

# Arecanut bunch segmentation using deep learning techniques

Anitha A C, R Dhanesha, Shrinivasa Naika C L  
Department of Studies in Computer Science and Engineering,  
University B.D.T College of Engineering,  
Davanagere, Karnataka 577004, India

Krishna A N, Parinith S Kumar, Parikshith P Sharma  
Department of Computer Science and Engineering,  
SJB Institute of Technology, Bengaluru-560060, India

Received: November 6, 2021. Revised: May 14, 2022. Accepted: June 17, 2022. Published: July 26, 2022.

**Abstract-** Agriculture and farming as a backbone of many developing countries provides food safety and security. Arecanut being a major plantation in India, take part an important role in the life of the farmers. Arecanut growth monitoring and harvesting needs skilled labors and it is very risky since the arecanut trees are very thin and tall. A vision-based system for agriculture and farming gains popularity in the recent years. Segmentation is a fundamental task in any vision-based system. A very few attempts been made for the segmentation of arecanut bunch and are based on hand-crafted features with limited performance. The aim of our research is to propose and develop an efficient and accurate technique for the segmentation of arecanut bunches by eliminating unwanted background information. This paper presents two deep-learning approaches: Mask Region-Based Convolutional Neural Network (Mask R-CNN) and U-Net for the segmentation of arecanut bunches from the tree images without any pre-processing. Experiments were done to estimate and evaluate the performances of both the methods and shows that Mask R-CNN performs better compared to U-Net and methods that apply segmentation on other commodities as there were no bench marks for the arecanut.

**Keywords-** Arecanut, Segmentation, Deep Neural Networks, U-Net, Mask R-CNN, Feature Pyramid Network.

## I. INTRODUCTION

Agriculture and farming is a backbone of many developing countries and provides food safety and security for the nation. It is a primary source of the mankind and take part an important role in the nation's economy where as in developed countries it becomes agribusiness. Agriculture is labor intensive, time consuming

and needs continuous monitoring for weed and disease control. A major improvement in agri production has to be achieved with the decreased agricultural land to meet increasing population throughout the world safeguarding the natural ecosystems by following sustainable farming procedure at the same time [1]. A complex and unpredictable multivariate agricultural ecosystem needs to be better understood by continuously monitoring, collecting and analyzing various environmental and physiological aspects. Precision agriculture and farming aims to optimal utilization of water, fertilizers and herbicides to maximize the yield and quality. Precision agriculture as an integrated crop management system plays a major role in soil mapping, plant phenotype detection, crop identification, disease detection, weed control, estimated production, yield monitoring, selective harvesting and grading based on the maturity level noticed. Several land-based and aerial-based applications using machine vision techniques have been developed for natural resource assessments, farming, growth monitoring, harvesting, grading and sorting. A large volume of data collected through different imaging sensors poses a variety of challenges. Hence, intelligent image analysis techniques for growth monitoring, disease and weed identification/classification and yield monitoring and estimation is required for a variety of agricultural applications.

Arecanut being a major plantation in India, it is grown largely in many parts of the country due to its high commercial value. It is also called betel nut and is used mainly for masticator purposes. Arecanut plays a vital role in Ayurveda and Veterinary medicines. Arecanuts takes a vital role in the religious, community and cultural functions and in the lucrative life of the mankind in India. India stands first both in production (57%) of arecanut and area (47%). The average production in India is 1.27 tones/hectare and is on par with the world productivity [2]. Arecanut trees are very tall and thin and grow to the height of 30-40 feet. So, needs skilled labors for growth monitoring, disease detection, spraying, harvesting. Obviously, the labor cost will become

very high and this increased cost burdens the small and middle level farmers life. It is also a very risk-taking for the workers who engaged in harvesting. Many instances were appeared that workers are injured falling from the tree. This creates a huge opportunity for vision based automated harvesting. Image segmentation being a fundamental step in machine vision it could be employed to automate tasks such as disease detection, selective spraying, yield estimation and predict harvest time. Crop growth information can assist farmers arrange compost, herbicides, harvest and better estimate of the yield.

Manual process of crop growth monitoring is labor intensive, expensive and often inaccurate especially for nursery crops, vegetables, fruits, dried fruits, tree nuts and horticulture. Robust and accurate segmentation of crops realizes automated monitoring of growth, health, maturity and yield. Segmentation during several growth phases have accomplish a little success so far as the crop color, shape and/or size varies with cultivar, season and also limited by occlusion, shadow and non-uniform background and illumination. Identifying matured crops is one of the most attracted research and is essential for estimating the yield and automated harvesting. A lot of literature exists for segmentation, classification, harvesting, grading, disease detection for crops like tobacco, potato, coffee, tea, mango etc. A very few attempts have been made for the crop arecanut. The purpose of this research is to present a method which is accurate and efficient for the segmentation of arecanut bunches by eliminating unwanted background information.

The rest of this article is arranged as follows. Section II reveals the work done on the related crop segmentation. Section III describes the methods applied for arecanut bunch segmentation. Data set and models training is described in section IV. Experimentation and performance evaluation is described in section V. Summary of the work done is accorded in the section VI.

