Enhancing Precision Agriculture with Machine Learning and Image Processing: A Comparative Evaluation of YOLO and RCNN for Weed Identification and Detection

Satya Kiranmai Tadepalli, Shobarani Salvadi, Akhil Kambhatla, and Kadiyala Ramana

Abstract Agriculture is one of the origins of mortal food in this world. Currently due to the growing population, we need the lesser productive capability of the husbandry to meet the demands. In olden days, humans employed natural styles as a kind of industrial poison, just like they did with cow manure in fields. This resulted in a rise in output sufficient to suit the needs of the populace. However, subsequently, people were permitted to generate more profits through further development. A revolt known as the "Green Revolution" therefore began. After this period operation of deadly venoms as dressings and fungicides have increased to a drastic position. By doing so we were successful in adding productivity, but we've forgotten the damage done to the terrain, which will raise a mistrustfulness in our food on this beautiful earth. So, in this design, we're going to apply a model using image processing and deep literacy that will be suitable to separate whether a crop is weed or not and lets us to spot the dressings on only weed crops and not prompt the surroundings where good crops may be present, by this we can enhance crop yield as we're scattering the dressings on only weed crops and not on other areas of the field.

Keywords Deep learning · Image processing · YOLO3 · RCNN

1 Introduction

One of the most severe obstacles to crop productivity is weeds, In order to increase crop output and supply more food for a growing global population, weed control is essential. Nonetheless, weed control could have a harmful impact on the ecosystem.

S. K. Tadepalli · S. Salvadi · A. Kambhatla · K. Ramana (✉)
Department of Artificial Intelligence and Data Science, Chaitanya Bharathi Institute of Technology, Hyderabad, India
e-mail: ramana.it01@gmail.com

V. K. Gunjan et al. (eds.), *Cybernetics, Human Cognition, and Machine Learning in Communicative Applications*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-97-8533-9_22

The use of herbicides has the potential to pollute the environment since, in the majority of situations, only a very small amount of the sprayed chemicals really reach their intended targets, while the majority of herbicides fall to the ground, and some may even drift away. Mechanical weed removal may cause erosion and harm beneficial creatures including soil-dwelling earthworms and soil-surface spiders. Some weed management techniques have additional drawbacks and frequently have a severe impact on the environment. Sustainable weed management techniques must be created to solely effect weed plants and cause the least amount of disruption to the environment. Before using any control methods, weeds should be discovered and recognized in real-time, which would improve and increase the sustainability of weed management. In order to quickly identify weeds in crop fields, convolutional neural networks are used to develop a deep learning model to detect weeds with high accuracy.

2 Literature Review

Many authors have claimed development credit for the creation of weed detectors. The list of recent innovations that have been published in papers is as follows:

In paper [1] the author describes "Deep Neural Networks to detect weeds from crops in Agricultural Environments in Real-time". The author began by going over many methods. The neural network was trained using backpropagation, the hyper-parameters were adjusted using the validation set, and the final model was evaluated using the testing set. After pre-processing the photographs, they used a pre-trained convolutional neural network named VGG16 to extract features from the pictures. A new fully connected layer was then trained to execute the classification job on top of the VGG16 network. Finally, they implemented the trained neural network on a Raspberry Pi computer equipped with a camera module to detect weeds in the field in real-time. This technology's primary flaws are its expensive price and limited dataset.

In paper [2] "Algorithm of Weed Detection in Crops by Computational Vision" is described by the author. The authors provided a computer vision-based approach for weed detection in crops. To capture pictures of the crops and weeds, they employed a digital camera mounted on a moving platform, such as a cart or a tractor. After that, filters were used to pre-process the images. Then, in 2020, they removed. These articles use features from the segmented regions of interest, such as colour, texture, and shape, to demonstrate how computer vision can precisely locate objects inside the home. Using a machine learning algorithm that was trained using a set of annotated photographs, scientists identified the separated areas of interest as either crops or weeds. Post-processing procedures, such as morphological operations to exclude small items and smoothing techniques to remove noise, were used to further the categorization findings. The system's main flaws include limited scalability and environmental concerns.

