Computer Vision Based Weed Removal System using Object Detection based on Convolutional Neural Network

[1] **Pramod M. Kanjalkar** [2]**Jyoti B. Kulkarni** [3]**Parth A. Basole**

[1] *Associate Professor, Department of Instrumentation Engineering, Vishwakarma Institute of Technology, Pune, India.*
[2] *Assistant Professor, Department Of Computer Engineering, Sinhgad College of Engineering, Pune, India*
[3]*Student, Department of Instrumentation Engineering, Vishwakarma Institute of Technology, Pune, India*

ABSTRACT

Agriculture is one of the oldest and certainly most important professions in history of humankind. Agriculture itself is embedded deep within our foundation of country. Due to ever growing demand and growth of technology, it is possible to help farmers to boost their production via some means of current technology that is may be via robotics. Weeding and harvesting in particular area due to large land side requires repetition of monotonous task. It does not help in resulting much efficiently for weeding i.e. (using excessive herbicide and margin for human error). There is also a lack of manpower noticeable in agriculture sector. To detect and remove weed from other crops is essential task for the farmers.

Hence, the aim of the proposed system is to detect weed from the other crops in the images captured by digital camera. This system uses an object detection technique by Convolutional Neural Network. The proposed system is very helpful in agriculture area.

Keywords: *Object detection, Darknet, Convolutional Neural Network, YOLO.*

1. INTRODUCTION

Agriculture is one of the oldest and arguably most important profession in history of humankind. Agriculture itself is embedded deep within our foundation of country. Due to ever growing demand and growth of technology it’s now possible to help farmers to boost their production via means of current technology that is via robotics. Weeding and harvesting in particular due to large land side requires repetition of monotonous task. Which results in not very efficient way of weeding i.e. (using excessive herbicide and margin for human error). There is also a lack of manpower noticeable in agriculture sector.

As mentioned above in agriculture sector, there are various methods of weed removal such as manual picking, Herbicide and Cutting. All of these solutions need labours doing physically the tasks. This results in many errors or bugs such as improper removal of weed, excessive use of herbicides which have quite undesirable side effects, such as soil losing its nutrients.

1.1 Object Detection:

Image classification is a method which involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in the image. Object detection is very challenging and combines these two tasks. It draws a bounding box around each object of interest in the image and assigns them a class label. It is a technique which detects various objects of some category in digital videos and images like buildings, cars, animals, persons, etc. There are many object detection application areas e.g. image and video surveillance, retrieval, computer vision, etc. Various methods of object detection use machine learning or deep learning based approaches.

All of these problems collectively are referred as object recognition. This problem requires an object detection approach.

An object detection is a technique which detects various objects of some category in digital videos and images like buildings, cars, animals, persons, etc. There are many object detection application areas e.g. image and video surveillance, retrieval, computer vision, etc. Various methods of object detection use machine learning or deep learning based approaches.

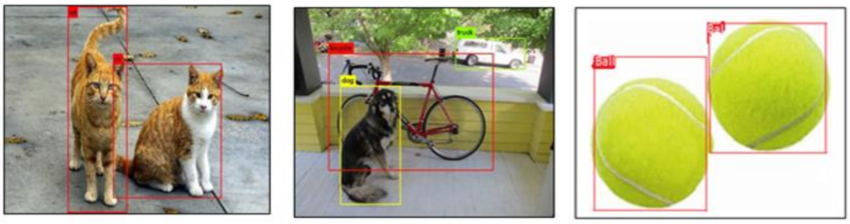


Figure 1: Objects in the images using CNN

CNN can be applied on images to analyse them. CNN has its own architectures which shares weights in every layer. It consists of more than one convolutional layers. It is used in image processing, classification, segmentation, etc. CNN is a powerful machine learning tool. It takes an image as an input, allocate some weights, calculate the biases to various image items, and differentiate within them.

