The system down sample the input images until the first detection layer, feature map associated with the layer of stride 32 has been used for recognition purpose. Further, layers are up tested by a factor of 2. Another layer with stride 16 has been used for recognition purpose. Up sampling method is repeated, till it reaches layer of stride 8 where the final recognition is made.

Fig.4 demonstrates the Faster RCNN architecture. The primary components of the system are:

- CNN.
- RPN (Region Proposal Network).

The Faster RCNN makes use of RGB images to perform the basic object recognition. The working of the above architecture mainly includes of two phases: (i) Region-Proposal; and (ii) Region-Classifier.

The goal of the trained FasterRCNN is to detect the cotton crop object in a given input image. Faster RCNN is trained to provide the detected image as its output along with confidence score, class name and the loss graph. The ConvNet model used is the Inception-v2 model, which provides the feature map for the RPN layer. The RPN layer generates the RPN classification and localization loss and generates the proposed regions. The NMS makes sure that each object in the image is detected only once that is each cotton crop in the image has a single bounding box. The IOU ratio is checked for the classification purpose and the desired output is generated.

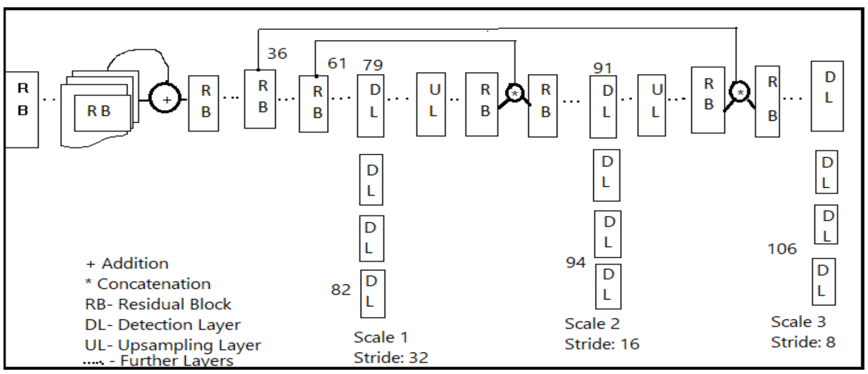

Figure.3. Yolo architecture

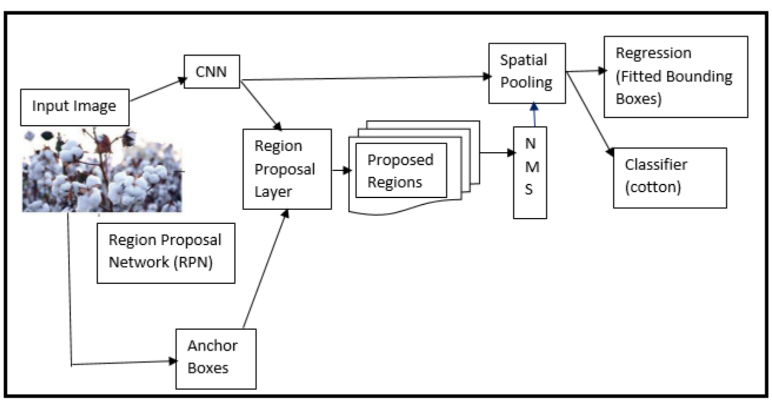

Figure.4. Faster RCNN architecture

Using the above architecture, we have developed a system that uses YOLOv3 and FasterRCNN that has been tested and executed on real time images and finally obtained a results

IV. RESULTS AND DISCUSSION

Based on the proposed method Faster RCNN was found to be most efficient in many research applications. However, in the proposed method we found that YOLOv3 technique having higher accuracy for the low iteration. Step size, number of both training iterations and examples are utilized in one iteration helps in obtaining the results that is prediction percentage and the accuracy of the bounding boxes. In YOLO, the structure got 97.73% mean average precision and loss is 0.2514 for Batch size of 64, for step size 0.0001 and for number of iterations 1000. Loss versus iterations are plotted as shown in the Fig 5 graph.