## II. RELATED WORK

Segmenting an image being a fundamental and classical problem in machine vision, it divides an image into many non-overlapping subsets in such a manner that each subset corresponds to a meaningful object in the image. Segmentation being a one of the important steps in any machine vision systems, it is used to analyze or interpret an image without human intervention and its performance largely impact on the performance of the whole vision system. Image segmentation is a perceptual grouping/clustering of data points based on similarity, proximity and continuation which reflect local and/or global properties of an image [3]. Developing an efficient and accurate automated segmentation method is essential as the manual segmentation is very difficult, consumes time, subjective and error-prone. General purpose segmentation is very challenging as it is domain dependent and inherently ill-posed. Another challenge is effective representation of object. Most of the existing segmentation techniques gives attention to a binary

classification approach i.e. crops vs. backgrounds. Background removal is an important step and has to be done in an most pertinent manner to keep away from misclassification. A very less work is done for the segmentation of arecanut. Review of different segmentation techniques for other similar commodities is directly related to our problem.

Crop detection and segmentation is challenging, because color of the crop and the sharp of the color varies in the outdoor field as the illumination changes. Moreover, shadows and inter-reflection also increases the complexity as the illumination changes. Segmenting immature crop is much more difficult as the immature crop color is mostly green in color and resembles the background foliage. Besides in a single crop bunch, there will be slight variation in colors of different parts and increases the complexity of crop segmentation. For example, arecanut bunch have nuts of green and yellow colors. In reality, segmentation approaches that uses color too possess their superiority. Primarily, color being the most powerful visual cue it is used to distinguish an object from others. Secondly, color is mostly unchanging to transition in size, direction and obstruction of the object under persistent illumination environments [4]. Color-based approaches are mainly grouped into two classes: pixel-based and region-based approaches. Images captured by digital camera in the electromagnetic spectrum are generally represented using RGB color space. Humans also perceive color in the RGB space. RGB space can be used to create other color spaces using linear or non-linear transformations. Alternative color spaces are used to address illumination variation issues while capturing images. A detailed study on various pixel-based approaches demonstrated ExGR color model gives better results for the green vegetation segmentation in the field compared to other color models for the images that were acquired in various environmental conditions (green house and field conditions) with different backgrounds [5].

Many yield estimation methods involve image segmentation/detection which include mango yield estimation by segmenting image pixels into an object and the surrounding using RGB and YCbCr color space and texture-based segmentation using variability of pixel adjacency [6], segmentation for apple yield estimation using HSV color space for red apple detection and HSI profile to detect green apple pixels in a controlled lighting at night [7], threshold, linear color model and histogram, Mahalanobis distance and Bayesian classifier are pixel-based segmentation methods applied to detect and estimate the yield of reddish grapes using RGB and HSV color spaces. Better results were obtained with threshold based segmentation using Otsu threshold applied to H layer [8]. Pixel-based approaches are simple and efficient, but inculcate more noise. Researchers turn their attention to region-based approaches which are mostly based on edge detection and shape fitting or clustering including apple image segmentation [9], segmentation of coconut bunch [10], segmentation of oil palm bunch [11].

Apples are identified using thresholding followed by edge detection, circular Hough transform and clustering [12] and K-nearest neighbors classification using color and texture features, seed area detection and growth followed by contour segmentation [13]. Citrus are detected by converting RGB to HSV followed by thresholding, color detection and watershed segmentation [14]. A very few attempts have been made for the segmentation of betel palm. Siddesha et al., [15] explored thresholding, clustering, Watershed and an interactive Maximum Similarity based Region Merging (MSRM). Better results were reported with MSRM compared with other techniques, but it is time consuming as it requires human intervention to mark foreground and background regions. Segmentation of arecanut bunches using active contours can be found in [16].

Segmentation is notably superior when we combine both color and texture features in contrast to conditions where only one feature is available [17]. Mango regions segmentation from its tree image is exploited using color and texture features of the objects [18]. Variable lighting conditions in the field influences the magnitude of reflected light and therefore reduces the effectiveness of threshold and color based approaches. In many applications, objects almost often have particular shape. Segmentation of objects based on the analysis of shape and size is often desired. Grape bunch detection using HOG, LBP as shape and texture descriptor respectively and SVM-RBF classifier followed by DBSCAN clustering is an attempt in this direction [19]. Segmentation of tomatoes based on shape and position information using active contours and comparing SIFT descriptors between two consecutive images computed at points of the contour proposed by [20]. Segmentation of overlapping objects remains challenging even using prior shape property as it hides the partial contours. Occlusion and overlapping affect the shape based segmentation methods. Another challenge is the influence of shadows on segmentation performance. Shadows prone to be classified as foreground during segmentation. Shadow identification and elimination using color and multilevel threshold improved the accuracy of segmentation of images captured in the field conditions [21].

Machine Learning (ML) based segmentation approaches gained more popularity. A method to segment matured grape bunches by finding edges using gradients of gray-scale images and then determining circles using Hough transform has been proposed [22]. Three grape cluster detection algorithms based on distribution of edges between the foliage and the grape clusters, training C5.0 decision tree algorithm using patches of foliage and grape areas and on shape matching were presented in [23]. Most segmentation techniques concentrates on the use of particular properties such as color, texture and shape of the target object to be segmented [24]. Segmentation based on these hand engineered features are not powerful. Researchers turned their attention to deep learning based methods such as Convolutional Neural Networks (CNN). Pixel-wise classification-based segmen-

tation of apple fruit in orchard using multi-scale Multi-Layered Perceptrons (MLP) and CNN followed by post-processing using Watershed segmentation and circular Hough transform to identify and estimate the yield of the fruits has been proposed [25]. Though better results were achieved using CNN and Watershed compared to traditional methods, they are susceptible to illumination and occlusion. In addition, pixel based methods labels one pixel at a time results in enhanced run time. A multi-class (fruit, leaves and branches) image segmentation to segment almond fruit using feature learning with a Conditional Random Field (CRF) has been proposed [24]. Fruit flower detection results on apple, peach and pear flowers without any pre-processing or data-specific training under different conditions using CNN demonstrated the robustness and broad applicability [26]. Detection and counting of mangoes using MagnoNet, a deep CNN based architecture and contour based connected object detection demonstrated that it is better compared to its peer methods and it is invariant to illumination variation, scaling, contrast and occlusion [27].