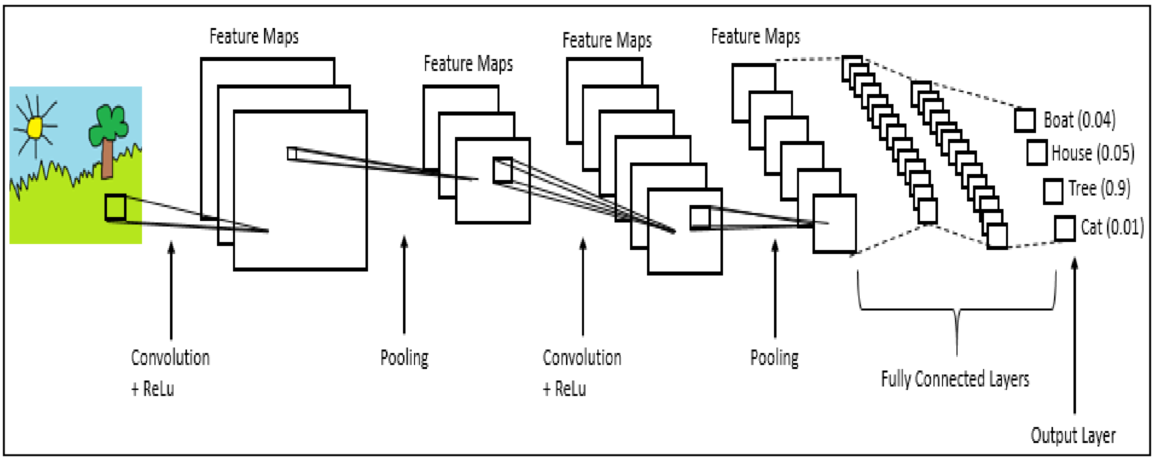


Figure 2: A Convolutional Neural Network

1) Convolution Layer: The main feature of CNN is a convolution layer. It consists of the weights of the features and the local links [1]. The objective of this layer is to recognize input function representations. The convolution layer contains several feature maps as shown in Figure 2. Every neuron of an identical feature map is used to eliminate the features of different locations in the previous phase named local features. A new feature is obtained by converting a trained kernel. Further, the result is moved to a nonlinear activation function. Likewise, the application of various kernels obtains the divergent feature maps [2].

2) The Pooling Layer: Generally, it is located in the middle of two convolutional layers. It is a second important layer used for feature extraction from feature maps. It minimizes the dimensions of the maps and increases the efficiency of feature extraction. The map size in this layer is decided by a moving phase of the kernel. Average and max pooling operations are performed upon it [2]. By storing a quantity of pooling and convolution layers, the upper level features can be extracted from of an input [3].

3) Fully Connected Layer: More than one fully connected layers can be present in CNN. Connection of each neuron to every another neuron of the immediate following layer makes the layer fully connected. The last fully connected layer is preceding to the output layer and it is not having any spatial facts in it [2]. There are many approaches for detection of objects like Regions with CNN (R-CNN), Fast Regions with CNN (Fast R-CNN), Faster Regions with CNN (Faster R-CNN), YOLO, YOLOv1, YOLOv2, YOLOv3, etc.

1.2 Types of Convolutional Neural Network

1) R-CNN: Basically, object detection using R-CNN includes three modules. The module 1 is concerned with a construction of region proposals that are not depending upon the category. The collection of candidate detections is explained by these proposals. The module 2, which is itself a huge CNN, is related with the extraction of a fixed sized feature vector from other region. The end module is formed with a stack of supervised SVMs [4].

2) Fast R-CNN: It takes an image with the objects as an input. It generates a convolutional feature map that works on the whole image with some convolutional layers. At the max, the system has two modules. First module is a fully convolutional network and this network employs various regions. Fast R-CNN detector is suggested in [5]. In between R-CNN and Fast-CNN, R-CNN provides a region wise detection proposals at the pixel level, whereas Fast R-CNN provides the regional proposals as an input at the feature map level.

3) Faster R-CNN: It performs object detection and region proposal generation tasks [6]. Object detection is faster using the design [6]. When Faster R-CNN is compared with all the previous CNN structures, then the advantage is that the computation time is minimum with Faster R-CNN [7].

4) YOLO: It is a real time system. It has its own neat architecture based on CNN. Due to this characteristic, it becomes better and faster. There are many versions of YOLO like YOLOv1,

YOLOv2 and YOLOv3. With some different approach, YOLO applies only one NN to a whole image. It splits the image in various regions and does prediction. Later, some weights are assigned by the predicted probabilities. These probabilities extract the features. YOLOv3 is more accurate and better than Fast R-CNN. Figure 3 shows an overall architecture for the proposed modifications in YOLOv3 for training.

