Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7765845

CROP IDENTIFICATION USING MASK R-CNN

[1] Mrs. V. Vidhya, [1] Dr. G. Victo Sudha George, [1] Dr. P. Dinesh Kumar
[2] Ms. Sushmitha.A,[2] Ms. Shalu kumari.
[1] Professor, Department of CSE, Dr.M.G.R. Educational and Research Institute, Chennai
[2] Final Year B.Tech-CSE,Dr.M.G.R. Educational and Research Institute,Chennai.
Vidhya.cse@drmgrdu.ac.in, victosudhageorge@drmgrdu.ac.in,
dineshkumar.it@drmgrdu.ac.in, sushmithaarivi@gmail.com, kumarishalu463@gmail.com

ABSTRACT – Fruits & vegetables provide an essential role as a food in our everyday life. It provides nutrients vital for our health and maintenance of our body. Those who devour greater Fruits & greens as part of a wholesome food plan are probable to have decreased danger of a few persistent illnesses. This paper describes One of the primary tasks of completing and developing many computer vision applications is to identify a method of figuring out crop image. We are using Mask Region-Based convolutional Neural Network is an Artificial Intelligence algorithm based on multi-layer neural networks that learn relevant features from images, being capable of performing several tasks like image classification, object detection, and segmentation. Our model efficiently detects objects or image data by generating high-quality segmentation masks for each instance. This can develop an automatic fruit and vegetable classification system with a data set of information about each fruit and vegetable with the help of artificial intelligence (AI). After successful training, the Mask RCNN model can predict the name of the fruit and vegetable.

KEY WORDS: Deep learning; Mask R-CNN; FPN; Detection and Segmentation; fruits and vegetables.

I. Introduction

Image identification is when a computer can analyze an image and determine which class it belongs to. Object detection is a computer vision method for locating illustrations of objects in images. Our paper describes about the approach of creating a system for identifying the images where the Mask Region-based convolutional neural network will predict the name of fruit and vegetable given in the image. Recently, deep learning-based techniques have shown impressive results for most tasks in computer vision including object detection and image classification techniques are able to automatically learn massive data of various conditions and extract optimal features that provide a high performance of crop identification. Many researchers tend to use CNN to extract features for image identification and classification. Our paper proposes to use Mask R-CNN deep learning framework in the detection and segmentation of crop detection. Using Mask R-CNN for classification and segmentation achieves several improvements compared with previous literature methods.

II. Related Works

In [1] A deep learning framework utilizing an improved Faster R-CNN was employed for detecting multiple fruit classes. The framework comprised creating a fruit image library, augmenting the data, generating an enhanced Faster R-CNN model, and evaluating its performance. The researchers constructed a library of fruit images and fine-tuned the convolutional and pooling layers of the model to optimize its performance. The framework achieved an accuracy rate of over 91% using the improved Faster R-CNN model. In [2] To automatically detect pecans, a Faster R-CNN with Feature Pyramid Network (FPN) was employed. The study tested a total of 241 pecan images, which underwent pre-processing such as exposure suppression and compensation for abnormal exposure, and illumination compensation for uneven lighting conditions, before inputting them into the network. The original VGG-16 network was replaced with ResNet-50 featuring a residual structure. The addition of FPN structure to the model addressed the issue of low accuracy in pecan detection due to similar fruit and background color and overlapping occlusion to some extent, as compared to the original Faster R-CNN model. In [3] They used machine learning and deep learning algorithms for visual symptoms to determine the detection and classification of crop leaf diseases. In 4] They used CNN Based Detection of Healthy and Unhealthy Wheat Crop. In [5] It is detailed explanation about Object Detection using Deep Learning. In [6] They made an Research on Instance Segmentation Algorithm of Greenhouse Sweet Pepper Detection Based on Improved Mask RCNN. In [7] They use instance segmentation method, mask region-based convolutional neural network (Mask RCNN), to detect cucumber fruits. They used Resnet-101 as the backbone of Mask RCNN with feature pyramid network (FPN). Cucumber Fruits Detection in Greenhouses. The RPN was improved on the original Mask RCNN by taking into of account the feature of the cucumber fruit. They proposed LG to filter the non-green background, constrain the anchor box to the green area, and adjust the size and aspect ratio of the anchor box to the shape and size of the cucumber fruit. In [8] They use faster RCNN for detecting Apple Fruits Based on Color and Shape Features. They used SLIC. With the help of that, an image is segmented to approximately 350 super-pixel blocks. The averages and variances of color components in RGB, HSV, Lab and YCbCr color space are used to make up the color vectors of super-pixel blocks. The SVM classifier with Gaussian kernel function is applied to classify color vectors. The blocks classified as fruit blocks are treated as candidate regions. Next, the shape feature described by HOG vectors are extracted by multi-scale sliding windows in the range of candidate regions and then classified by a linear SVM classifier. Finally, fruits are detected by rectangle boxes. In [9] they used Pure CNN PCNN. They used 360 dataset which contains fruits images. It can be successfully trained to classify various types of fruit images. In [10] Detection and breakdown of matured green tomatoes is implemented based on Mask R-CNN. Their model Score of bounding box and mask region for test set both reach 92.0%.

III. Existing System

They fail to encode the position and orientation of objects. They have a hard time classifying images with different positions. A lot of training data is needed for the CNN to be effective. CNNs have a habit of to be much sluggish due to maxpool like methods. In case the convolutional neural network is made up of multiple layers, the training process could take a particularly long time if the computer does not have a good GPU. Convolutional neural networks identify images as clusters of pixels arranged in different patterns. They don't understand them as components present in the image.

IV. Proposed System

Region- based convolutional neural networks are developed for applying deep models to identify and classify multiple objects from the input image. R-CNN starts by employing a selective search approach for extracting multiple region proposals where each proposal is a bounding box around the interesting boundary of an object in the image. also, each region proposal is fed into CNN to generate output features. Eventually, a set of Support Vector Machine (SVM) classifiers use these features to determine the class of each object within each region offer. Fast Region- based Convolutional Neural Network (Fast R- CNN) improves the performance of R- CNN by rooting features from the whole image rather of rooting features on each region offer. The region proposals extracted through selective search also applied to ROI pooling scheme along with the CNN output features to classify and identify ROIs in the image. For further enhancement, faster R- CNN used a (RPN) region proposal network rather than the selective search utilized in Fast R- CNN. The use of RPN decreases the generated number of proposal regions while achieving accurate and fast object discovery. Mask RCNN has the same frame as Faster R- CNN but introduced a (FCN) fully convolutional network for locating objects on the pixel position by generating a double mask for each object. Mask R-CNN improved accuracy of detecting objects by ROI alignment scheme. The proposed system exploits the effectiveness of Mask R- CNN in object localization, bracket, and segmentation at a pixel position to detect and

A. Architecture Diagram

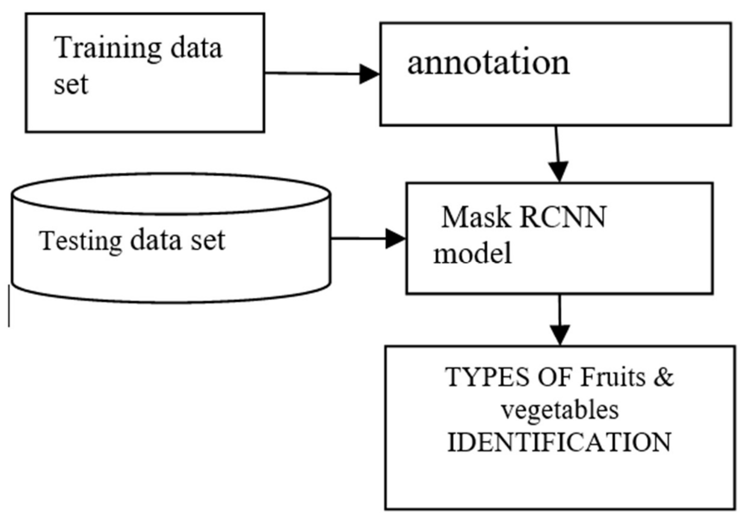

Figure 1: Architecture diagram.

- In the figure 1 the first step is acquiring the Dataset of different fruits and vegetable.
- Second step is Annotation. On an image, annotators tag the objects in the image and provide more information so that the algorithm can interpret the data and learn how to solve the problem.
- Third step is annotated image is send to Mask R-CNN. We will give testing dataset to the algorithm. It identifies which type of Fruit and Vegetable it is.

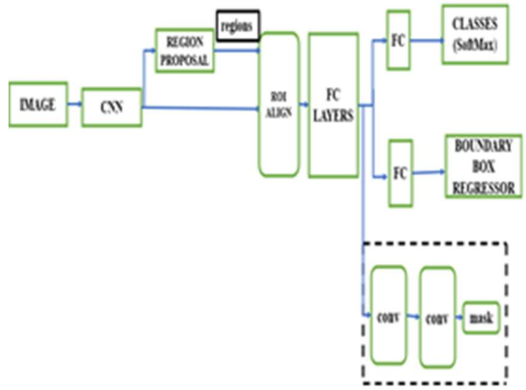

Figure 2: Mask-RCNN Architecture

Figure 2 shows the Architecture diagram of Mask R-CNN.

B. Algorithm Used

Mask Region-Based Convolution Neural Network

1. Backbone Network.
2. Region Proposal Network.
3. Mask Representation.
4. RoI Align.

C. CONVNET (Convolutional Neural Network)

A ConvNet is suitable to successfully capture the Spatial and Temporal dependences in an image through the operation of applicable filters. By reducing the number of parameters involved and the ability to reuse weights, the architecture is better suited to image datasets. In other words, the network can be trained to understand the complication of the image more.

D. Backbone Network

There are two kinds of backbone network.

1) ResNet architecture (ResNet- C4).
2) ResNet with point collective network.

The typical ResNet architecture was comparable to the Faster R-CNN architecture, but ResNet-FPN provided some disparity. Consists of multi-layer RoI generation. It is a layered pyramid network with features, which improves the accuracy of the previous ResNet architecture by inducing ROI of various sizes.

E. RPN (Region Proposal Network)

From Faster R-CNN takes the feature maps from CNN and passes them on to the Region Proposal Network. Sliding window protocol is used in RPN over these feature maps and then it creates k-anchor boxes of different forms and dimensions in each window. The first is the possibility that the anchor is an object (not taking into account which class the object belongs to). An alternate is a bounding box regressor that alters the anchors to improved fit the object. Now we have bounding boxes of different forms and dimensions that are passed to the RoI pool layer. There can now be proposals after the RPN phase where there is no assigned class.

We can take each proposal and crop it so that each proposal contains an object and this is what a RoI pooling tier ensures. It extracts static sized feature maps for respective anchor.

F. Mask Representation

A mask contains three-dimensional data about the object. Therefore, unlike the classification and bounding box regression layers, we couldn't collapse the output to a fully connected layer for improvement because it requires pixel to pixel identical on the lowest layer. Mask R-CNN uses a fully related network for mask identification. This ConvNet takes RoI as input and outputs an m * m mask representation. We also rescale this mask to output as an input image and diminish the channels to 256 using 1*1 convolution. To generate input for a fully connected network that predicts masks, RoIAlign is utilized. The main goal of RoIAlign is to transform the variable sized feature maps created by the region proposal network into a fixed size feature map. The Mask R-CNN paper proposed two variations of the architecture. In one variation, the mask generating convolutional neural network input is passed through RoIAlign after it is applied (ResNet C4). In another variant, the entry passes just before the fully connected layer (FPN).

G. ROI

For fully connected layers the input should be in fixed size. In object recognition, each application will be of a different shape. So, we use ROI Pooling for changing all the input proposals in the fixed shape.

H. Fully Connected Layers

Fully connected neural network layers apply linear transformations to input vectors using matrices. The resulting product is then subjected to a non-linear transformation through a non-linear activation function f.

V. IMPLEMENTATION

A. Dataset Preparation

The dataset frames are annotated as pixel-wise semantic segmentation where each pixel in the image is labeled with a class. These classes are different type of fruits and vegetable. For training Mask RCNN the dataset images are converted into COCO format and saved as JSON train. JSON format for object discovery and segmentation is a collection of attributes" images" contains information about the image, "categories" contains information about the classes and their unique identifier. The model substantially concentrates on (x, y) equals of each ROI. This paper introduces a simple system to convert the ground truth dataset from labeled images to JSON train format. The first step to convert the ground verity dataset from labeled images to JSON train format is to divide the dataset into three different datasets training, confirmation, and testing with a rate of 60, 20, and 20, independently. There's only one JSON train for each one of the three datasets. The coming step is to produce the JSON train for every image in the ground verity of each dataset through chancing the external boundary of each ROI. also, equals (x, y) for each boundary are listed as a vector of pixels and each match is characterized with an ID and class marker. Eventually, this data is written in. JSON train.

B. Training and Testing

In training, we will train the datasets of different fruits and vegetable images and then we will go to testing part. Training and testing are conducted using the source codes of Mask R-CNN implemented with Python 3, TensorFlow, and Keras, the proposed framework is trained online on Google Co-lab with free GPU and memory resources. The frame work produces bounding boxes containing the segmented masks for each class in the image. ResNet-101 network is used as backbone architecture for feature extraction in the proposed framework. To avoid the model overfitting and enhance the performance of the model, the transfer learning from the pretrained Mask R-CNN model on COCO dataset is used to initialize the weights of the CNN feature extractor. This paper shadows the open-source implementation of Mask R-CNN and executes parameters fine tuning to fit the pre-trained model with the particular fruits and vegetable dataset. The input frames of the network have 3 channels (i.e., R, G, and B). For training Mask R-CNN, the optimal parameters fine-tuning is determined experimentally.

C. Annotation of Image

To enable algorithms to interpret data and overcome challenges, an annotator tags objects within images to make them more informative. The labeling image annotation tool was utilized for this purpose. First step is we have to set default saved annotation in file. Then open Dir where the images are saved. There is the option RectBox to annotate the image. Annotations are saved as XML files in selected file.

1) After annotation it is given to the Mask R-CNN model

2) We will input the testing data set to the R-CNN model.

3)Mask R-CNN will predict the name of the fruits and vegetable classify crop images.

VI. Result & Discussion

In Figure 3 shows the input image of fruit Mango.

Figure 3: Input image.

In Figure 4 shows the Output image of fruits which is identified with their respective name as Mango.

Figure 4: Output image.

Confusion Matrix: A table 1 to summaries categorization model performance is called a confusion matrix. On test data for which the output label is already known, it is based.
Confusion matrix provides insight into whether a model is producing the intended results or not. Additionally, it is simple to spot model flaws

true label \ predicted label	apple	banana	carrot	cauliflower	cucumber	ginger	mango	onion	pineapple	potato
apple	19	0	0	0	0	0	1	0	0	0
banana	0	20	0	0	0	0	0	0	0	0
carrot	0	0	18	2	0	0	0	0	0	0
cauliflower	0	0	2	18	0	0	0	0	0	0
cucumber	0	0	0	0	20	0	0	0	0	0
ginger	0	0	1	0	0	19	0	0	0	0
mango	0	0	0	1	0	0	19	0	0	0
onion	0	0	0	1	0	0	0	19	0	0
pineapple	0	0	0	1	0	0	0	0	19	0
potato	0	0	0	1	0	0	0	0	0	19

Table 1: Confusion Matrix.

Loss Curve: It is mostly used to debug a neural network is a during the training. In figure 5 it gives us the training process and the track in which the network learns.

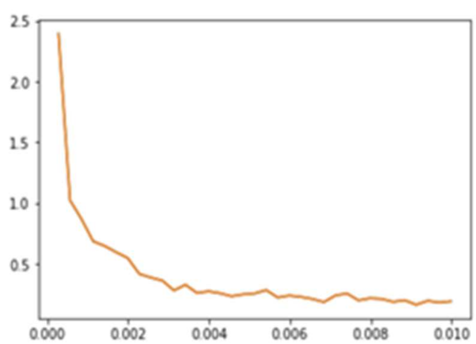

Figure 5: Loss Curve.

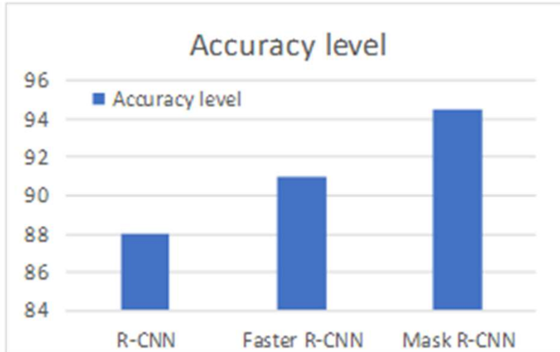

Figure 6: Accuracy.

Figure 6 shows the Comparison of RCNN, Faster RCNN, Mask R-CNN. Mask R-CNN has higher accuracy. Detection accuracy is 94.5, Mask accuracy is 81.25. In figure 7 RCNN Running speed is 6.34 and Faster R-CNN is 5.73, Compared to those two algorithm Mask R-CNN Running Speed is 4.26 which is faster.

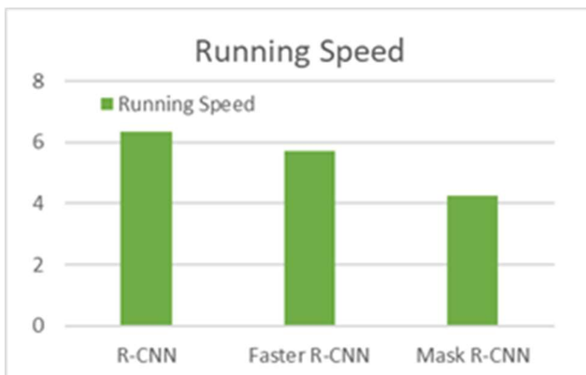

Figure 7: Running Speed.

ALGORITHM	DETEC-TION ACCUR-ACY	MASK ACCURA -CY	RUNN -ING SPEED
R-CNN	88	-	6.34
FASTER R-CNN	91	-	5.73
MASK R-CNN	94.5	81.25	4.26

Table 2: Algorithm accuracy and running speed.

VII. Conclusion

In this image identification, we can create an artificial intelligent model using python which identifies the Fruits & vegetables image with their name. This model can be able to take an image as input and display the name of fruit or vegetable given in the image as output. By using Mask region-based Convolutional Neural Network we were able to achieve a maximum accuracy.

VIII. References

[1] Wan,S.; Goudos, S. (2020)Faster R-CNN for multi-class fruit detection using a robotic vision system [ELSEVIER] Volume 168, 26 February 2020, 107036,

[2] Chunhua Hu, Zefeng Shi, Hailin Wei, Xiangdong Hu, Yuning Xie, Pingping Li(2022). Automatic detection of pecan fruits based on Faster RCNN with FPN in orchard [IJABE] Vol 15, No 6.

[3] Pallepati Vasavi, Arumugam Punitha T. Venkat Narayana Rao. Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms (IJECE) Vol. 12, No. 2, 2022.

[4] Anshuman Singh, Monika Arora. CNN Based Detection of Healthy and Unhealthy Wheat Crop, [IEEE] 2020.

[5] Anushka. Object Detection using Deep Learning, [IOP] Vol 7, 12(I), 2021, ISSN: 2277-70672021.

[6] Peichao Cong, Shanda Li, Jiachao Zhou, Kunfeng Lv and Hao Feng. Research on Instance Segmentation Algorithm of Greenhouse Sweet Pepper Detection Based on Improved Mask RCNN Agronomy 13(1):196 2023.

[7] Xiaoyang Liu; Dean Zhao; Weikuan Jia; Wei Ji; Chengzhi Ruan; Yueping Sun. Cucumber Fruits Detection in Greenhouses Based on Instance Segmentation,2019, [IEEE] Vol.7.

[8] X. Liu, D. Zhao, W. Jia, W. Ji and Y. Sun.A detection method for apple fruits based on color and shape features [IEEE] vol. 7, pp. 67923-67933, 2019.

[9] Asia Kausar, Mohsin Sharif, Jinhyuck Park Sungkyunkwan. Pure-CNN: A Framework for Fruit Images Classification [IEEE]2018.

[10] Linlu Zu,Yanping Zhao,Jiuqin Liu,Fei Su,Yan Zhang ,Liu. Detection and Segmentation of Mature Green Tomatoes Based on Mask R-CNN with Automatic Image Acquisition Approach [Sensors] 2021, 21, 7842 2021.