CROP AND WEED DETECTION FROM IMAGES USING YOLOv5 FAMILY

Katakam Koushik[1] , S. Venkata Suryanarayana[2]

[1]PG Scholar, CVR College of Engineering/IT Department, Hyderabad, India

Email: 20B81DB004@cvr.ac.in

[2]Professor, CVR College of Engineering/IT Department, Hyderabad, India

Email: suryahcu@gmail.com

Abstract

Weed is the main field element in agriculture that has an impact on crop quality and productivity. So it's crucial to find and categorize weeds in the field when they're still in the early stages of development. Farmers often use cultural, biological, and mechanical approaches to prevent weed development in their fields. Later, as technology developed, farmers started use chemical substances like herbicides and insecticides to control pests and weeds in their fields. Farmers also sprinkle herbicides on the crops after evenly spraying them throughout a field. Crop growth, crop quality, and crop output are all impacted by the herbicides' chemical composition. Therefore, it is crucial to find weed in the field when it is still in the early stages of development. Herbicides must be sprayed selectively on weeds in order to prevent damage to crops from the herbicides' chemical components. This allows for site-specific weed control. We are proposing YOLOv5 model to detect Crop and Weed from the images. In this paper we compared the performance of versions with the various existing deep learning-based object detection methods like YOLOv3, YOLOv3-tiny, YOLOv3-spp with three different parameters named map 0.5, map 0.5:0.95 and data set used. This information will be helpful for practitioners to select the best technique for the Crop and Weed dataset.

Keywords: weed detection system, YOLOv5s, YOLOv5l, YOLOv5l6,YOLOv5x6.

INTRODUCTION

Wild vegetation known as weed develops in fields in between crops, competing with the crop for space and nutrients. Weed consumes crop-supplied water, sunlight, minerals, and nutrients, and it also takes up area designated for crop output; as a result, the field's crop development is hampered. Farmers suffer significant losses due to weed since it lowers crop growth, quality, yield, and land value. Farmers spray herbicides consistently across the field to keep the weeds under control, but this is expensive for them to do only to keep the weeds under control. A weed detection system [1] that particularly identifies and categorises weeds in the field is created using information technology as part of a precision agriculture strategy called "site-specific weed management." This method allows herbicides to be sprayed just on weeds, which reduces the amount of herbicides required in agriculture.

Deep learning technology may be utilized to create a successful weed detection system because it is already being used for facial recognition, natural language processing, object identification, picture categorization, sentiment analysis, and automated self-driving cars. Convolution neural networks (CNN), a deep learning technique frequently employed in object recognition and picture categorization, can be applied to construct a weed identification system. One or more robots, drones, computer vision systems or image recognition systems might be used to accomplish the weed detecting system. For autonomous self-driving cars, object identification, sentiment analysis and picture classification, deep learning technology is already in use. It might be implemented to develop a successful weed detection system.

In the field where a wide range of crops are grown, different forms of weed development can be noticed depending on the soil quality. It's crucial to first identify the types of weeds and crops present in the land. The data required to train the model must then be extracted using this information.

LITERATURE SURVEY

In [2] the authors developed a Faster R-CNN detector and the fully convolutional model that trains the detector on shared characteristics from several layers of a deep neural network. A Region Proposal Network (RPN) [3] accepts any size picture and generates a set of rectangle bounding boxes. Faster R-CNN employs anchor boxes instead of picture pyramids. A scale and aspect ratio reference box is known as an anchor box. When numerous anchor boxes are used, there are many different scales and aspect ratios for the same place. Each zone is then transferred to each reference anchor box, enabling for the detection of objects of varying sizes and aspect ratios. The ROI [4] Pooling layer extracts each proposed region in the image.