### III. METHODOLOGIES

This article presents the capability of two deep learning based techniques for the segmentation of arecanut bunches from images taken in the field environment. To assess the methodologies, the data set made available by R. Dhanesha et al. [28] has been used to train and test the performance of the networks. The use of deep neural networks improves the accuracy of segmentation, because features are derived automatically from the images to resulting in the best representation of the inputs. Results were compared with state-of-the-art techniques in terms of well-established metrics. The following sections will describe more details about the architectures of U-Net and Mask R-CNN used for segmentation of arecanut bunches.

#### A. U-Net

U-Net is basically a CNN developed for biomedical image segmentation [29]. The success of CNN was limited by the amount of training data and the network size [30]. The principal idea of CNN is to learn mapping an image onto feature space by converting it into a vector which is further used for classification. CNN is basically developed for classification function, where the output of CNN is a single class label for each image. The desired output in image segmentation is a localization of object of interest to be segmented and the assignment of class label to each pixel. The idea of developing U-Net is centered around this problem. U-Net was built on top of a more opulent architecture called Fully Convolutional Network (FCN) and has been extended to work with fewer training examples to yield improved segmentation performance. It uses the same process of feature mapping of an image into a vector to recreate an image from this feature vector. This preserves the structure of an image which would reduce distortion enormously. This is a challenging task because it is very difficult to convert a vector into an image.

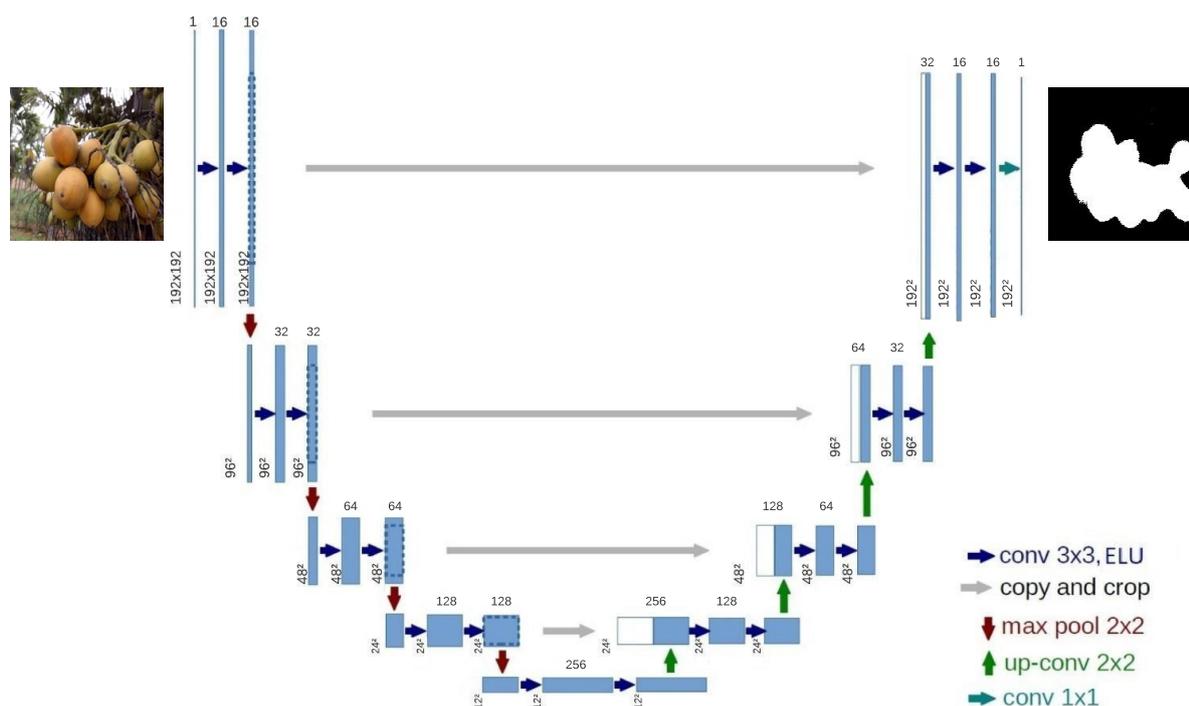


Fig. 1: architecture of U-Net: Blue boxes indicates multi-channel feature maps and white box correspond to copied feature maps. Arrows indicates different operations

U-Net architecture shown in Figure 1 can be conceptually thought of encoder network accompanied by a decoder network. The Encoder-contraction layer, the first half of the architecture is an experienced classification network such as VGG/ Residual Network (ResNet) in which convolution blocks followed by a max-pool down sampling is done to map the input image onto feature vectors at multiple different levels. Each block in the contraction layer applies two 3x3 convolution operation followed by a 2x2 max pooling to an input. The number of feature maps doubles by going through every block in view of the architecture learns the compound structures successfully. The last layer establishes the link between contraction and the expansion layers. It is done using two 3x3 CNN layers and a 2x2 up convolution layer. The Decoder-expansion layer, second half of the architecture essentially carries out the reverse of the down sampling. It is the symmetric extension path to semantically extrapolate the discriminative features learned from the encoder onto the pixel space to acquire precise localization employing transposed convolutions. Therefore it is an end-to-end FCN. It contains only convolutional layers and no dense layer so it can accept image of any size. Similar to encoder network, each block in the decoder network drives the input to two 3x3 CNN layers and to a 2x2 up-sampling layer. Feature maps to convolution layer is down sampled by two after passing through each block to maintain symmetry. At the same time, input

get supplemented by feature maps of the corresponding contraction layer. The above step ensures the attributes learned during contraction of the image are used to recreate it. Number of contraction blocks is same as that of number of expansion blocks. Finally, the resultant blocks are passed through 1x1 CNN layer to obtain the desired segments which is equal to the number of feature maps. Pixel multiplication adding pixels between and around the existing pixels will give the final segmentation result. The flow chart of arecanut segmentation using U-Net is shown in Figure 2.