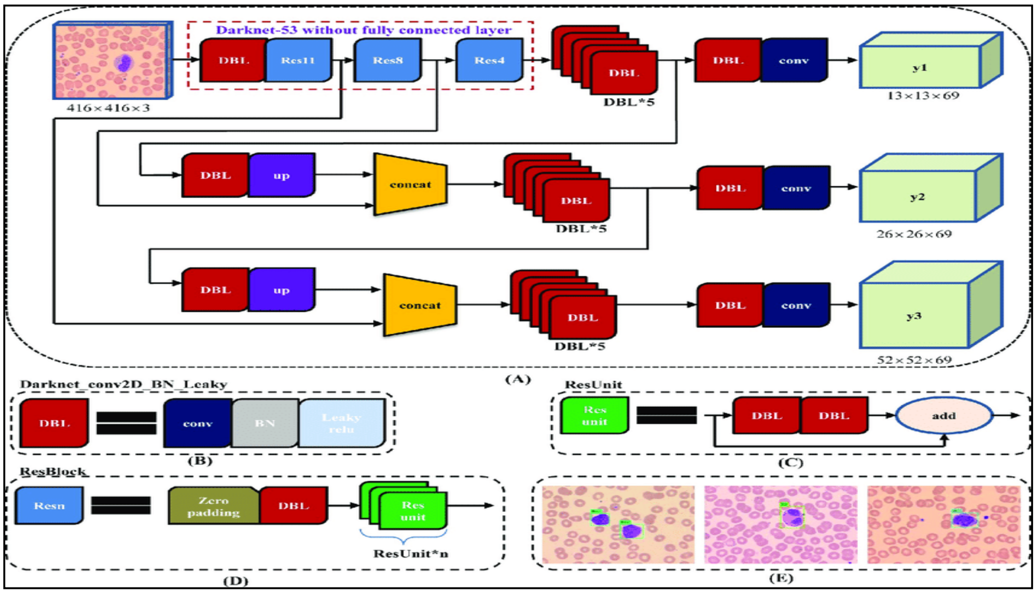


Figure 3: YOLOv3 Architecture

5) YOLOv1: For the input image, N*N grid is created. The grid is responsible for object detection [89]. The grid cell performs a bounding box prediction, scores of a confidence and a class probability for every grid modelled in terms of a tensor [14]. The accuracy level is indicated by the confidence scores. The network architecture of YOLOv1 inspired by a GoogleNet network model and image classification is achieved by it. There are of 24 convolutional layers and 2 fully connected layers. In YOLOv1 [15] and [16].

6) YOLOv2: It is a novel technique for combining a huge amount of classification data. This technique accurately detects the objects and uses these detected objects to increase its accuracy [17]. In YOLOv2, a training is performed on image dataset with a 448*448 resolution by considering 9-10 epochs roughly. The filters are adjusted in the network to work properly on the inputs. The training is performed and the trained network gives an increment of 4%. In YOLO, the fully connected layers are not used, but the prediction is done by boxes. It is operated on 448*448 input images instead of 416*416 size [17].

7) YOLOv3: YOLOv3 uses a technique of Logistic Regression for predicting maximum score to detect the object category and the bounding boxes [18]. In YOLOv3, a prediction is performed in three scales. The features extraction is carried out which is analogous to the pyramid network of features, and unsampling of the feature map is performed twice. This technique helps to obtain semantic information from previous feature maps. Convolutional extra layers are added which work on feature map and predict bounding boxes leading to object detection [94].

2. LITERATURE SURVEY

There are many other approaches for object detection from images over CNN. Using machine learning techniques object detection has been done by many researchers.

The authors proposed an innovative depth estimation method in [9] that uses BL for object detection and MSA [9]. Training dataset of six attributes is used [9] and classified the objects into four different types. A relative depth value is assigned to the object depending on the type [9].

To fit a depth map distribution from source data, a learning is performed using CAN. The conversion is made from the sparse points to dense the depth maps [11].

The authors [12] used an aperture camera that is phase-coded providing depth related colour features of an image [12]. A fully CNN is used for depth estimation [12].

The authors, in paper [10], proposed a new quality criteria set by focusing on particular characteristics of depth maps [10].

The authors [19] investigated a detections of cars by training and testing on dataset of cars. The authors analysed that YOLOv3 object detection technique proved better than Faster R-CNN in processing time and sensitivity.

A deliberation about the combination of NN and Wavelet Transform [13] is performed. A mixture of two techniques, i.e. DWT and NN on IRS-1D images, used and improved results are obtained.

The above survey elaborates on many object detection techniques using machine learning and NN approach.

3. PROPOSED WORK

This dataset contains 1300 images of sesame crops and different types of weeds with each image label.

Each image is a colour image of 512 × 512 size. Labels for images are in YOLOv3 format. This dataset contains two classes of images; one as crop and another as weed which are written in YOLOv3 format i.e. having its class, name, centre positions and the bounding box width and length.

Darknet is an open source neural network framework written in C and CUDA. It is fast, easy to install, and supports CPU and GPU computation.

The deep learning framework is described in C programming language. Once the network gets trained, then there is no need of Darknet for the inference. OpenCV has a built-in support for the formats of Darknet, hence the model and the weights which are trained are directly used where OpenCV is used, also from Python.

The important part of this network is that there is some documentation to train the own data set and how to execute the inference on the own input. Other popular frameworks are sometimes much optimized for training and validation against various existing data sets that it is difficult to break out of this golden enclosure and build a usable product. By building the source code in python, the YOLO model can be trained.

This results in a YOLO weight file which would be used to test it on our Test data. This weight file is then used to feed forward the YOLO network with taking input as the images in test file and it gives a matrix as we discussed before of type [x, y, c, w, h, p]. This matrix is later used to draw bounding box around the pictures.

Further, this trained model is loaded into another jupyter notebook which would use trained weights and by using opencv library would forward propagate the neural network to get the matrix mentioned before then by using rectangle function of opencv bounding box would be drawn on the selected image.

3.1 Algorithm of YOLOv3 CNN Training for Object Detection

Step 1. Read colour image of size 416 × 416 ×3.
Step 2. Number of Layers are from 0 to 106
Step 3. As size increases, feature map size also increases.
Step 4. Use of YOLOv3 with feature map of size 13×13×69, 26×26×69 and 52×52×69 as outputs.
Step 5. Perform Up-sampling at layers 85, 97.
Step 6. Route at layer 85, 61, 91, 97 and 36
Step 7. Filters or strides used of size 32, 64, 128, 256, 512 and 1024.

3.2 Algorithm of Proposed System

Step 1. Read colour image of size 512×512
Step 2. Apply YOLOv3 algorithm for image object Detection.
Step 3. Detection of weed from other objects is done
Step 4. The results are analysed based on sesame crops as a dataset.
Step 5. End

4. RESULTS AND DISCUSSION

Below are the results of the image which was fed into the forward prop and the result which we got in the end.

Table 1: Some Object Detection accuracies using YOLOv3 approach

Sr. No.	Image Size	Object Detection time (seconds)	Accuracy (%)
1	3966x2420	32	97
2	769x577	33	96
3	774x513	32	98
4	769x761	34	97
5	901x889	33	96
6	336x287	34	96

Table 5.2: Comparison of Object Detection accuracies using CNN Approaches [1]

Sr. No.	**Object Detection techniques**	**Detection accuracy (%)**
1	Fast RCNN [1]	70.0
2	YOLO [1]	63.4
3	SSD 500 [1]	76.8
4	YOLOv2 [1]	78.6
5	CNN using YOLOv3	96

