iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.41442

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Fruit Detection and Localization using UNET

D. Veda Swaroop[1], D. Mahesh[2], Gogula N S S K. Pavan Dheeraj
[1, 2, 3]*School of Mechanical Engineering, Sastra Deemed to be University, Tirumalaisamudram, Thanjavur — 613 401*

***Abstract:** Accurate yield estimation helps the farmers to make better decision in crop management. The primary task for accurate estimation of fruit yield is to detect and localize the fruit in the field. This problem is explored in our work by using U-Net architecture. It is one of the deep learning based on semantic segmentation architecture used for better object detection and localization. U-Net consists of two paths namely contraction and expansion path. Contraction path acts as an encoder and extracts features from the source object (i.e., mango images) and expansion path acts as a decoder that recovers the resolution of the image for better localization. The mango fruit was chosen for fruit detection and used the ACFR mango dataset. The dataset was split in to train, validation and test images. The training options of epoch 10, batch size 16 and an Adam optimizer were used. The train dataset wastrained with the above said training options using U-Net model and evaluated with the performance metric accuracy and binary cross entropy. The developed model was tested with the test images which was not the part of training. The prediction accuracy, loss of test image were 98.66% and 0.0268% respectively.*

I. INTRODUCTION

Background: Farming consists of some very important wholesome processes which includes crop selection, land preparation, seed selection, sowing, irrigating the land, crop growth, fertilizing and harvesting. These processes can be taken as a chain steps required for the production of yield, the yield can be any thing like fruit, rice, paddy etc. Every step as its own weightage depending on the complexities that occur during the completion of that process. And India is a country which gives more importance to the agriculture sector other than any sector. Agriculture and its sub sectors, are the biggest source of livelihood in India. And about 70% of rural households in India primarily depends on the agriculture. The only way that these people will get benefited by introducing various smart techniques using modern technologies to minimise the cost of investment for production. By adopting to these techniques we can eliminate the need of more labour, requirement of special equipment, need of expert opinion etc., which can minimise the cost of investment for the yield and a farmer can make some more profits than earlier. For this we concentrated our project on automating the harvesting step in the farming.

Harvesting is a step which includes collection of fruits from the trees, quality check and finally counting the fruits. Usually this step requires more labour, expert opinion and it is also a time consuming process. And also this step has more weightage in terms of cost terms required for the completion of this step.

Because this step is a labour intensive implies more capital on wages to labour, requires expert opinion this should also include some capital and more importantly time consuming. All these problems are eliminated by automating the yield estimation process and by this we can also automate the harvesting process with the help of harvesting machines. By adapting above we can obtain a precise value of yield that will obtain from the field. And now we can undoubtedly focus on the marketing step, which is the final and an important step in farming. Now the farmer can attain more profit compared to earlier, as the above technique ensures less capital on labour wages and expert opinion. And to conclude our problem statement would be the 'Fruit detection and localisation in orchard's'. Our project 2 work is mainly focused on the detection and localisation of fruits in orchard's. We used deep learning techniques for detection and localisation of fruits, as we there are vast number of deep learning approaches to solve our problems. For our problem, we can obtain the solution by choosing the segmentation deep learning model like U-Net, deeplab etc. And the image segmentation is of two types, they are semantic segmentation and instance segmentation. Image segmentation is a pixel wise classification or masking on each object in the image, in simple words classifying each pixel of whether it belongs to a object class or a background.

The segmentation is mainly deal in classifying the pixels, in other words it is also known as labelling the image data. The type of segmentation, semantic segmentation is the task of clustering the pixels in the image together in belong to the same object class. It is a process of predicting the each pixel in the image because each pixel in an image is classified to a object class or background class. And the other type, instance segmentation is also similar to semantic segmentation in masking the pixels belonging to a object class. It is the task of labelling the pixels in the image with different weights irrespective of the object class. And the semantic segmentation model used in our project is U-Net. The below image depicts the semantic segmentation and instance segmentation.