### B. Mask R-CNN

Mask R-CNN is a variant of deep neural network intended to perform instantaneous segmentation. Mask R-CNN detects objects in an image simultaneously, provides objects bounding box, class label and segmentation mask for each instance. Two-stage framework of Mask R-CNN is shown in Figure 3. Region Proposal Network (RPN), first stage of Mask R-CNN scans the image and produces proposals i.e areas where an object might be present. The second stage is a binary mask classifier which classifies the proposals, produces bounding boxes and the masks. In principal, Mask R-CNN is an extension of Faster R-CNN with an addition of RoIAlign which does pixel-to-pixel alignment. Faster R-CNN has two stages: RPN being first stage used to propose contestant entity bounding boxes. The second stage extracts fea-

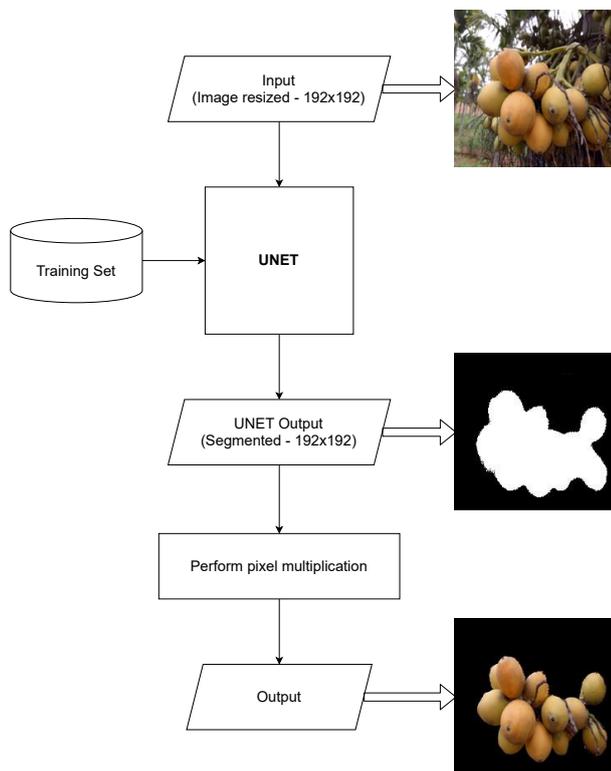


Fig. 2: U-Net pipeline of arecanut image segmentation

tures from bounding boxes and performs bounding-box regression and classification. The Mask R-CNN works in the same way but in the second stage, it outputs the mask for each Region of Interest (RoI) in addition to class and box offset. Given an image with multiple objects, classification tells to which each of the object in an image belongs. Semantic segmentation predicts labels for each pixel in the image to which class it belongs. Object detection detects bounding box co-ordinates of the object along with class label [31]. The blocks with letter 'C' are feature maps of ResNet101 and the blocks with letter 'P' are feature maps of Feature Pyramid Network (FPN).

The Mask R-CNN being a technique for instance segmentation, it identifies each pixel of the object in the image and classifies them. It does this using multiple networks. Firstly, it consists of FPN which is an extension of CNN, which is used to represent objects in different scales and is shown in Figure 4. The first pyramid generates the high level features and passes it to the second pyramid. The second pyramid passes them to lower layers. These two pyramids allow access to both low-level and high-level features at every level. Then Region Proposal Network (RPN) is used to propose regions by using anchors. The regions are proposed by RPN scanning the FPN from top to bottom, predicting anchor class and bounding box offsets for each anchor. Then a RoIAlign network is used to find appropriate areas of the feature map. It outputs multiple possible bounding boxes computed using bi-linear interpolation. RoIAlign network was used to overcome the misalignment problem

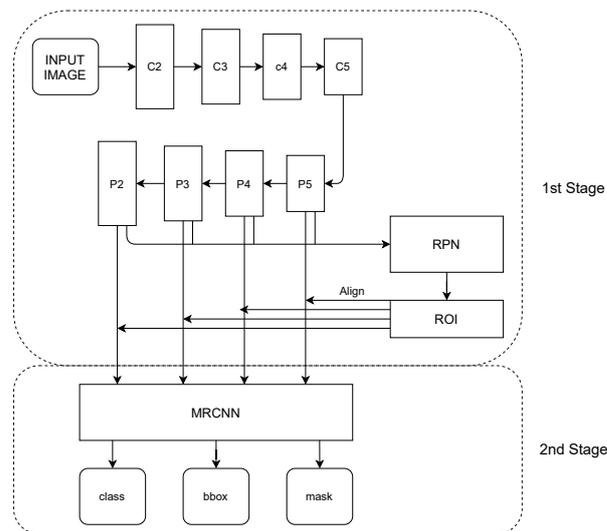


Fig. 3: Architecture of Mask R-CNN ([32])

of RoI pooling network. Then a CNN is used to produce masks taking the regions selected by RoI classifier as input. These masks are soft and are low-resolution (28x28 pixels) represented by float-point numbers holds more information than binary masks. The smaller mask size helps keep the mask part light. Mask R-CNN is a pre-trained model follows transfer learning approach and is trained using COCO data set [14]. A method that converts unmasked parts of the image to transparent pixels and masked pixels to its original values has been added. Thereby, segmenting the arecanut bunch from the background. The flow chart for arecanut segmentation using Mask R-CNN is shown in Figure 5.