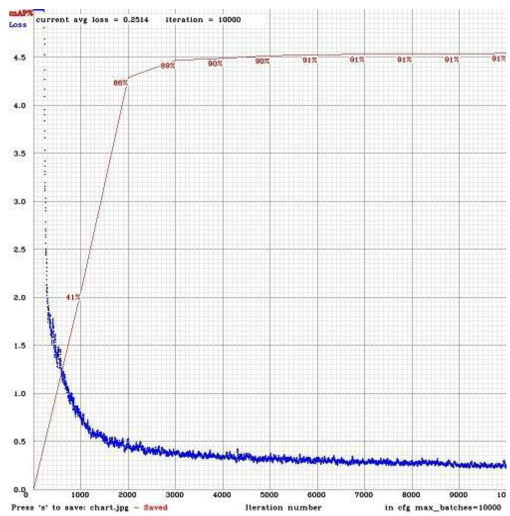

Fig 5. Loss versus iterations graph

Figure 6. Tested Image for YOLO with 10000 iterations

When the single images are tested the resultant images are produced with labeled bounding boxes as shown in the fig 6 and fig 7.

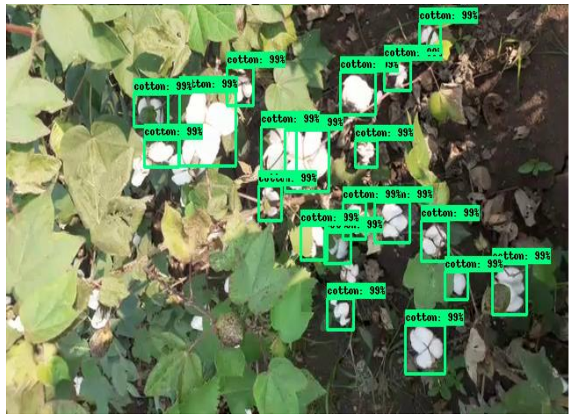

Figure 7. Tested Image for Faster RCNN with 100000 iterations

Fig 8 and 9 depicts the loss value for the classification of detected objects into various classes (cotton crop). The x-axis denotes the number of iterations and the y-axis specifies the values.