Fig. 1.1 Semantic segmentation

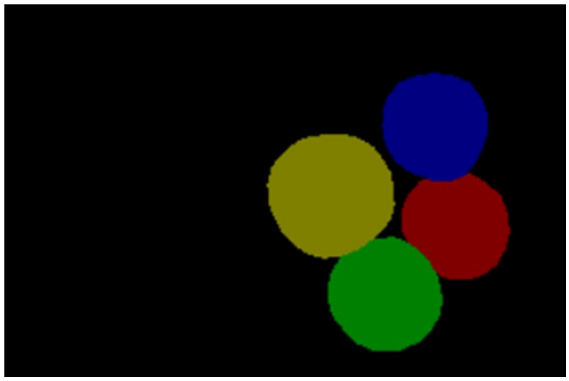


Fig. 1.2 Instance segmentation

The figure Fig. 1.1 is a semantic segmented image, in which the objects are made foreground according to the object class. And in the Fig. 1.2 instance segmented image, in which each object is masked in different colors representing a instance object class. Image segmentation has enormous application in today's industry such as Self-driving cars, object localization and nuclei analysis in bio-medical studies etc. Image segmentation is an important computer vision technique that is driven to understand the information in the image.

II. LITERATURE SURVEY

1) Wenkang chen et al., 2020, Detecting citrus in orchards using improved YOLOv4,
Discussed about the fruit detection of citrus fruit in orchard environment using YOLO object detection algorithm and got better results for occluded regions.

2) Suchet Bargoti et al., 2017, Deep fruit detection in orchards
Mango and almond fruit using Faster R-CNN object detection algorithm, in which the model obtained good accuracy for every fruit 0.85 on average.

3) Hanwen Kang, Chao Chen, 2019, Fruit detection and segmentation for apple harvesting using virtual sensors in orchards.
Discussed about the DaSNet architecture to detect fruit and about the role of data augmentation for improving accuracy and Image segmentation was also performed for output.

4) Kushtrim Bresilla et al., 2019, Single shot CNN's for real time fruit detection within the tree
Used single stage YOLO object detection algorithm for detection of apple in orchards and the obtained accuracy is 0.9.

5) Harshall Lamba, 2019, Understanding semantic segmentation with U-Net
Discussed about semantic segmentation applications with problems in real industry and about the architecture, training of the U-Net model.

6) Ayyuce kizrak, 2019, Deep learning for Image segmentation
Discussed about the importance of the segmentation and detailed about the various layers present U-Net architecture for semantic segmentation.

7) Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation.
This is the base paper for the U-Net architecture that is used for segmenting the pixels in the image.

8) Bargoti, S., & Underwood, J. (2016). Deep Fruit Detection in Orchards.
This the dataset, that we used to train model, to obtain the fruit areas in the image.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

III. METHODOLOGY

This project is about detection and localisation of object, which involves deep learning models to solve the problem. Usually any deep learning technique involves these steps to achieve the problem, this project too consists of those steps to obtain the required solution. The steps that consist our methodology are data acquisition, data pre-processing, data annotations and data augmentation, model building, training the model, prediction on test images and evaluating the performance of the model. We use several python libraries to accomplish these steps, they are tensoflow, os, OpenCV, split-folders, skimage, numpy etc.

A. Data Acquisition

The dataset used for this project is ACFR multifruit 2016 dataset, which contains images of three different tress in orchards, they are apples, mangoes and almond trees. And the dataset also contains the annotations i.e., the location or co-ordinates of the fruit in the image these are circle annotations for apples and rectangular annotations for the mangoes and almonds. The annotations were in the CSV format, which gives data for making the mask images of the original images for the training of the model. In this project we only took the mango images for our experiment. This mango dataset contains 1964 images in PNG format, 500x500 size, 16-bit/color RGB images and the camera sensor used to capture these images is Prosilica GT3300c strobes. The orchard in these images is located at Bundaberg, Australia. The variety of the mango tree is Calypso, and the pictures were taken in low light conditions. The rectangular annotations are in (x,y,dx,dy), which gives the length of the rectangle as x+dx and the breadth of the rectangle as y+dy. And by knowing the length and breadth we can form a rectangle to obtain the mask data of the original images.