“Weed Identification Using Deep Learning and Image Processing in Vegetable Plantation” is the topic of paper [3] by the author. The authors discussed several approaches, such as employing a camera to take pictures of the weeds and plants, pre-processing the images to enhance their quality, and identifying the weeds using a convolutional neural network (CNN). The CNN, trained on a big dataset of image data, is more accurate thanks to transfer learning. After the categorization results are shown on a user interface, farmers can either manually or automatically identify and remove the weeds. Overall, the application of this technology may result in a reduction in the amount of time and labour required for weed management in vegetable plantations. The system’s flaws are its expensive cost, small dataset, and ineffectiveness in all vegetable crops.

In paper [4] “Weed Detection in Farm Crops Using Parallel Image Processing” is discussed by the author. In order to begin the procedure, photos of the crops are taken using a high-resolution camera placed on a tractor. The photographs are then pre-processed to raise their quality and reduce noise. The pre-processed pictures are then divided into different sections using a clustering technique. The segmented areas are separated into crop and weed categories using a support vector machine (SVM) classifier. The SVM classifies the regions using characteristics such as colour, texture, and form after being trained on a dataset of tagged pictures. Farmers can use the resulting weed map to find and remove weeds. The system is appropriate for large-scale farming operations since it is created to be effective, precise, and scalable. The system’s two main drawbacks are that it uses a lot of computer resources and that not all types of farm crops can be detected for weeds using it.

In paper [5] the study “Deep Learning-Based Approach for Weed Detection in Potato Crops” aims to develop an automated weed detection system using deep learning techniques for potato crops. The procedure comprises gathering photos of weeds and potato plants from several fields, pre-processing the photos, and using a Convolutional Neural Network (CNN) to train a deep learning model. A sizable collection of photos with either “potato” or “weed” labels is used to train the model. The trained model is then put to the test on fresh photos to gauge how well it can identify weeds. The findings demonstrated that the suggested method obtained a high level of weed detection accuracy, making it a potentially useful tool for managing weeds in potato fields.

In the manuscript [6] the goal of the project “Deep Learning-Based Approach for Weed Detection in Potato Crops” is to create an automated system for weed identification in potato crops by utilizing deep learning methods. The technique entails collecting images of weeds and potato plants from several fields, preparing the images, and then training a deep learning model with CNN. The model is trained using a large set of images labelled as “potato” or “weed”. Next, the trained model is evaluated on new images to see how effectively it recognizes weeds. The results showed that the recommended strategy achieved a high degree of accuracy in weed detection, indicating that it might be a helpful tool for controlling weeds in potato fields.

3 Proposed Work

Weed detection is highly valued in agriculture since it is crucial to maintaining the health and yield of crops. Large-scale weed identification is challenging since conventional techniques require a lot of physical labour and take a long time. Deep learning and image processing are currently used in promising weed detection techniques. The suggested method utilizes deep learning and image processing to identify and classify weeds successfully and automatically. To determine whether the input image contains weeds, convolutional neural networks (CNN) are used to extract features from the image.

Figure 1 depicts the proposed system The dataset, which is in CSV format, needs to be converted into a text file format that YOLO can read after we manually identify each object of interest with a bounding box and extract the photographs from the dataset. The YOLO model is then trained using the dataset, which entails changing the model's weights in response to errors it makes. The model can be used to extract features from fresh photos after it has been trained, and these features can then be used to build a scan window or box around each object that is spotted. Last but not least, the YOLO model delivers a precise output that includes the object's label and the coordinates for its enclosing box. Throughout this process, a convolutional neural network (CNN) is used to learn to extract relevant features from the input images, which is crucial for accurate object detection.