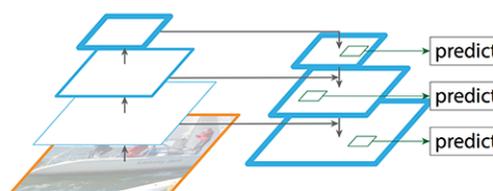


Fig. 4: Feature pyramid network [33]

#### IV. EXPERIMENTATION

The objective is to present an efficient and accurate technique for segmentation of arecanut bunches by eliminating unwanted background information. U-Net and Mask R-CNN are good compared to CNN and other traditional methods for complex images though they are successful ones for simpler images. U-Net classify each pixel where as CNN is largely used when the entire image needs to be classified. Mask R-CNN is easy to train, very efficient as it does simultaneous object detection and instance segmentation of each instance precisely in an image, easy to generalize and out performs other ex-

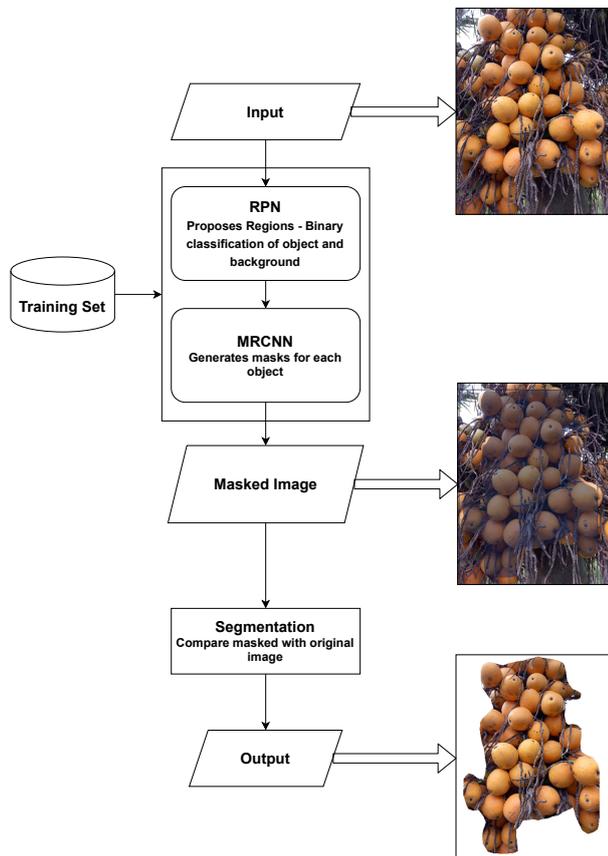


Fig. 5: Mask R-CNN pipeline of arecanut image segmentation

isting methods. To assess the methodologies, the data set made available by R. Dhanesha et al. [28] was used to train and test the performance of the networks. It consists of 1017 images which includes 388 ripe and 629 unripe images. All the images are of 4160x3120 resolution saved in jpeg format. Sample input of both ripe and unripe images are shown in Figure 6. Experiments were carried out on a database of 1017 images of which 388 are ripe and 629 are unripe. Both U-Net and the Mask R-CNN models were trained on a ripe data set of 388 images of which 310 images were used for training and 78 images were used for testing (80:20 split) and evaluated the performance of both methods on different sets of data. Deep learning based supervised techniques for semantic image segmentation requires adequate data set and proper annotations thereby requires longer training time compared to other machine learning based methods and less testing time.

Some of the labeling tasks are more complex and require experts to annotate input images. Most of the existing annotation tools depends on polygonal approximation of object boundaries. Here the method opted is encoding the object boundary of each input image by a mask and extracting the spatial structure of the object. Each mask is represented by a set of polygon points. All the annotations for Mask R-CNN and U-Net were done using the Labelme tool with each annotation saved as a json file. labelme2coco was used to convert them into

COCO json format for Mask R-CNN. labelme2voc was used to convert json to binary images needed for the U-Net model-white representing area of arecanut and black representing the background. Annotation of sample input images are shown in Figure 7. Input images and the corresponding annotations are used to train the models. The input images used are of size 192x192x3 and modified Rectified Linear Unit (ReLU) layer into Exponential Linear unit(ELU). U-net has been trained from the scratch-no transfer learning. The model has been trained for 50 epochs having set the learning rate to 0.01 for a batch size of 32 and the saved model consists of weights from the epoch with the least loss. The pixel values are normalized to [0, 1].

Unlike U-Net, Mask R-CNN has been trained using pre-trained weights for a ResNet101 used in the COCO data set. The process of transfer learning allows us to save training time and mainly allows us to work with less data. ResNet is an Artificial Neural Network (ANN) used as a backbone in many deep learning techniques. Researchers tend to add more layers to neural networks to solve complex problems and to improve the accuracy. Training becomes difficult and the accuracy starts saturating and then degrades as we go on adding layers. ResNet helps us to solve this problem and allows us to train with 150+layers successfully. Low-level features such as corners and edges are detected by beginning layers and high-level features such as nut, fruit or tree are detected by delayed layers. This backbone network converts image from 1024x1024x3 (RGB) to a shape feature map of 32x32x2048 resolution and it becomes the input for FPN. Images are resized to 800 pixels with a one sampled RoI. The model was trained for 10 epochs and the best model was saved. RoI is treated as positive if the IoU between RoI and its ground truth is greater than or equal to 0.5 and negative otherwise. The models have been trained using one GPU for a batch of size one image assigned per GPU with a learning rate of 0.01. The learning momentum is set to 0.9 and a weight decay to 0.0001. Table 1 summarizes the training accuracy of both methods on different data sets.

Table 1: Training accuracy

Model	IoU	Pr	Re	F1
U-Net	57.04%	65.11%	85.63%	70.78%
MRCNN	67.88%	81.07%	83.26%	79.36%

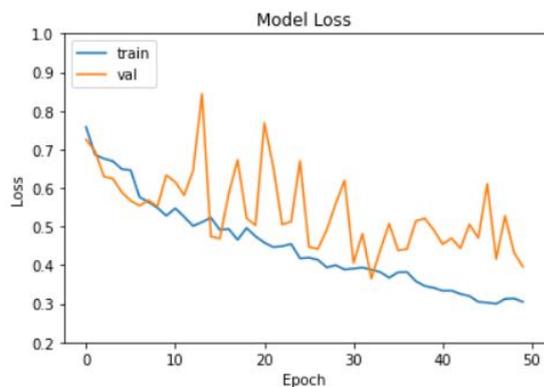
Machine learning algorithms learn from training data set incrementally and learning curves are widely used to diagnose the machine learning algorithms. U-Net and Mask R-CNN have been trained for 50 and 10 epochs respectively to get the full learning curve of the models. The models learning performance are depicted in Figure 8. Graphs describes the changes in learning performance for different epochs over time. The U-Net validation loss reached its lowest at 32<sup>nd</sup> epoch and Mask R-CNN validation loss reached its lowest at 6<sup>th</sup> epoch. Therefore, U-Net and Mask R-CNN models have been chosen with



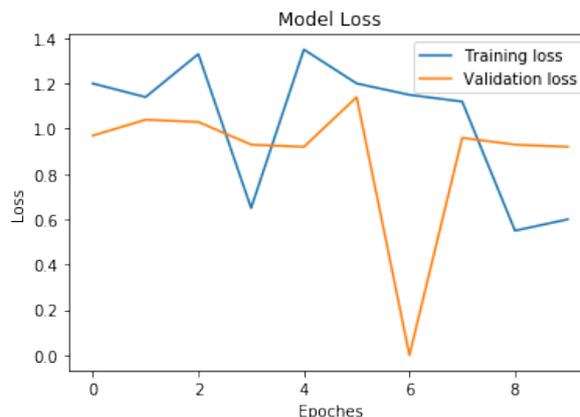
Fig. 6: Sample input images



Fig. 7: Annotation of sample input images



U-Net



Mask R-CNN

Fig. 8: Training and validation loss

Table 2: Comparison of the results obtained

Author	Method	Commodity	Data size	IoU	Pr	Re	F1
Ron et al [23]	GDA1	Grapes	100		55.16%		
	GDA2				53.54%		
	GDA3				53.39%		
D S Guru et al [18]	Thresholding	Mango	44		72.77%	71.43%	66.7%
	Binarization						
	Template Matching						
Philippe et al [26]	CNN	AppleA	147	71.4%	83.3%	87.7%	79.4%
		AppleB		63.0%	77.3%	91.2%	67.1%
		AppleC		59.0%	74.2%	64.8%	86.8%
		AppleD		75.4%	86.0%	79.2%	94.1%
Proposed	U-Net	Ripe areca	388	54.61%	61.53%	87.07%	68.26%
		Unripe areca	629	58.07%	74.71%	77.15%	72.95%
	MRCNN	Ripe areca		61.01%	73.57%	81.84%	72.95%
		Unripe areca		65.98%	89.86%	73.14%	78.68%