B. Data Pre-processing

This process includes splitting the dataset, resizing the images and necessary operations that should made on the data before feeding in to the built model for training. Initially the whole dataset is split in to three folders namely train, validation and test folders in any ratio giving greater to train folder and remaining to the validation and test folder equally. The ratio we used is 8:1:1 i.e., 8 parts to the train and two equal parts to each validation and test folders. After splitting there are 1571 images in train, 196 and 197 images in validation and test folders respectively. This task is achieved by using the ratio method in the split-folders python library. And for resizing the images, read the image in same mode as it is in the dataset 5 using OpenCV library, resize the images across the channels of the image using the resize method in transform class in skimage python library.

C. Data Annotation

Annotating the images is also called as labelling the images. It is a task of labelling the each pixels whether the pixel is belonging to any of existing object class or background class. There are so many opensource platforms for performing annotations on the data like labelme, cvat etc. And there so many export options of the annotated data via these annotating platforms, the format that we use in our project segmentation mask 1.1. As the required annotations are given in the dataset in CSV format. With help of python libraries, we have made the segmentation masks of the original images from the data stated above in data acquisition. Some of the original image and masked image pairs are as shown below

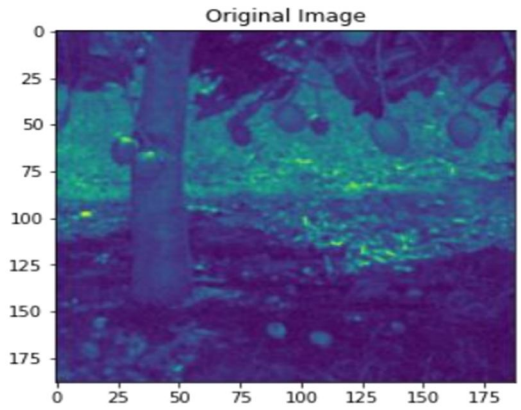


Fig. 2.3.1.Sample Image 1

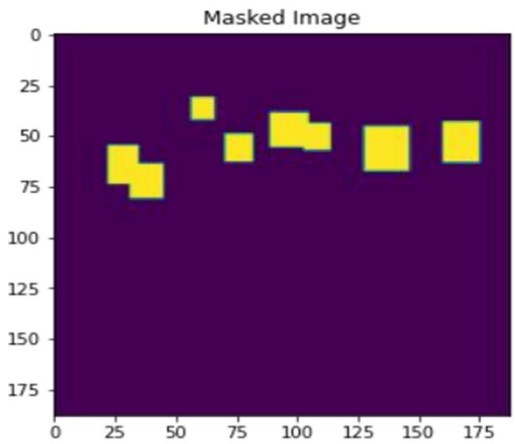


Fig. 2.3.2.Sample Image 1

Those extracted coordinates of the rectangles are drawn on a blank image of same size of original image. A blank image is created using numpy python library by initializing a numpy array of zeros and of same size as the original image. Now the rectangle can be drawn on these blank images using the rectangle method in OpenCV python library. Once the completion of drawing all the rectangles on the blank image, the mask image is saved with name of original image. After finishing this task of masking images and saving them, repeat the data preprocessing step to make the mask images ready to feed in to the model along the original images of the trees. By this we can conclude the data annotation step.