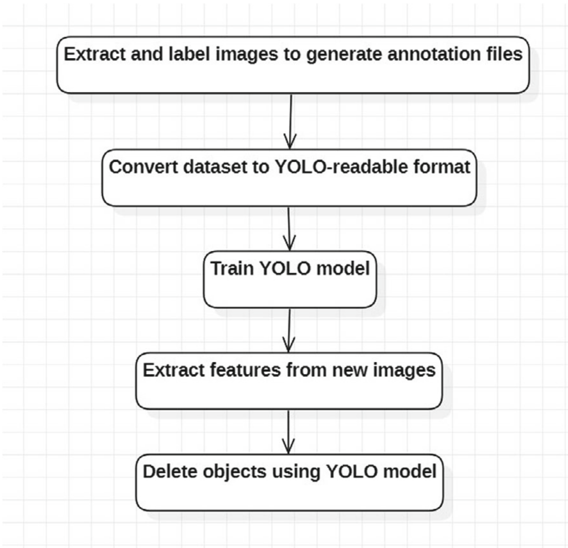

Fig. 1 Proposed system architecture

4 Convolutional Neural Network

Convolutional neural networks (CNNs), a type of deep learning architecture, were developed especially for analyzing images.

Tasks include segmentation, object detection, and image classification. It has a number of layers, including convolutional layers, pooling layers, and fully connected layers, and uses the convolution approach to extract attributes from the input image. The core idea behind CNNs is to train a set of filters—sometimes called kernels—that can identify important patterns or features in a picture. These filters are applied using convolution to the input picture, resulting in a collection of output feature maps that depict different facets of the image. In addition, CNNs use fully connected layers to complete the final classification task and pooling layers to down sample the output feature maps and lower the spatial dimensionality of the data (Fig. 2).

Different Layers of CNN are:

1. Convolutional Layer: A convolutional layer is often the first layer in a CNN. This layer applies a variety of filters on the input image to extract features. The input image is convolved with the weight matrix of each filter to produce a feature map.
2. Pooling Layer: A pooling layer is frequently added following each convolutional layer. By picking the highest or average value within a frame of pixels, this layer decreases the feature map's spatial dimensions. This makes the features more resilient to little changes in the input and lowers the computational cost of the network.
3. Activation Function: The output of each layer is subjected to an activation function to introduce nonlinearity to the network. The most common activation function, Rectified Linear Unit (ReLU), zeros out all negative values.
4. Fully Connected Layer: Similar to the layers in a conventional neural network, the final layers of a CNN are frequently completely linked layers. These layers

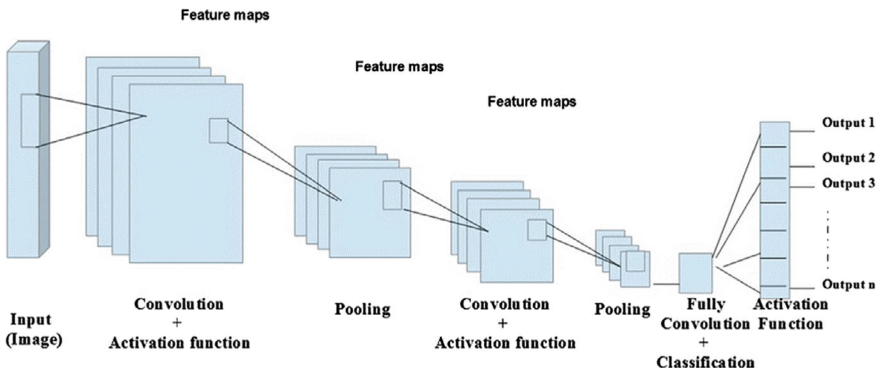

Fig. 2 CNN architecture

apply a set of weights to the previous layers' flattened output to create the final output.
5. Dropout Layer: After the completely connected layer, a dropout layer may be added to avoid overfitting. In order to lessen the co-adaptation of neurons, this layer randomly removes part of the neurons during training.
6. Softmax Layer: Similar to the layers in a conventional neural network, the final layers of a CNN are frequently completely linked layers. These layers apply a set of weights to the previous layers' flattened output to create the final output.

In general, a CNN's architecture is created to draw out useful elements from the input image and use them to generate predictions. The network may learn progressively complicated characteristics that are more and more invariant to minute changes in the input by stacking numerous convolutional and pooling layers.

5 Proposed Algorithm for Weed Detection

Using the YOLO Algorithm, we attempt to distinguish the weed images from the other crops for this project. Convolutional networks are used by YOLO, which was chosen because of its outstanding performance in item and sample recognition. This is why, in addition to monitoring moving objects, it now has real utility in areas like transportation and animal identification.