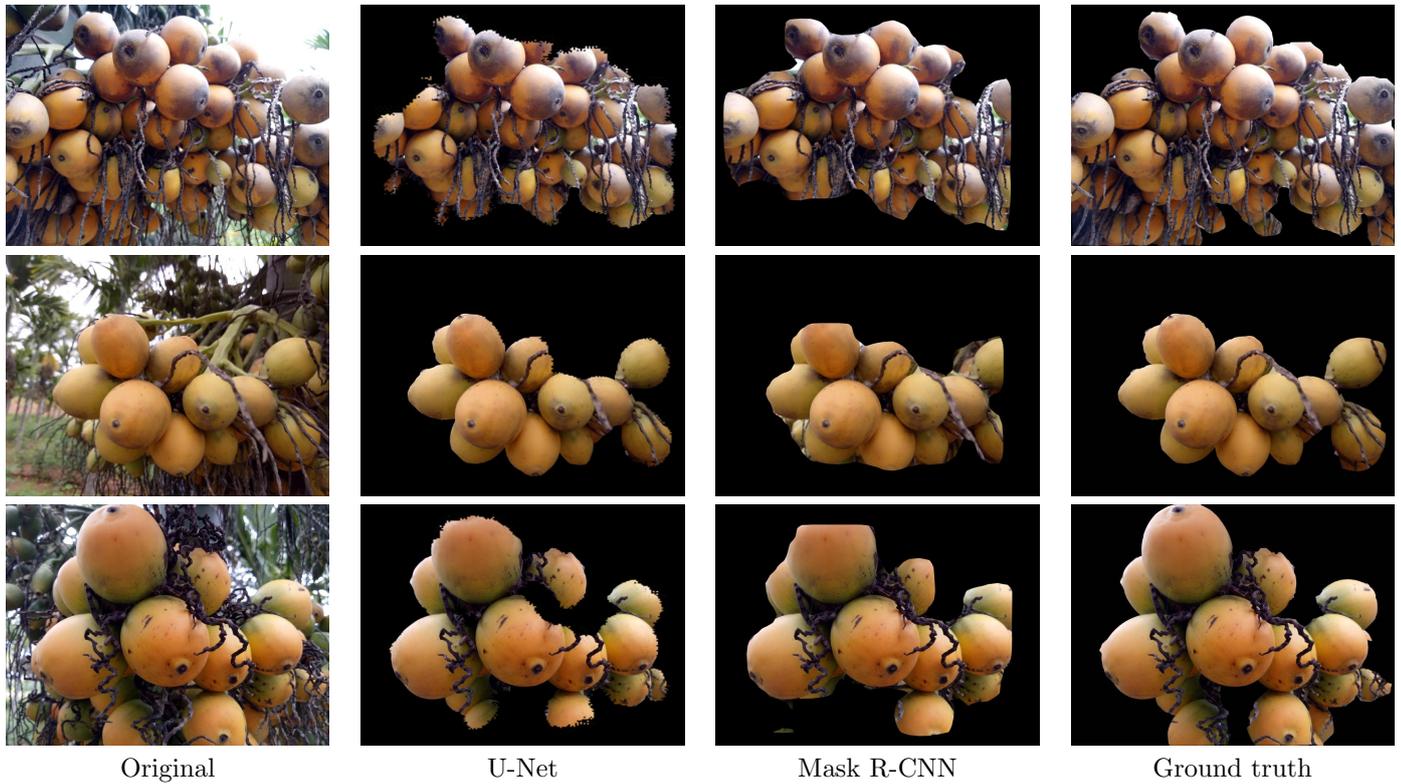


Fig. 9: Comparison of results of both the methods against ground truth for ripe images



Fig. 10: Comparison of results of both the methods against ground truth for unripe images

these lowest validation losses for our experiments.

## V. RESULTS AND EVALUATION

The objective is to present an accurate and efficient method for arecanut bunch segmentation from an image acquired in field conditions. To evaluate the per-

formance of above methods, experimentation has been carried out on both the data sets. Both the presented methods do not needed any pre-processing. Four standard measures were used for check the performance of both the models: IoU, Precision (Pr), Recall (Re) and F1-Score (F1) depicted by (1) to (4) computed at pixel-level. Though, there are very limited attempts made for the segmentation of arecanut [15] [16] and also the performance evaluation done only for very few images, methods that apply segmentation on other commodities are used as baselines for comparison. Table 2 summarizes the test performance of both the methods in comparison with the other commodities. The higher values of above measures indicates the greater performances of segmentation algorithms. The applicability of the methods are exhibited by their high segmentation accuracy across both the data sets that differ with regard to color, background and resolution. Our segmentation performance is better compared to grape [23] and mango [18] segmentation performance. This may be improved further with more training examples. Here, the challenges are more due to inflorescence and occlusion compared to other agricultural crops. The sample segmentation outputs achieved by the models are presented for qualitative differentiation of both ripe and unripe images in Figure 9 and Figure 10 respectively. Results obtained using Mask R-CNN is better compared to U-Net which is closest to the ground truth images in the fourth column. Both the models are implemented on Google Colab (virtual machines on cloud) using Intel Xeon CPU@2GHz with 12GB RAM and NVIDIA Tesla T4 GPU with 16GB RAM, MRCNN takes 15 seconds and U-Net takes 3 seconds to segment each image.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

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

$$Re = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 \bullet \frac{Pr \bullet Re}{Pr + Re} \quad (4)$$

## VI. CONCLUSIONS

In this paper, two deep learning based state-of-the-art techniques are presented for segmentation of arecanut bunch. The methods illustrated greater segmentation accuracy across both the data sets that differ with regard to background, illumination, image resolution, nuts density and color. The models are trained using only ripe arecanut and succeeded in generalizing for unripe arecanut which illumination, are considerably vary in terms of color without any pre-processing. Experiments were conducted to evaluate the accuracy, efficiency and robustness of the methods. Outcomes shows that methods gives favorable performance in contrast to other methods that apply segmentation on other commodities. Future

research direction is to improve the segmentation performance and find the yield count of arecanuts in a bunch.

## ACKNOWLEDGEMENTS

The authors would like to thank R Dhanesha and Shrinivasa Naika C L [28] for making available the data set for sharing their data set. The link for the data set is <http://davangereuniversity.ac.in/arecanut-database/>.