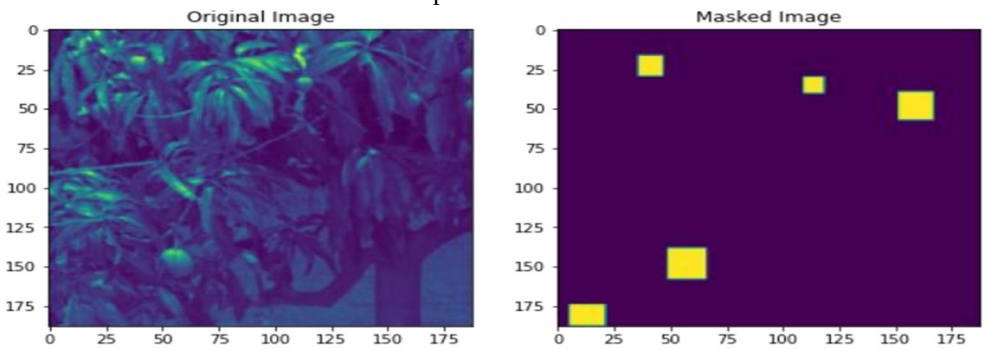


Fig. 2.3.3. Sample image 2

D. U-Net

It is a deep learning architecture used to solve segmentation problems in industry. This model is named as the shape of the architecture resembles the alphabet 'U'. This architecture can perform both semantic and instance segmentations depending upon the application and the activations in layers used to built the architecture. Mainly this architecture consists of layers like conv2D, maxpool2D, dropout, concatenate, conv2D transpose layers. These layers combinedly forms two phases in the architecture namely contraction phase and expansion phase. The clear picture of the architecture is shown in Fig.2.3. This is mainly founded by making intention to solve the biomedical problems in localisation of various cells in images captured by the microscopes. As we discussed earlier the left part of the architecture is called contraction or encoder path and the right part is called as expansion or decoder path. The encoder path consists of sequential 3x3 feature size convolutional layers of same padding with some features and down sampling of these features is done max pooling layers of stride 2x2, additionally dropout layer is added before max pooling layers to prevent overfitting of data.

The decoder path deals in up sampling the data followed by 2x2 up-convolutional layers which is conv2D transpose layer in keras, that halves the number of feature channels. These maps are concatenated with the corresponding down sample feature map in encoder path and followed by same 3x3 feature size convolutional layers. This concatenation of features maps from the encoder path with the one in decoder is responsible for localising the objects and made semantic segmentation possible. And finally a convolutional layer of feature size 1x1 is applied to the 7 feature map and the resulting map is the probabilities map of pixels of the image, in belonging to the object classes present in the image. This model can be built using layers from keras application programming interface in tensorflow python deep learning library.

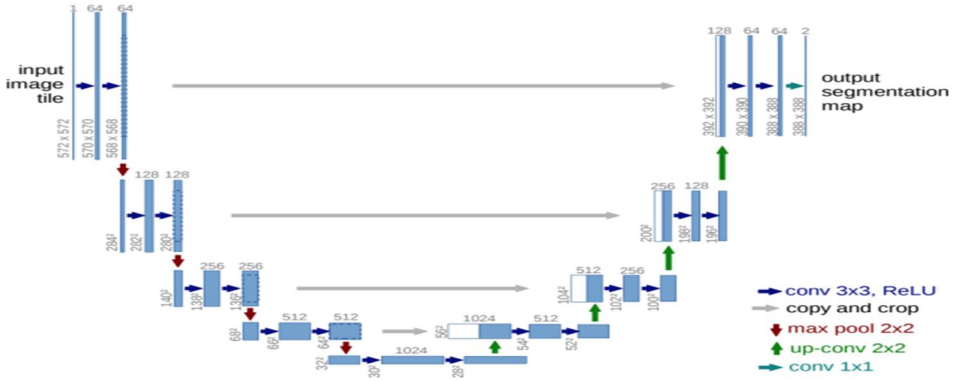


Fig. 2.4. U-Net architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Now we have all the model, training data and validation data ready to perform the training operation for obtaining the segmentation results. And to make the model ready for training, one should compile the model with a suitable optimizer, loss function and training metrics. The optimizer which can be used for our problem can be adam, sgd etc., with some learning rate which achieves stochastic gradient descent conditions with the existing features. And the loss function used is binary cross entropy, the training metrics used can be accuracy of the prediction of pixels or any depending on the loss function.