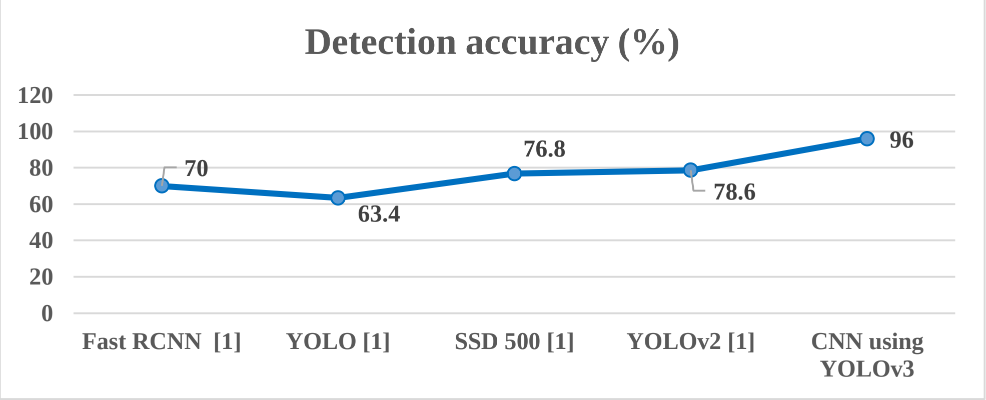
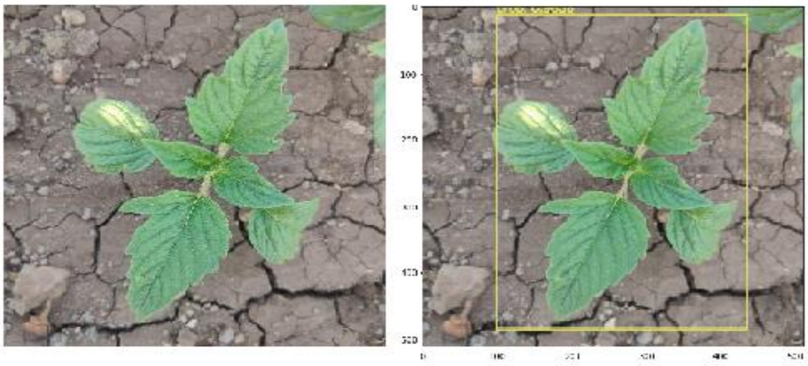


Figure 4: Object Detection Accuracy



(a) (b)

Figure 5: a) Input image 1; b) output image 1

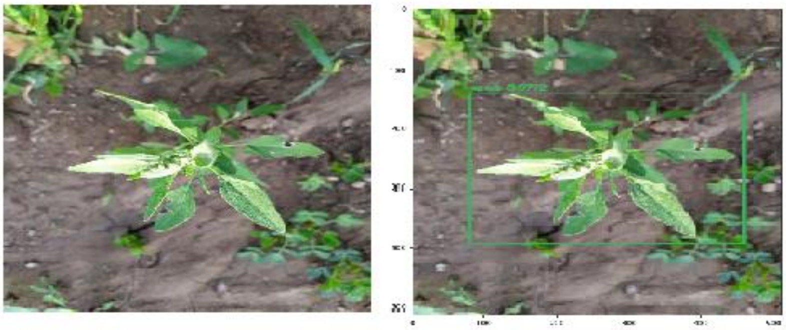
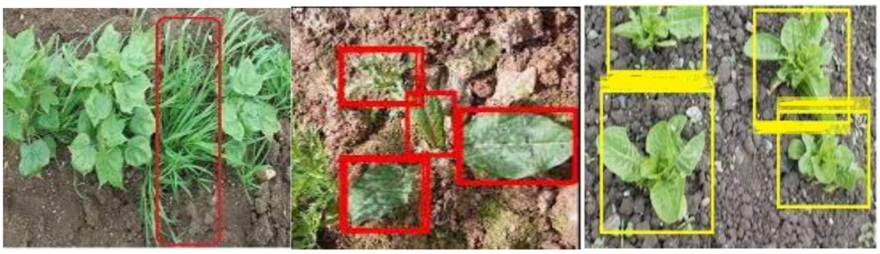


Figure 6: a) Input image 2; b) output image 2



(a) (b) (c)

Figure 7: Weed detection using YOLOv3

5. CONCLUSION

The proposed work presented here works on various crop images. The image training is performed to detect various objects in the images. YOLOv3 framework is used for the training and object detection. The experimental results in Table 2 show that the training accuracy using YOLOv3 increased as compared to other methods. Images in Figure 5, 6 and 7 show that the weed from the images get separated from normal crop.

YOLOv3 model in particular helps with speedy detection, which could be employed on embedded boards such as BeagleBone AI and JetsonNano. In Future, a robust system which could use one shot learning to identify certain crops from a single image, can be employed to make the whole solution more modular and less jarring.

REFERENCES

[1] K. G. Shreyas Dixit, Mahima G. Chadaga, Sinchana S. Savalgimath, G. Ragavendra Rakshith, Naveen Kumar M. R., (2019). Evaluation and Evolution of Object Detection Techniques YOLO and R-CNN, International Journal of Recent Technology and Engineering, (IJRTE), ISSN No.: 2277-3878, Vol. No.8, IssueNo.-2S3, 824-829.

[2] Shanshan Guo, Shiyu Chen and Yanjie Li, (2016). Face Recognition Based On Convolutional Neural Network and Support Vector Machine, International Conference on Information and Automation Ningbo, China, Proceeding of IEEE, 1787-1792.

[3] Tianmei Guo, Jiwen Dong, Henjian Li Yunxing Gao, (2017). Simple Convolutional Neural Network on Image Classification, IEEE second International Conference on Big Data Analysis, 978-1-5090-36196, 721-724.

[4] Hyeok-June Jeong, Kyeong-Sik Park, Young-Guk Ha, (2018). Image Preprocessing for Efficient Training of YOLO Deep Learning Networks, 2018 IEEE International Conference on Big Data and Smart Computing, 635-637.
[5] Jonathan Hui, (2018). Object Detection: speed and accuracy comparison (Faster R-CNN, R-FCN, SSD, FPN, RetinaNet, medium.com.
[6] H. Khalajzadeh, M. Mansouri, and M. Teshnehlab, (2014). Face Recognition Using Convolutional Neural Network and Simple Logistic Classifier, Soft Computing in Industrial Applications, Advances in Intelligent Systems and Computing, Springer International Publishing Switzerland, 197–207.
[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, (2016). You Only Look Once: Unified, Real-Time Object Detection, Computer Vision and Pattern Recognition, 1–10.
[8] Rohith Gandhi, (2018). R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms, https://towardsdatascience.com.
[9] Jae I Jung, Yo Sung Ho, (2010). Depth Map Estimation from Single View Image using Object Classification based on Bayesian Learning, IEEE Explore, https://doi.org/10.1109/3DTV.2010.5506603.
[10] Tobias Koch, Lukas Liebel, Friedrich Fraundorfer, Marco Körner, (2019). Evaluation of CNN-based Single-Image Depth Estimation Methods, Computer Vision, ECCV 2018 Workshops. Lecture Notes in Computer Science, Vol. No. 11131, Springer.
[11] Yaoxin Li, Keyuan Qian, Tao Huang and Jing kun Zhou, (2018). Depth Estimation from Monocular Image and Coarse Depth Points based on Conditional GAN, MATEC Web of Conferences, IFCAE-IOT 2018, Vol. 175, 1-5.
[12] Harel Haim, Shay Elmalem, Raja Giryes, (2018). Depth Estimation from a Single Image using Deep Learned Phase Coded Mask, IEEE Transactions on Computational Imaging, 298-310.
[13] C. M. Sheela Rani, V. Vijaya Kumar, B Sujatha, (2012). An Efficient Block based Feature Level Image Fusion Technique using Wavelet Transform and Neural Network, International Journal of Computer Applications (0975 – 8887), Vol. 52, No.12, 13-19.
[14] D. Foley and R. O'reilly, (2018), An Evaluation of Convolutional Neural Network Models for Object Detection in Images on Low-End Devices, AICS 2018, Vol. 2259, 1–12.
[15] Dhruv Parthasarathy, (2017). A Brief History of CNNs in Image Segmentation: From R-CNN to Mask R-CNN, https://blog.athelas.com.
[16] Ankit Sachan, (2018). Zero to Hero: Guide to Object Detection using Deep Learning: Faster R-CNN, YOLO, SSD, CV-Tricks.com.
[17] Sik Ho Tsang, (2018). Review: R-CNN (Object Detection), Coinmonks-Medium.
[18] Sik Ho Tsang, (2018). Review: Faster R-CNN (Object Detection), Towards Data Science.