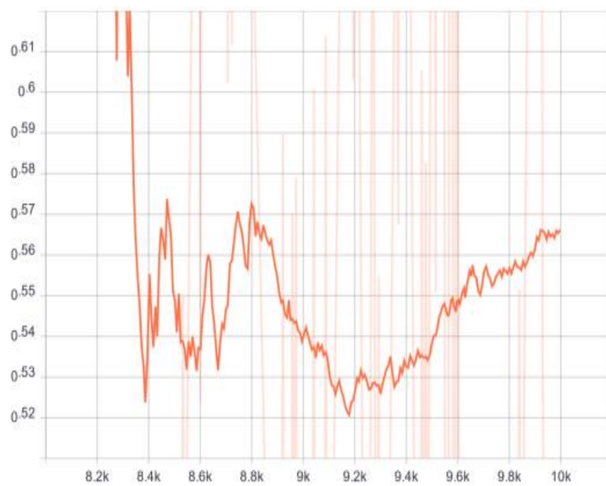

Figure 8. Localization loss for 10000 iterations in YOLO

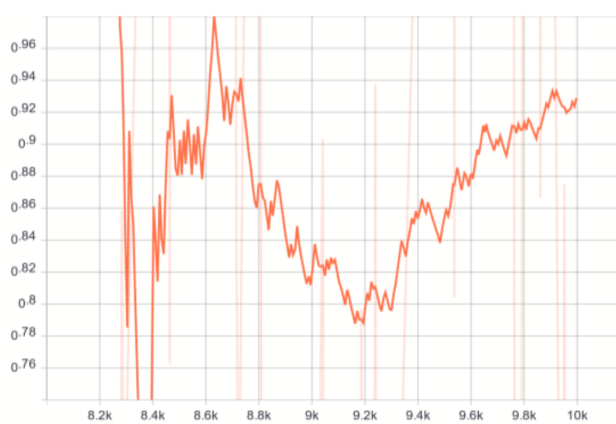

Figure 9. Localization loss for 10000 iterations in Faster RCNN

V. CONCLUSION

In order to detect and localize the cotton crop in the given image the proposed method used is based on Image based learning architecture. The dataset generated for the system consists of images or videos with cotton crop as the main class. The model YOLOv3 network and Faster RCNN was trained for 1600 cotton crop images. The networks were trained with different learning rates and for varying numbers of iterations. Further, the output weights obtained were analyzed with higher prediction percentages and lower error rates were implemented for the final network testing. The Faster-RCNN network takes less time to train but requires large number of iteration to produce better confidence score. The Faster RCNN network trained for ten thousand iterations approximately consumed 1 hour of duration, however produced lower confidencescore between 58-85%. Confidencescore is highly sensitive and directly proportional to the number of iteration. For an example in order to increase the confidencescore, it was trained for 100000 iterations for approximately 7 hours and 30 minutes generating a confidence score of 98-99%. The YOLOv3 network when trained for 10000 iterations consumed a longer time approximately 5 hours, but it produced a confidence score of 97.73% which is considerably a very good score. Therefore, we concluded that the YOLOv3 network takes a long time to train but generates higher accuracy for lower iteration count. Improving the tools and techniques of system further will result in more optimal and hassle-free operations.

REFERENCES

[1] SuchetBargoti and James Underwood., Deep Fruit Detection in Orchards, IEEE International Conference on Robotics and Automation (ICRA) Singapore, 2017.

[2] KushtrimBresilla, Giulio Demetrio Perulli, Alexandra Boini., Single-Shot Convolution Neural Networks for Real-Time Fruit Detection Within the Tree, Frontiers in plant science, Volume 10, Atricle 611, 2019.

[3] Inkyu Sa, Zongyuan Ge, FerasDayoub, Ben Upcroft, Tristan Perez and Chris McCool., Deep Fruits: A Fruit Detection System Using Deep Neural Networks, Sensors Article, mdpi journal, 2016.

[4] Maryam Rahnemoonfar and Clay Sheppard., Deep Count: Fruit Counting Based on Deep Simulated Learning, Sensors Article, mdpi journal, 2017.

[5] Philipe A. Dias, Amy Tabb, Henry Medeiros., Apple flower detection using deep convolutional networks, Computers in Industry Elsevier journal, Volume 99, Pages 17-28, 2017.

[6] Jun Deng, Xiaojing Xuan, Weifeng Wang , Zhao Li , Hanwen Yao , Zhiqiang Wang., A review of research on object detection based on deep learning, Journal of Physics: Conference Series, 2020.

[7] Tao Yongting, Zhou Jun., Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking, Computers and Electronics in Agriculture Elsevier Journal, Pages 388-396, 2017.

[8] Steven W Chen1, Shreyas S. Shivakumar1, Sandeep Dcunha2, Jnaneshwar Das., Counting Apples and Oranges with Deep Learning, Ieee Robotics And Automation 2016.

[9] Nasser Behroozi-Khazaei, Mohammad Reza Maleki., A robust algorithm based on colour features for grape cluster segmentation, Computers and Electronics in Agriculture Elsevier Journal, Volume 142 Part A, Pages 41-49, 2017.

[10] Jun Lu, Won Suk Lee, HaoGan, Xiuwen Hu., Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis, Bio systems Engineering Elsevier Journal, Pages 78-90, 2018.

[11] Zaw Min Khaing, Ye Naung, PhyoHylamHtut., Development of Control System for Fruit Classification Based on Convolutional Neural Network, IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, 2018.

[12] Xiaochun Mai, Hong Zhang and Max Q.-H. Meng., Faster R-CNN with Classifier Fusion for Small Fruit Detection, IEEE Transactions on Automation Science and Engineering, Volume 17, Issue 3, 2018.

[13] CHUNHUA HU, XUAN LIU., Automatic Detection of Single Ripe Tomato on Plant Combining Faster R-CNN and Intuitionistic Fuzzy Set, IEEE Access, 2019.

[14] Ce Yang, Won Suk Lee, Paul Gader., Hyperspectral band selection using kullback-leibler divergence for blueberry fruit detection, IEEE conference, USA, 2017.

[15] Ehud Barnea, RotemMairon, Ohad Ben-Shahar., Colour-agnostic shape-based 3D fruit detection for crop harvesting, Bio systems Engineering Elsevier Journal, 2018.

[16] Motion Xu Liu, Steven W. Chen, Shreyas Aditya., Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure, IEEE International Conference on Intelligent Robots and System, 2018.

Conflicts of interest

We Annapoorna B R, Dr. S. Raviraja, Dr. Ramesh Babu D R declare(s) that there are no conflicts of interest regarding the publication of this article.