Table. 2.4.1 Compilation parameters

Parameters	Values
Optimizer	Adam
Loss function	Binary crossentropy
metrics	Accuracy

E. Training

After compiling the model with required parameters, now our model is ready for training and to learn from the feed data using supervised learning. Fit function in the tensorflowmodel library enables us to feed the train data, validation data and specifying the batch size, epochs, verbose, callbacks and whether to shuffle the data or not, in our case we do not shuffleour data because our data slices should contain original image and corresponding mask imageof the original. In some cases, one cannot suffice the validation data separately this can be tackled by using validation split parameter in the fit function. It usually enables us to use a split of training data as a validation data. The parameters used in the fit function are as shown belowTable 2.1

Parameter	Value
X	X_train
Y	Y_train
Batch size	16
Epoch	25
Verbose	1
Validation data	(X_val, Y_val)
Shuffle	False

Table. 2.4.2. Training parameters

IV. RESULTS AND DISCUSSION

After completing the training process, we can predict the results by passing a stack ofimage to the model and can visualize it. The training process should take some time as it tries to learn from the data and performs prediction on validation data. And for every epoch in training process, the model will output some parameters namely loss, accuracy, validation lossand validation accuracy. And the parameters that got in final epoch, while training are considered as the results of the training. In our training the parameters that we got in the final epoch are as shown in Table. 3.1.

Parameter	Value
Train loss	0.0222
Train accuracy	0.9872
Validation loss	0.0190
Validation accuracy	0.9892

Table. 3.1 Training results

And we can plot the summary of whole training process using graphs between epoch and accuracy, epoch and loss. All the parameters of the training process will be get stored in adictionary, those values can be retrieved and can be used for plotting the results. These graphsare used to evaluate the performance of the model. The model is said to be good if the epoch vs accuracy graph is exponential and if the epoch vs loss graph is a hyperbolic curve.

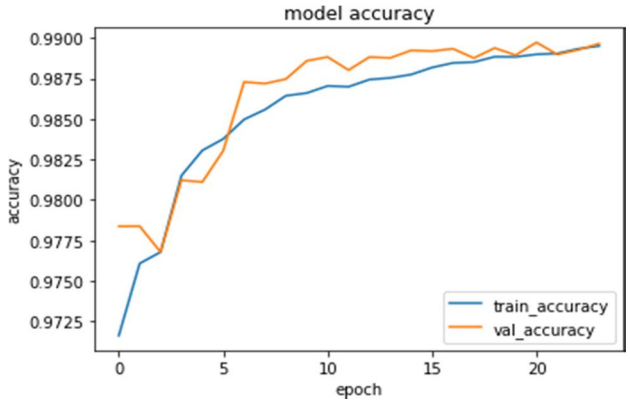


Fig. 3.1 Epoch vs Accuracy

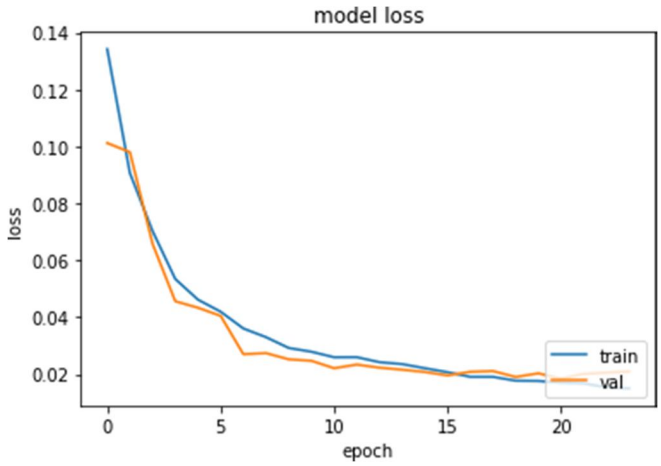


Fig. 3.2 Epoch vs Loss

The above are the graphs that are obtained from the history of the model training.Now we can predict the results of the test images for this we have to perform all the data pre- processing steps that are done to the train and validation datasets to the tests dataset. We can also evaluate the loss and accuracy in predicting the test image stack using evaluate function in directory of model. By performing this, we obtained loss of 0.0268 and accuracy of 98.65%.And the predicting is done by using predict function in model directory. The predict function expects the input images in a stack, for this we can stack the images using some numpy operations. Now we got a stack of probability maps, these maps are indexed as per the indexesin the input stack. These probability maps will contain probabilities of each pixel in that position in belonging to object class. We should make one threshold value for the probabilities, the pixel of more probability than threshold value will be made as ones and all other as zeroes.And thus we obtain a binary image, which is mask image of the passed test image.