In [5] the authors developed R-FCN which is based on region of interest. The Region-of-Interest (RoI) pooling layer divides a popular family of deep networks used for object recognition into two sub networks. They utilized classification architectures like as AlexNet [6] and VGG Nets [7] to divide the sub network. Between two sets of convolutional layers, a deeper RoI-wise sub network is placed to address this issue. Then RoI's are classified into object categories and background using the R-FCN architecture. Convolutional weight layers are used for all learnable weight layers and they are computed on the complete picture. SPPnet and Fast R-CNN are "semi-convolutional" that means one sub network computes the full picture while another analyses selected areas. Sliding multi-scale windows on shared feature maps are used by OverFeat to recognize objects. similarly, sliding windows that replace region suggestions are examined. For obtaining holistic item identification results on a whole picture, another family of object detectors uses fully-connected (fc) layers. They achieved the mAP of the class-specific RPN as 67.6 percent, which is around 9 points lower than the 76.4 percent of the regular Faster R-CNN.

Jafari.et.al, [8] evaluated two methods based on farm features investigated with different stratagies to create a weed identification system. The photos were shot when the weeds had four leaves, and they had a resolution of 960 × 1280 pixels to capture the farm characteristics of the plant. The photos were recorded in RGB format, and there are 60 sets of 120 images in the collection. The photos are first analysed to identify any green content, and then transformed to grayscale images so that features may be extracted. For the discrimination procedure, texture characteristics were added in ANN [9], and PCA [10] was utilized to condense the dimensionality of the input data. To categorize plant types, the support vector machine is utilized, and its accuracy is assessed using R^2 and RMSE values. They constructed the confusion matrix in order to assess how well the ANN performed. Both ANN [9] and SVM [11] achieved accuracy of 86% and 88%, respectively.

Adnan Faroq.et.al, [12] analyzed weed detection systems for classifying and detecting weeds based on spectral bands and spatial resolution which are essential in agriculture if area-specific weed management is to be achieved. The author's primary goal is to decrease the use of herbicides in fields because they are bad for both crops and people's health. and contrasting, The effectiveness of weed detection is measured by contrasting CNN and HOG. For the purpose of classifying weeds, the spectral bands and spatial resolution of the hyperspectral pictures were examined. Convolution neural networks, a type of deep learning technique, can automatically extract features and effectively learn high-level characteristics from hyperspectral pictures. Using a Brimrose VA210 filter and a JAI BM-141 camera, the collection's photos were captured.. 200 picture patches of each type of weed were created using a data augmentation method during the preprocessing stage, and the spatial resolution of the hyperspectral image was assessed. They used two criterias to analyse the weed categorization: one based on batch size and band number sensitivity, the other based on patch size and resolution sensitivity.

Junfeng Gao.et.al, [13] developed machine-learning technique to classify various types weed and crop . Hyperspectral and NIR cameras are used to take the pictures in the institute for agriculture and fisheries research's plant laboratory. Arvensis, mays, and rumex are the three weeds listed in this study.. The raw picture is divided into five 5x5 single-band subimages as part of the pre-processing, and each of these images is then further cropped to determine the region of interest for the reflectance calibration. They categorised Weeds and crops using the random forest method. When estimating the significance of permuted characteristics, the random forest technique offers useful information. They were able to classify three weeds with accuracy of 0.785, 0.663, and 0.713, respectively, whereas maize was classified with 94% accuracy. Since the classifier is unable to distinguish between the three separate types of weeds, the findings show that crops are better identified than weeds. The major drawback of this model is, it will work only when the data distribution was uncertain and untested.

In [14] the author revealed that weed detection using a convolution neural network can be done well without data pre-processing.. As a result of experimental hardware limitations, they considered 5000 pictures with 45600 labelled patches randomly selected for training and validation. During the pre-processing step, they changed the pictures' RGB to HSV scalar coordinate systems to normalise them. They opt to categorise the data using the CNN method, which comprises of a final SoftMax layer, three convolutional layers, three fully connected layers, and three additional layers. The model produced two models after 40 hours of training, one of which was trained using the original photographs and the other using color-normalized picture data. Finally, the validation score only varies by 0.2%, having no effect on the model classification result, and both models have the same accuracy. In future investigations, they suggested using GPU processors, which have a larger computing capability than CPU systems, to complete the work with more patches.

3 METHODOLOGY

The primary factor in the agricultural land that affects the yield and quality of a crop is weed. Therefore, it's essential to identify weeds in the field when they are still in the incipient phases of development. We use YOLOv5 model to detect weed in agriculture lands because YOLOv5 model has a focus layer witch cable to detect low level features accurately. The Workflow of YOLOv5 for the Crop and Weed dataset is as shown in Figure 1.

<u>YOLOv5</u>

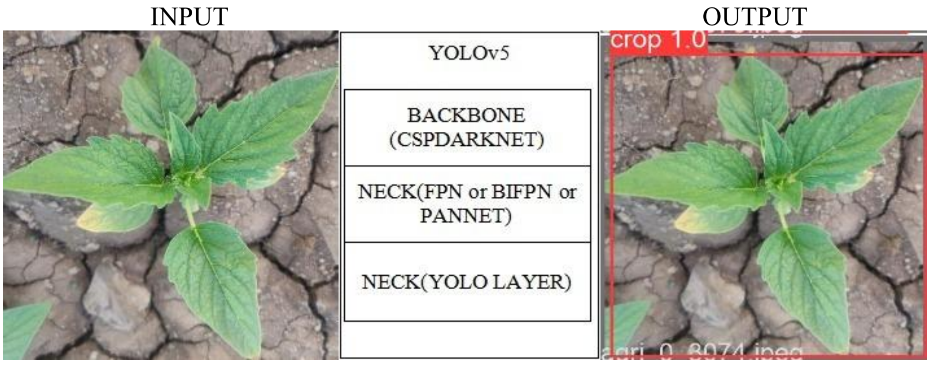

Figure 1 : Workflow of YOLOv5

YOLOv5 which uses the MS COCO [15] with AP 0.5:0.95 and AP50 is a cutting-edge detector that is both more accurate and faster (FPS) than any other detector in the market. YOLOv5 use feature pyramid network (FPN) and path aggregation network (PAN) identification of the same object in various sizes and scales. In FPN top-down augmentation path is used and in PAN bottom-up data augmentation path is used. YOLOv5 uses Focus structure along with CSPdarknet53 as a backbone. The major goal of the Focus layer is to minimize layers, parameters, FLOPS, CUDA memory and boost forward and backward performance with the least amount of mAP effect.

The MS COCO dataset demonstrates that various object detection techniques make use of anchor boxes. In reality, older versions of YOLO, such YOLOv2, only k-Means clustering was used for it.

Whereas, YOLOv5 uses an alternative approch with genetic algorithm to create anchor boxes. Autoanchor is the term for this process, which recomputes the anchor boxes to suit the data if the default ones are insufficient. The Architecture of YOLOv5 is as shown in Figure 2.

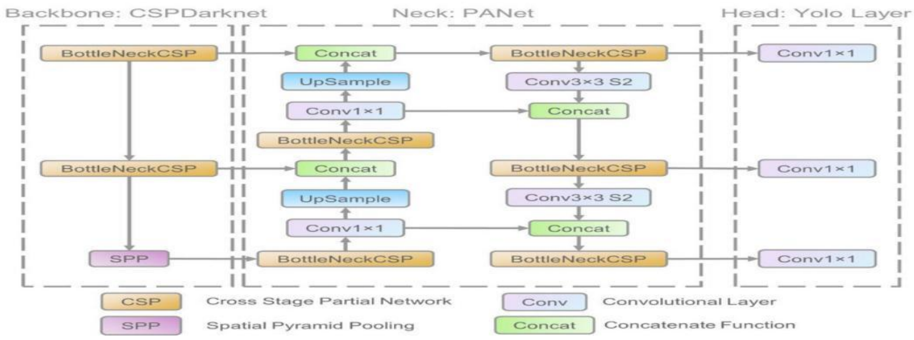

Figure 2 : YOLOv5 Architecture

YOLOv5 architecture

1. Backbone:

Model Backbone's primary function is to draw out crucial details from an input image. The CSP (Cross Stage Partial Networks) architecture is used in YOLO v5 to extract highly advantageous characteristics from an input picture.

2Neck:

Using the Model Neck, features are frequently arranged into pyramids. Feature pyramids improve models' ability to generalize to the target context when it comes to object scaling. It facilitates the recognition of the same entity in various scales and sizes.

Feature pyramids enable models to perform well on undiscovered data. Other models use other feature pyramid techniques, such as FPN, BiFPN and PANet. PANet is used in YOLOv5 as a neck to get feature pyramids.

3Head:

The model Head is mostly in charge of the final detecting step. With the use of anchor boxes, bounding boxes, objectness scores, and class probabilities, it creates final output vectors and the different versions of YOLOv5 based on p5 and p6 model is as shown in Table1.

Table 1 : YOLOv5 versions

S.No	YOLOv5p5 Models	YOLOv5p6 Models
1	YOLOv5n	YOLOv5n6
2	YOLOv5s	YOLOv5s6
3	YOLOv5m	YOLOv5m6
4	YOLOv5l	YOLOv5l6
5	YOLOv5x	YOLOv5x6

RESULTS

Experimental Setup

To Implement YOLOv5 model we require the following Hardware and Software support.

Hardware requirements: Processor: i3 or better,RAM: 8GB or More,Storage: 120GB or More,Nvidia GPU Recommended are used.

Software requirements:OS: Windows,Python 3.7 or later,Pytorch,OpenCV Other necessary Python Modules are used.

Precision, Recall, F1 score, map 0.5, map 0.5:0.95 are used to evaluate the performance of the model.

DATASET:(Crop and Weed)

The Crop and Weed dataset contain images from agriculture lands collected using UAV's, the dataset collected https://www.kaggle.com/datasets/ravirajsinh45/crop-and-weed-detection-data-with-bounding-boxes website.

The Crop and Weed detection dataset contains 701 images across two classes namely Crop and Weed. The whole dataset is divided into train/test sets by ratio 70% and 30% within each event class so 501images used for training and 200 images used for testing.

Sample output of crop and weed for YOLOv5l6 shown in Figure 3 with probability values.

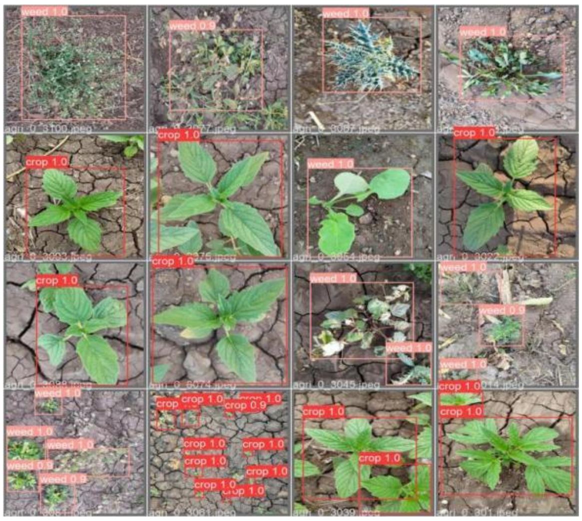

Figure 3 : sample output crop and weed for YOLOv5l6

The performance of YOLOv3 and YOLOv5 family for the given image dataset is listed in the Table 2, Table 3 and Table 4. We also compared the accuracy of all models in the form of bar chart as shown in Figure 4.

Table 2 : Precision,Recall,F1 score,map 0.5 and map 0.5 to 0.95 values for Crop

YOLOv3 and YOLOv5 models	Crop				
	map 0.5	**map 0.5:0.95**	**Precision**	**Recall**	**F1 score**
YOLOv3-tiny	94.2	63.3	0.86	0.956	0.9054
YOLOv3	99.5	93.2	0.997	0.997	0.997
YOLOv3-spp	99.5	95.8	0.997	1	0.99849
Yolov5n	96.1	67.2	0.905	0.928	0.9163
Yolov5s	96.7	68.1	0.924	0.952	0.9378
Yolov5m	99.3	89.1	0.98	0.981	0.9804
Yolov5l	99.5	94.7	0.994	0.999	0.99649
Yolov5x	99.5	95.3	0.991	0.995	0.99299
Yolov5n6	99.2	89.8	0.982	0.991	0.98647
Yolov5s6	99.5	94.7	0.994	0.999	0.99649
Yolov5m6	99.5	98.3	0.996	1	0.99799
Yolov5l6	99.5	99.2	0.999	1	0.99949
Yolov5x6	99.5	99.4	0.999	1	0.99499