YOLOV3 Algorithm

First, we process the input image and create a blob from it by swapping the red and blue colour channels, shrinking the image to a fixed size of (512, 512), and scaling pixel intensities to the range [0, 1]. The blob is then designated as the YOLO network's input. The YOLO network is then operated to provide a list of outputs from the network's output layers called layer Outputs. Each output layer is in charge of identifying items of various sizes and corresponds to a particular scale in the image. The class ID and confidence ratings of the discovered objects are extracted when we cycle through each layer's output and each detection contained inside it. The method scales the bounding box coordinates back in relation to the size of the input image if the confidence score is higher than a predetermined threshold. It also updates the lists of bounding boxes, confidences, and class IDs. Finally, we use the class name and confidence score to create bounding box rectangles and labels on the input image for each object that was detected.

A one-stage object detection model, the YOLOV3 (You Only Look Once version 3) algorithm predicts object bounding boxes and class probabilities from input photos in a single pass. This is how it goes:

- Input image: At first it takes an input image and divides it into a grid of cells.
- Feature extraction: To extract information from the input image, a deep convolutional neural network (CNN) architecture processes the image. A feature pyramid

network (FPN) is used by the YOLOV3 architecture to detect objects at various scales and resolutions.

- Information about the object's width, height, centre coordinates, and a confidence level that expresses how likely it is that the box contains an object. The possibility that an object falls into one of several predetermined categories is represented by the class probabilities.
- Non-maximum suppression: The predicted bounding boxes are then subjected to a non-maximum suppression (NMS) algorithm by the YOLOV3 algorithm in order to remove redundant detections and keep only the most precise bounding boxes.
- Output: The final output of the YOLOV3 algorithm is a set of predicted bounding boxes and class probabilities for each object detected in the input image.

RCNN Algorithm

The second approach we've employed is known as RCNN (Regions with Convolutional Neural Networks), and it identifies objects in an image by combining region suggestions and convolutional neural networks (CNNs). It entails picking areas of an image that could contain an item, using CNNs to extract features, and then classifying the object with an SVM.

First, using a selective search method, we pre-process the images and annotations by scaling them, changing the format of the comments, and creating region suggestions. To make the data compatible with the RCNN architecture, which necessitates inputs of a specific size and format, this data pre-processing step is required. After creating the region suggestions, we used intersection over union (iou) to compare the generated regions with the labels on the ground truth. Any region with iou > 0.5 is saved as a positive example (it may object), and any region with iou 0.2 is preserved as a negative example.

There are three parts in RCNN:

1. CNN fine-tuning: First, using my created region suggestions, we adjusted the VGG16 model, which has three output classes (Crop, Weed, and Background) and an input size of $224 \times 224x3$. By updating the weights of the fully connected layers of the CNN while maintaining a weight-free state for the convolutional layers, this procedure entails training the CNN on the detection task. The CNN is taught to categorize object suggestions as either foreground (object) or background (background) and to fine-tune their placements during fine-tuning. The input image's features must be extracted in this stage in order to be utilized later on by the RCNN.
2. CNN + SVM training: The retrieved features from the improved CNN are used to train a support vector machine (SVM) classifier in the second stage. From the refined model, we eliminated the final two fully linked layers and used the CNN model as a feature extractor. Each object suggestion is represented as a fixed-length feature vector in the CNN output, which is a feature map that represents the image. In order to categorize each proposed object as a certain object class, the SVM classifier is trained on these feature vectors. A multi-class SVM or multiple

binary SVMs can be used to train the SVM to categorize different object classes. The SVM can be used to classify object suggestions and eliminate false positives once it has been trained.

3. Bounding box regression: It is a technique for enhancing the bounding boxes generated by early object proposals. Each sample includes an image, item suggestions, and the bounding box for the associated ground truth. The model is trained using a set of labelled data. By minimizing a loss function that penalizes the difference between the anticipated offsets and the actual offsets, the model learns to improve the bounding boxes. By better localizing items in an image, bounding box regression can considerably increase the accuracy of object detection. Even though we didn't apply this tactic during the training method, we still got good accuracy.

6 Results

We have used Sesame crops and weed dataset which was obtained from Kaggle and also we captured some images from the field and each image is of 512×512 coloured image (Fig. 3).

Fig. 3 Dataset images

YOLOV3 Weed Detection Performance

The graph in Fig. 4 shows the relation between the number of iterations and the average loss value. As the number of iterations increases the average loss value decreases the less the average loss value the more is the accuracy. We put the no of batches as 4000 which represents 4000 iterations and we run our training model up to 3200 iterations, we got the current average loss as 0.3496 which is low and as we increase the number of batches; the average loss value will decrease.

Once the model is trained, it can be used to extract features from new images, which are then used to create a scan window or box around each detected object. Finally, the YOLO model provides an accurate result, including the label of the object and the coordinates of its bounding box.

By considering several factors and based on the output obtained in order to assess the performance of the suggested YOLOV3 algorithm, the following counts were reached after considering 100 images (Table 1).

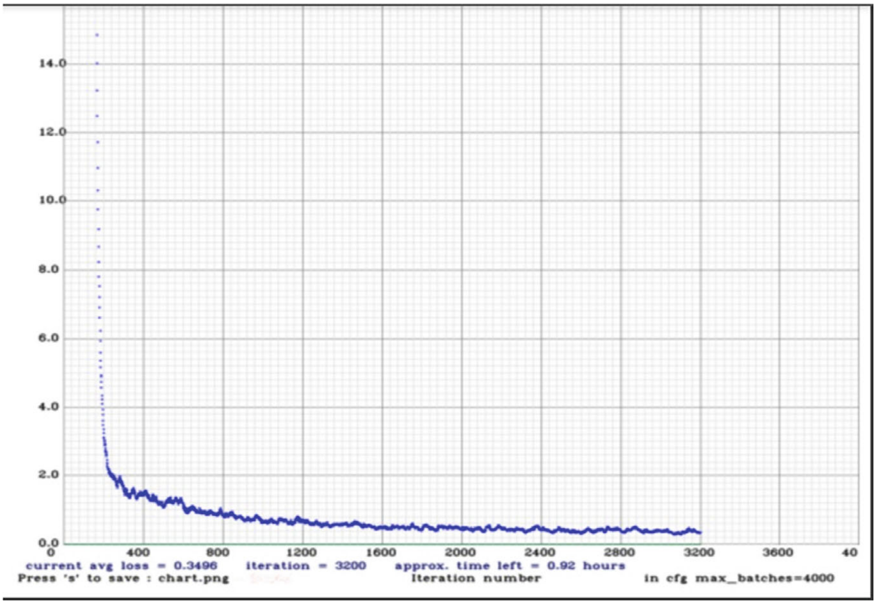

Fig. 4 Analysis of results

Table 1 Truth Table for YOLOv3

		Actual (True) values	
		Positive	Negative
Predicted values	Positive	49	2
	Negative	2	47

Table 2 Truth table for RCNN

		Actual (True) values	
		Positive	Negative
Predicted values	Positive	48	3
	Negative	2	47

With the use of the aforementioned variables, we discovered that the average accuracy was 96%, which is good and gives the YOLOV3 algorithm a good indicator of how well it is performing in comparison to other algorithms.

RCNN Weed Detection Performance

With the use of the aforementioned variables, we discovered that the average accuracy was 95%, which is good and gives the RCNN algorithm a solid indicator of how well it is functioning in comparison to other algorithms (Table 2).

7 Comparison of YOLOV3 and RCNN

Popular object detection techniques used in computer vision to find items in images include YOLOv3 and RCNN. They take different approaches to object detection, though.

A single-stage object detection method called You Only Look Once v3 (YOLOv3) divides the input image into a grid of cells before starting to look for things. Each cell forecasts a particular number of bounding boxes based on their class probability and confidence scores. Since the system only employs a single convolutional neural network (CNN) to generate predictions, it is speedier and more efficient than earlier object detection systems.

RCNN, on the other hand, is a two-stage object identification method that creates region proposals (i.e. potential item locations) before using a CNN to classify each proposal and enhance its bounding box. As a result, RCNN is able to handle a wider range of object sizes and forms. However, it is slower and uses more computer resources.