The white colour area on the images denotes the area is belonged to the fruit region. At first lets comparethe results obtained by passing trained image as the input, for that we should stack the image using vstack function in numpy library. And now pass the stack to the predict function in model directory, set a threshold probability value to get the binary image. We can plot the results and compare the obtained binary image with the ground truth mask image of the original image. Now we can see that the fruit regions were exact rectangles in the ground truth mask, but in the predicted mask will be not the exact rectangles because we should make labels in the groundtruth mask as the exact shape of the fruit in the image. But we used those rectangular annotations in the dataset and generated the masks. If we use the labels in masks image as exactshape as of the fruit in the images, we will get more accurate results.

For plotting the results we used methods in pyplot class in matplotlib library in python.The below are predictions on some train images.

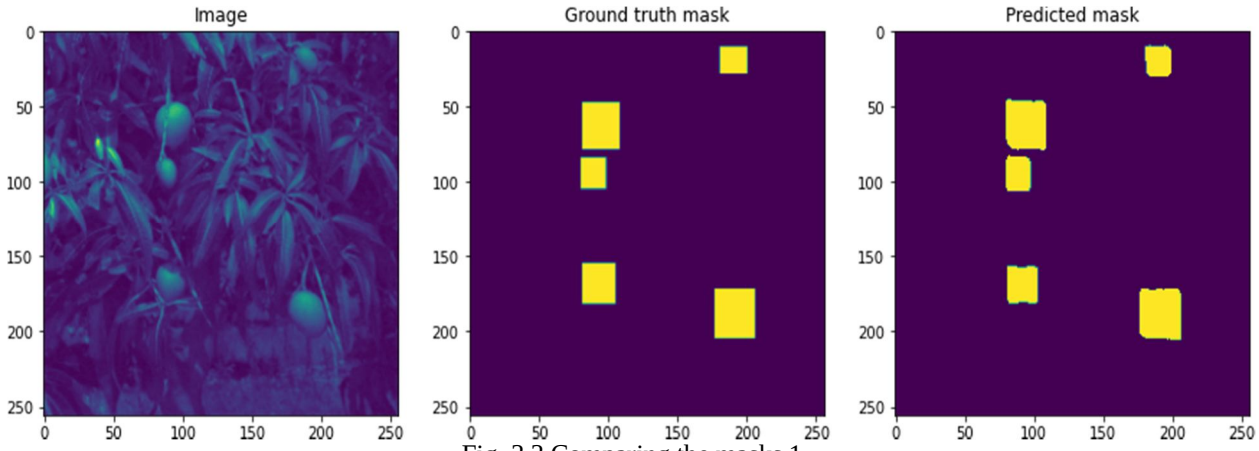


Fig. 3.3 Comparing the masks 1

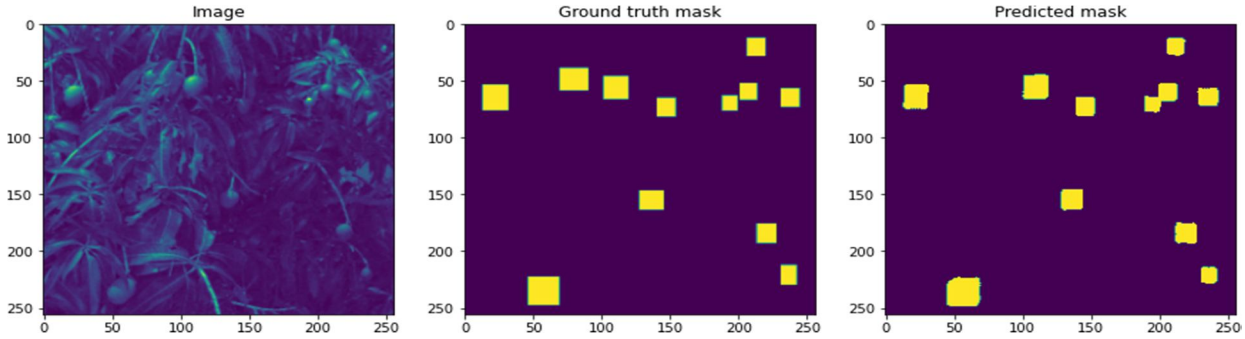


Fig. 3.4 Comparing the masks 2

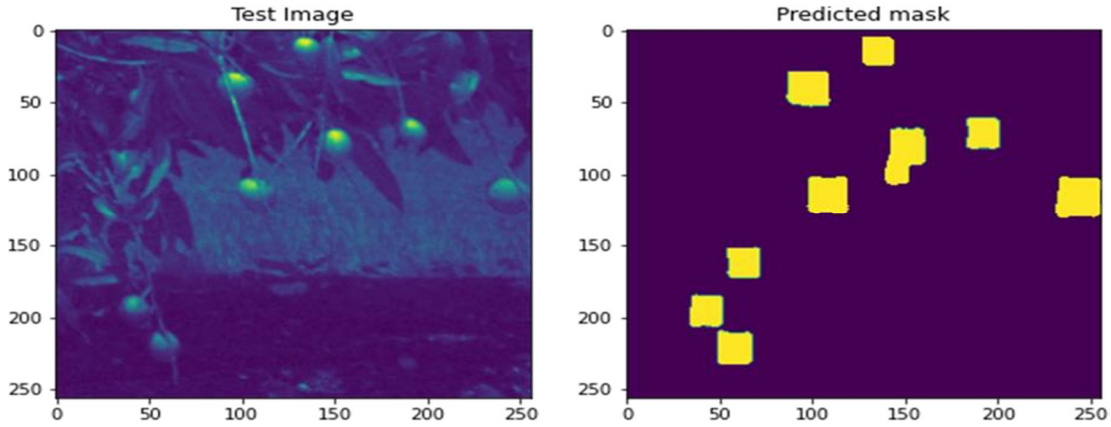


Fig. 3.5 Test image prediction

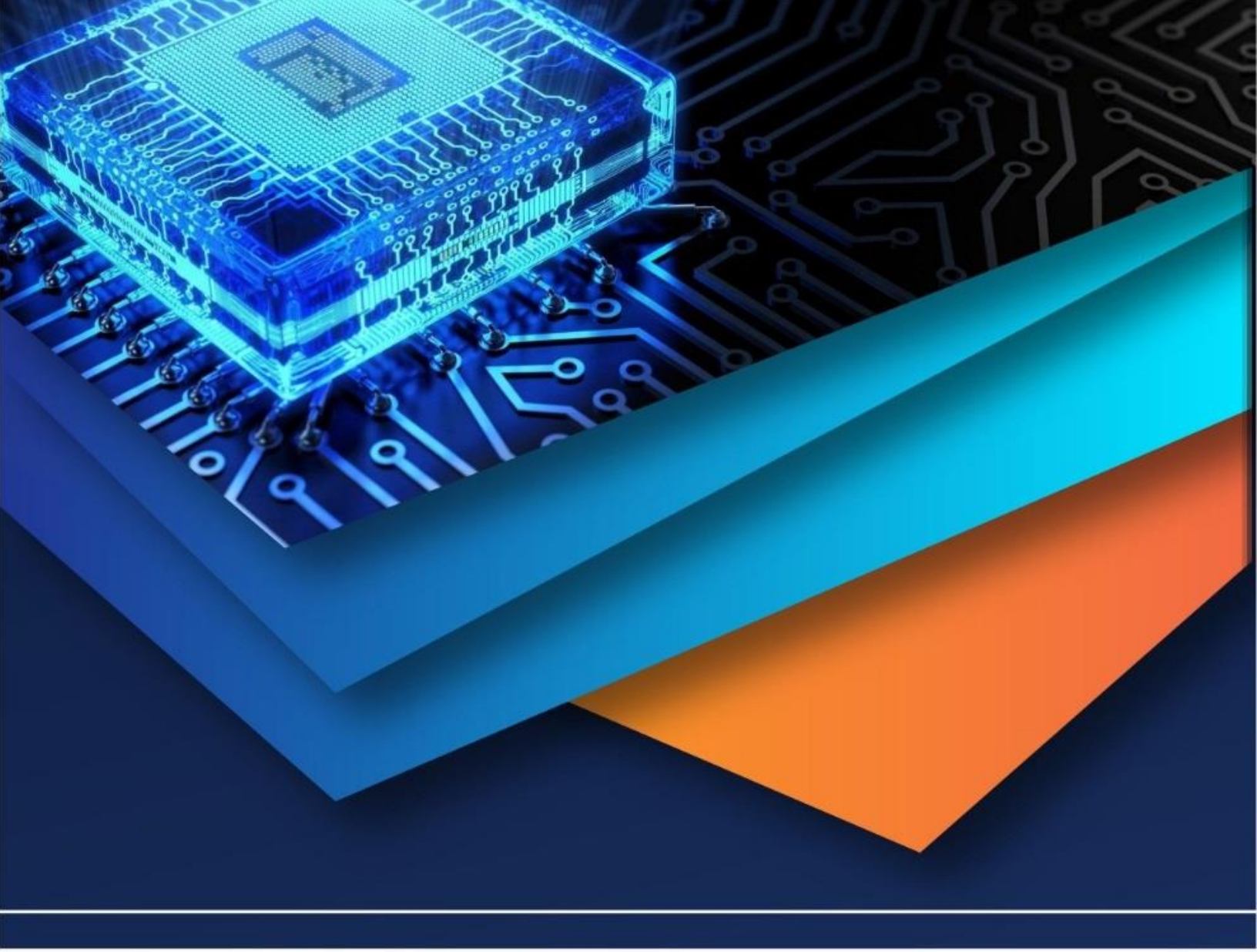
V. CONCLUSION