Table 3 : Precision,Recall,F1 score,map 0.5 and map 0.5 to 0.95 values for Weed

YOLOv3 and YOLOv5 models	Weed				
	map 0.5	**map 0.5:0.95**	**Precision**	**Recall**	**F1 score**
YOLOv3-tiny	93.5	62.7	0.821	0.952	0.8816
YOLOv3	99.5	93.9	0.995	0.998	0.9964
YOLOv3-spp	99.5	96.2	0.996	0.999	0.9974
Yolov5n	96.3	69.6	0.889	0.938	0.9128
Yolov5s	96.5	69.8	0.911	0.944	0.9272
Yolov5m	99.4	91.4	0.982	0.987	0.9844
Yolov5l	99.5	95.4	0.995	0.996	0.9954
Yolov5x	99.5	95.9	0.998	0.995	0.9964
Yolov5n6	99.4	90.9	0.987	0.992	0.9894
Yolov5s6	99.5	95.6	0.996	0.996	0.996
Yolov5m6	99.5	98	0.998	0.999	0.9984

Yolov5l6	99.5	99	1	0.999	0.9994
Yolov5x6	99.5	98.6	1	0.999	0.9994

Table 4 : Precision,Recall,F1 score,map 0.5 and map 0.5 to 0.95 values for All(crop and weed)

YOLOv3 and YOLOv5 models	**All**				
	map 0.5	**map 0.5:0.95**	**Precision**	**Recall**	**F1 score**
YOLOv3-tiny	93.5	62.7	0.821	0.952	0.8816
YOLOv3	99.5	93.9	0.995	0.998	0.99649
YOLOv3-spp	99.5	96.2	0.996	0.999	0.99749
Yolov5n	96.3	69.6	0.889	0.938	0.91284
Yolov5s	96.5	69.8	0.911	0.944	0.9272
Yolov5m	99.4	91.4	0.982	0.987	0.98449
Yolov5l	99.5	95.4	0.995	0.996	0.99549
Yolov5x	99.5	95.9	0.998	0.995	0.99649
Yolov5n6	99.4	90.9	0.987	0.992	0.9894
Yolov5s6	99.5	95.6	0.996	0.996	0.996
Yolov5m6	99.5	98	0.998	0.999	0.99849
Yolov5l6	99.5	99	1	0.999	0.99949
Yolov5x6	99.5	98.6	1	0.999	0.99949

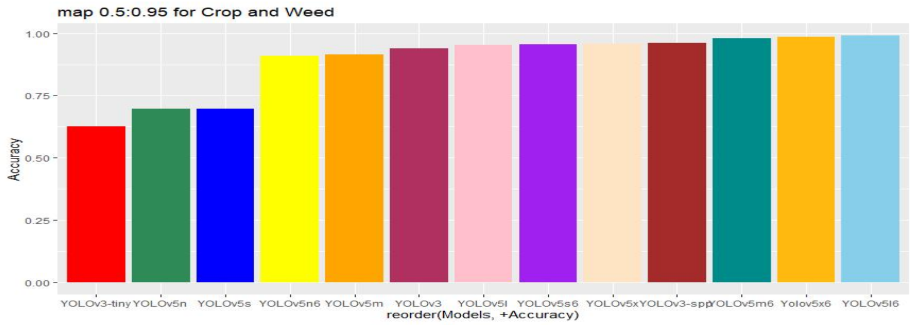

Figure 4 : Accuracy comparisons of all models for All(Crop and Weed) category

Conclusion

In most computer and robot vision systems, object detection is crucial. Despite recent improvements and the implementation of some current approaches into various consumer devices or assistance driving systems, we are still a long way from human-level performance, especially in open-world learning. Object detection isn't extensively employed in a variety of scenarios when it may be quite useful. Object detection systems are becoming increasingly vital as mobile robots and autonomous gadgets become more widely used (e.g: quadcopters, drones). Accuracy (map 0.5), Accuracy (map 0..5 to 0.95), Precision, Recall and F1 score for YOLOv5l6 model are 99.5, 99, 1, 0.999, 0.99949 respectively. Accuracy (map 0.5), Accuracy (map 0..5 to 0.95), Precision, Recall and F1 score for YOLOv5x6 model are 99.5, 98.6, 1, 0.999, 0.99949 respectively. For crop and weed classes Accuracy (map 0.5), Accuracy (map 0..5 to 0.95), Precision, Recall and F1 score results show that YOLOv5l6 achieved the best performance.

Summary

On Comparing the YOLOv3 family and YOLOv5 family on the Crop and Weed dataset, YOLOv5l6 can able to predict Crop and Weed accurately than the other models.

Future Enhancement

YOLOv7 [16] is the latest version of the YOLO series network and was developed in August 2022. YOLOv7 improves speed and accuracy by introducing several architectural reforms like E-ELAN (Extended Efficient Layer Aggregation Network), Model Scaling for Concatenation based Models and several Trainable BoF(Bag of Freebies) like Planned re-parameterized convolution, Coarse for auxiliary and Fine for lead loss. Expand, shuffle, and merge cardinality is a technique used by E-ELAN to constantly improve the network's capacity for learning while preserving the original gradient route.

References

[1] S.V. Suryanarayana A Survey on Weed Detection System Using Deep Learning Turkish Online Journal of Qualitative Inquiry (TOJQI) Volume 6, July 2021: 6147- 6151

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

[3] Vu, T., Kim, K., Kang, H., Nguyen, X. T., Luu, T. M., & Yoo, C. D. (2021). Sphererpn: Learning spheres for high-quality region proposals on 3d point clouds object detection. 2021 IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/icip42928.2021.9506249

[4] Rossi, L., Karimi, A., & Prati, A. (2021). A novel region of interest extraction layer for instance segmentation. 2020 25th International Conference on Pattern Recognition (ICPR). https://doi.org/10.1109/icpr48806.2021.9412258

[5] Rasmussen, C. B., Nasrollahi, K., & Moeslund, T. B. (2017). R-FCN object detection ensemble based on object resolution and image quality. Proceedings of the 9th International Joint Conference on Computational Intelligence. https://doi.org/10.5220/0006511301100120

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90. https://doi.org/10.1145/3065386

[7] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep Convolutional networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2016.182

[8] Bakhshipour and A. Jafari, "Evaluation of support vector machine and artificial neural networks in weed," Computers and Electronics in Agriculture, vol. 145, pp. 153-160, 2018.

[9] Alexandre de Brebisso,Etienne Simon,Alex Auvolat,Pascal Vincent,Yoshua Bengio Artificial Neural Networks Applied to Taxi Destination Prediction arXiv:1508.00021v2 [cs.LG] 21 Sep 2015

[10] Mu, n., & qiao, d. (2019). image classification based on convolutional neural network and support vector machine. 2019 6th international conference on information, cybernetics, and computational social systems (iccss). https://doi.org/10.1109/iccss48103.2019.9115443

[11] Etika KartikadarmaL, Sari Wijayanti, Sari Ayu Wulandari, Fauzi Adi Rafrastara Principle Component Analysis for Classification of the Quality of Aromatic Rice International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 8, Augus 2017

[12] Farooq, "Analysis of spectral bands and spatial resolutions for weed classification via deep convolutional neural network," IEEE GEOSCIENCE AND REMOTE SENSING , pp. 183-187, 2019.

[13] J. Gao, D. Nuyttens, P. Lootens and Y. He, "Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared mosaic hyperspectral imagery," Biosystems Engineering, vol. 170, pp. 39-50, 2018.

[14] J. Wu, "Weed detection based on Convolutional Neural network," IT and Cognition, pp. 1-13, 2018.

[15] Ning, Z., Wu, X., Yang, J., & Yang, Y. (2021). MT-yolov5: Mobile terminal table detection model based on YOLOv5. Journal of Physics: Conference Series, 1978(1), 012010. https://doi.org/10.1088/1742-6596/1978/1/012010.

[16] Chien-Yao Wang , Alexey Bochkovskiy, and Hong-Yuan Mark Liao Institute of Information Science, Academia Sinica, Taiwan 6 Jul 2022 arXiv:2207.02696v1.