Both systems had excellent results for weed and crop detection. In our case, YOLOV3 provided an accuracy score of 96% and RCNN provided a score of 95%, both of which are excellent. The Detailed Comparison can be seen as shown in Table 3 which shows the different parameters.

Both YOLOv3 and RCNN can be efficient for weed and crop detection, although how well they perform may vary depending on the application and dataset. While RCNN may be preferable for smaller-scale applications where accuracy is more crucial, YOLOv3 may be more appropriate for real-time detection in large-scale areas where speed is crucial. To decide which model performs better for the current application, it is advised to compare the two models on the relevant task and dataset.

Table 3 Comparison of YOLOv3 and RCNN

Performance metrics	YOLOv3	RCNN
Accuracy	96%	95%
Precision	96%	94.11
Recall	96%	96
Specificity	95.91%	94
F1 score	0.96	0.95

8 Conclusion and Future Enhancements

In precision agriculture, weed detection is a crucial activity that can assist farmers in minimizing the effect of weeds on crop yields and reducing the usage of herbicides. Farmers can target weed infestations more efficiently and apply targeted weed control techniques for accurate weed detection. Weed detection is significant because it has the ability to increase agricultural yields, decrease the usage of herbicides, and lessen the environmental effects of weed management techniques.

We compared the effectiveness of YOLOv3 and RCNN for weed and crop detection in agricultural environments in this work. Using a collection of aerial photos of crops and weeds, we assessed both algorithms' performance in terms of accuracy and speed. Our findings demonstrated that both algorithms had superior sensitivity to small item detection. Overall, both models showed excellent performance in identifying crops and weeds, demonstrating the potential of both for use in precision agriculture.

Our study emphasizes the significance of choosing a suitable model for the particular application and dataset. While RCNN may be preferable for real-time applications where accuracy is more crucial, especially when recognizing tiny objects, YOLOv3 may be more appropriate for real-time applications where speed is crucial.

Future advancements can be done by combining deep learning models with other technologies like GPS, robotics, and sensors to provide a complete weed management system. Future research can focus on developing integrated systems that can provide real-time weed detection, mapping, and treatment.

References

1. Rakhmatulin I, Kamilaris A, Andreasen C (2021) Deep neural networks to detect weeds from crops in agricultural environments in real-time: a review. Remote Sensing 13(21):4486
2. Tejeda AI, Castro RC (2019) Algorithm of weed detection in crops by computational vision. In: 2019 International Conference on Electronics, Communications and Computers (CONIELECOMP). IEEE, pp 124–128
3. Jin X, Che J, Chen Y (2021) Weed identification using deep learning and image processing in vegetable plantation. IEEE Access 9:10940–10950

4. Umamaheswari S, Arjun R, Meganathan D (2018) Weed detection in farm crops using parallel image processing. In: 2018 Conference on Information and Communication Technology (CICT). IEEE, pp 1–4
5. Khan F, Zafar N, Tahir MN, Aqib M, Saleem S, Haroon Z (2022) Deep learning-based approach for weed detection in potato crops. Environ Sci Proc 23(1):6
6. Phung, Rhee Figure 1. Schematic diagram of a basic convolutional neural network. ResearchGate, 03-Jun2021. [Online]. Available https://www.researchgate.net/figure/Schematic-diagram-of-a-basicconvolutional-neural-network-CNN-architecture-26_fig1_336805909. [Accessed: 23-Dec-2021]
7. Zhao L, Li S (2020) Object detection algorithm based on improved yolov3. MDPI. [Online]. Available: https://www.mdpi.com/2079-9292/9/3/537. [Accessed: 24-Dec-2021]
8. Adarsh P, Rathi P, Kumar M (2020) YOLO v3-Tiny: Object detection and recognition using one stage improved model
9. Pham T, Dao S (2020) Plant leaf disease classification based on feature selection and deep neural network
10. Elstone L, How KY, Brodie S, Ghazali MZ, Heath WP, Grieve B (2020) High speed crop and weed identification in lettuce fields for precision weeding. Sensors 20(2):455