## REFERENCES

- [1] Rasmus Houborg Matthew F. McCabe and Arko Lucieer. High-resolution sensing for precision agriculture: from earth-observing satellites to unmanned aerial vehicles. *Remote Sensing for Agriculture, Ecosystems and Hydrology*.
- [2] T.N. Prakash B.J. Rajkumar and B.N. Pradeepa Babu. Indigenous technical knowledge (itk) and farmers willingness to practice- an economic study of arecanut peeling machines in karnataka. *International Research Journal of Agricultural Economics and Statistics*, 6(2):307–316, 2015.
- [3] Lei Zhang Bo Peng and David Zhang. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 46:1020–1038, 2013.
- [4] Y. Wang Z. H. Yu X. F. Zhang C. N. Li X. D. Bai, Z. G. Cao. Crop segmentation from images by morphology modeling in the cie l\*a\*b\* color space. *Computers and Electronics in Agriculture*, 99:21–34, 2013.
- [5] Martin Glavin Esmael Hamuda and Edward Jones. A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, 15:184–199, 2016.
- [6] P P Subedi D Jarvis A B Payne, K B Waish. Estimation of mango crop yield using image analysis segmentation method. *Computer and Electronics in Agriculture*, 91:57–64, 2013.
- [7] Marcel Bergerman Sanjiv Singh Qi Wang, Stephen Nuske. Automated crop yield estimation for apple orchards. *Experimental Robotics*, 344:745–758, 2013.
- [8] Dani Martinez Javier Moreno Eduard Clotet Davinia Font, Marcel Tresanchez. Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors*, 3:8284–8301, 2015.
- [9] Huaibo Song Chang Liu Dandan Wang, Dongjian He. Combining sun-based visual attention model and saliency contour algorithm for apple image segmentation. *Multimedia Tools and Applications*, 2019.
- [10] V N Manjunath Aradhya S Siddesha S, S K Niranjana. Segmentation of coconut crop bunch from tree images. *Second Int. Conf. on CCIP*, 2016.
- [11] Manjunath Aradhya S Siddesha S, S K Niranjana. Segmentation of oil crop bunch from tree images. *Int. Conf. On Smart Technology for Smart Nation*, pages 1621–1626, 2017.

- [12] S. Amatya M. Karkee Q. Zhang K. Lewis A. Gongal, A. Silwal. Apple crop-load estimation with over-the-row machine vision system. *Computers and Electronics in Agriculture*, 120:26–35, 2016.
- [13] Amos Naor Oded Cohen, Raphael Linker. Estimation of the number of apples in color images recorded in orchards. *IFIP*, 344:630–642, 2011.
- [14] Malrey Lee Ulzii-Orshikh Dorj and Sang seok Yun. An yield estimation in citrus orchards via fruit detection and counting using image processing. *Computers and Electronics in Agriculture*, 140:103–112, 2017.
- [15] S K Niranjana Siddesha S and V N Manjunath Aradhya. A study of different color segmentation techniques for crop bunch in arecanut. *Handbook of Research on Advanced Hybrid Intelligent Techniques and Applications*.
- [16] Shrinivasa Naika C.L. Dhanesha R. A novel approach for segmentation of arecanut bunches using active contouring. *Integrated Intelligent Computing, Communication and Security. Studies in Computational Intelligence*, 771:677–682, 2019.
- [17] Toni P. Saarela Landy and Michael S. Combination of texture and color cues in visual segmentation. *Vision Research*, pages 56–67, 2012.
- [18] H G Shivamurthy D S Guru. Segmentation of mango region from mango tree image. *LNAI*, 8284:201–211, 2013.
- [19] Fernando Auat Cheein Giancarlo Troni Rodrigo Perez-Zavala, Miguel Torres-Torriti. A pattern recognition strategy for visual grape bunch detection in vineyards. *Computers and Electronics in Agriculture*, 151:136–149, 2018.
- [20] Isabelle Bloch Ujjwal Verma, Florence Rossant. Segmentation and size estimation of tomatoes from sequences of paired images. *EURASIP Journal on Image and Video Processing*, 33:1–23, 2015.
- [21] Jan Willem Hofstee Hyun K. Suh and Eldert J. van Henten. Improved vegetation with ground shadow removal using an hdr camera. *Precision Agric*, 19:218–237, 2018.
- [22] Ashfaque Andrew Hellicar. Identification of mature grape bunches using image processing and computational intelligence methods. *IEEE*, 2014.
- [23] Ohad Ben Shahar Ron Berenstein and Amir Shapiro. Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer. *Intel Serv Robotics*, 3:233–243, 2010.
- [24] Zachary Taylor James Underwood Salah Sukkarieh Calvin Hung, Juan Nieto. Orchard fruit segmentation using multi-spectral feature learning. *IEEE Int. Conf. on Intelligent Robotics and Systems*, 77:5314–5320, 2013.
- [25] James P. Underwood Suchet Bargoti. Image segmentation for fruit detection and yield estimation in apple orchards. *Computers and Electronics in Agriculture*, 2016.
- [26] Henry Medeiros Philippe A. Dias, Amy Tabb. Multispecies fruit flower detection using a refined semantic segmentation network. *IEEE Robotics and Automation Letters*, 3:3003–3010, 2018.
- [27] Omkar Narasipura Ramesh Kestur, Avadesh Meduri. Mangonet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open field. *Engineering Applications of Artificial Intelligence*, 77:59–69, 2019.
- [28] Shrinivasa Naika C.L R Dhanesha. and Y. Kantharaj. Segmentation of arecanut bunches using ycgcr color model. *1st International Conference on Advances in Information Technology (ICAIT)*, pages 50–53, 2019.
- [29] Thomas Brox Olaf Ronneberger, Philipp Fischer. U-net: Convolutional networks for biomedical image segmentation. *MICCAI 2015, LNCS 9351*, page 234241, 2015.
- [30] Sutskever I. Hinton G.E. Krizhevsky, A. Imagenet classification with deep convolutional neural networks. *In: NIPS*, page 11061114, 2012.
- [31] T.-Y. Lin et al. Microsoft coco: Common objects in context. *In ECCV*, 2014.
- [32] Kaiming He et al. Mask r-cnn. *Computer Vision and Pattern Recognition*, 2018.
- [33] Tsung-Yi Lin et al. Feature pyramid networks for object detection. *Computer Vision and Pattern Recognition*, 2017.

## **Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)