Now we can conclude that we have built a model that can detect and localize the fruits in the images that are taken from orchards using cameras with specified sensors. Further studieswill be focused on counting the fruits in the trees of orchards for estimating the yield. Our studycan satisfy up to detecting and localization of fruit areas in the tress and for counting the we can use computer vision techniques to predict the count of the fruits in the trees. By counting the fruit areas in each test image and summing up them gives a prediction of yield in orchards.We can get the count of the yield before harvesting the fruits. By estimating the yield, we canswitch to automatic harvesting techniques using harvesting machines with out any fear of getting loss. Our built model along with counting algorithm can be embedded to make a roboticsystem for precise estimation of yield in orchards.

And we can use several other parameters while training the model like changing the training metrics according to the loss function. We took this problem as a semantic segmentation problem, there are many other techniques that can solve this problem using various other object detection algorithms. If we have larger images that should feed in to the model, we can make slices of the input image along with the masked image to predict the results. By adapting these techniques in regular farming, the farmer can get good profits as it eliminates the requirement of skilled labor in estimating the yield than earlier.

REFERENCES

[1] Wenkang Chen, Shenglian Lu, Binghao Liu and Tingting Qian, Detecting Citrus in OrchardEnvironment by Using Improved YOLOv4 https://www.hindawi.com/journals/sp/2020/8859237/
[2] Suchet Bargoti, James Underwood, Deep Fruit Detection in Orchardshttps://ieeexplore.ieee.org/document/7989417
[3] Hanwen Kang, Chao Chen, Fruit Detection and Segmentation for Apple Harvesting UsingVisual Sensor in Orchards. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6832306/
[4] Kushtrim Bresilla, Giulio Demetrio Perulli, Alexandra Boini, Brunella Morandi, Luca Corelli Grappadelli and Luigi Manfrini, Single Shot CNN's for Real-Time Fruit Detection Within the Tree. https://www.frontiersin.org/articles/10.3389/fpls.2019.00611/full
[5] Harshall Lamba, Understanding Semantic Segmentation with UNET. https://towardsdatascience.com/understanding-semantic-segmentation-with-unet- 6be4f42d4b47
[6] Ayyuce kizrak, Deep learning for Image Segmentation. https://heartbeat.fritz.ai/deep-learning-for-image-segmentation-u-net-architecture-ff17f6e4c1cf
[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks forBiomedical Image Segmentation. https://arxiv.org/abs/1505.04597v1
[8] Bargoti S, James Underwood, Deep Fruit Detection in Orchards. https://arxiv.org/abs/1610.03677v2



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY