Efficient Weed Detection in Agricultural Landscapes using DeepLabV3+ and MobileNetV3

Renuka Devi Rajagopal, Manthena Rishit Varma
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Heshalini Rajagopal
Department of Electrical & Electronic Engineering, Mila University, No. 1, Persiaran MIU, 71800 Putra Nilai, Negeri Sembilan Darul Khusus, Malaysia

Email: renukadevi.r@vit.ac.in, m.rishitvarma@gmail.com

Abstract

Weed detection is a crucial task in precision agriculture, significantly impacting crop yields and reducing the dependency on herbicides. Effective weed management enhances agricultural productivity by ensuring that crops receive adequate nutrients, water, and sunlight, which weeds would otherwise consume. Traditional weed control methods are labor-intensive and often rely heavily on chemical herbicides, which can have detrimental environmental effects. This paper presents a deep learning approach for weed detection, utilizing the DeepLabv3+ model with a MobileNetv3 backbone. This study underscores the potential of integrating advanced deep learning techniques into agricultural practices, paving the way for more sustainable and efficient weed management strategies.

Keywords: Deep Learning, Semantic Segmentation, Agriculture, Weed Detection, Image Processing, Computer Vision

1. Introduction

Weeds are a significant challenge in agriculture, competing with crops for essential resources such as nutrients, water, light, and space. This competition can drastically reduce crop yields, thereby impacting food production and the overall agricultural economy. Weeds can also serve as hosts for pests and diseases, further threatening crop health and productivity. The presence of weeds often necessitates increased labor and financial input for their control, making weed management a critical concern for farmers globally. Traditionally, weed management has relied heavily on chemical herbicides and manual labor. While effective, these methods come with several drawbacks. Chemical herbicides can have harmful environmental impacts, contaminating soil and water sources and affecting non-target plant and animal species. Moreover, over-reliance on herbicides can lead to the development of resistant weed species, complicating control efforts. Manual weeding, on the other hand, is labor-intensive and time-consuming, making it impractical for large-scale farming operations. To address these challenges, precision agriculture has emerged as a promising approach. Precision agriculture aims to optimize field-level management regarding crop farming. One of the key aspects of precision agriculture is the development of advanced technologies for weed detection and control. Accurate and timely identification of weeds is crucial for effective weed management, allowing for targeted interventions that minimize resource use and environmental impact. This paper presents an innovative solution for weed detection using deep learning, leveraging the capabilities of using DeepLabv3+ with a MobileNetv3 backbone. DeepLabv3+ is a state-of-the-art deep learning model designed especially for semantic segmentation, which involves classifying each pixel in an image into a predefined category. In this context, it is used to distinguish between crops and weeds within agricultural images. MobileNetv3, on the other hand, is a lightweight and efficient neural network architecture optimized for mobile and embedded vision applications. Its integration with DeepLabv3+ enhances the model's efficiency, making it suitable for real-time applications in the field where computational resources may be limited. The proposed system works by capturing images of the agricultural field using cameras mounted on drones or ground-based vehicles. These images are then processed by the DeepLabv3+ model, which uses MobileNetv3 as its backbone to perform real-time weed detection. The model analyses the images, segmenting the weeds from the crops with high accuracy. The integration of DeepLabv3+ and MobileNetv3 presents a powerful tool for precision agriculture, offering an efficient and accurate method for weed detection. This approach not only enhances the sustainability of weed management practices but also supports the broader goals of precision agriculture by optimizing resource use and minimizing environmental impact. According to the various papers referred and studied, there are many ways of implementation like i) a method for weed detection using low-level features like color and area to distinguish weeds from crops using Random Forest classifier (used in the CWFI dataset), ResNet-50, Inception-v3 models, ConvNets, SVM, AdaBoost, Random Forest models [1]. ii) an algorithm which detects weeds by segmenting green plants, using median and morphological filters, and area-based thresholding by using binary classification using image-processing techniques in MATLAB [2]. iii) a deep neural networks process crop images to detect weeds, which are then eliminated using a robotic arm with a weed cutter which uses YOLOv5 and CNN models [3]. iv) using Erosion and Dilation approach with Raspberry Pi where weeds are detected by converting images to binary, applying erosion and dilation, and counting white pixels exceeding a threshold [4]. v) a model which uses TensorFlow Lite where weeds are identified using TFL Classify app, which processes images from ESP32 AI Cam and directs herbicide spraying [5]. vi) a method utilizing YOLOv5, for accurately identifying and localizing weeds in agricultural fields [6]. vii) another method which works on YOLOv4, a state-of-the-art object detection algorithm which accurately identifies and classifies weeds in agricultural fields [7]. viii) with help of Yolov5, weed detection is done where the modified model replaces the 3x3 convolution in Yolov5's backbone with multi-head self-attention (MHSA) to improve weed detection accuracy [8]. ix) a method which proposes usage of YOLOv3 object detection algorithm for accurately identifying weed crops by applying CNN [9]. x) an approach where three deep learning models R-CNN, YOLOv3, and CenterNet are used for weed detection and measuring weed growth by calculating length and breadth [10]. The major challenges observed here are: (i) capturing image without any movements, (ii) good intensity level must exist, iii) there must be a standard height maintained between the crop top and the camera.

2. The Model's Architecture

DeepLabv3+ with MobileNetV3 is a powerful and efficient model for weed detection, combining the segmentation capabilities of DeepLabv3+ with the lightweight architecture of MobileNetV3. This combination enables accurate weed identification with reduced computational resources, making it suitable for real-time applications in agriculture. Its use of Atrous convolution and depth wise separable convolutions ensures high performance and fast inference times.

2.1. *MobileNet V3*

From Fig. 1, we can see the architecture of MobileNetV3, which can serve as a backbone network for models like DeepLabV3+ used for tasks such as weed detection. The architecture of MobileNetV3 consists of multiple layers. First comes the NL,Dvis layer where this layer represents a non-linear, depth wise convolutional layer. In this layer, the input image is processed using depth wise separable convolutions, which apply a single convolutional filter per input channel. This approach is computationally efficient and helps reduce the number of parameters. The "NL" likely stands for "non-linear," indicating that a non-linear activation function (e.g., ReLu) is applied after the depth wise convolution. Next comes the 3x3 block which represents a standard convolutional layer with 3x3 kernels. It takes the output feature maps from the previous layer and applies a set of 3x3 convolutional filters to extract higher-level features. This layer helps capture spatial and structural information in the input image. Next comes the Pool layer which down samples the spatial dimensions of the feature maps. It applies a sliding window operation, selecting the maximum or average value within each window. This reduces computational complexity and introduces translation invariance. The pooled features capture essential information while reducing spatial resolution. Pooling creates a spatial hierarchy, enabling detection of complex patterns. Then we have the FC, Relu and FC, Hard. These blocks represent fully connected (FC) layers with different activation functions. The "FC, Relu" branch uses the ReLU (Rectified Linear Unit) activation function, which introduces non-linearity by setting negative values to zero. The "FC, Hard" branch likely uses a different activation function, such as the hard swish or hard sigmoid, which can potentially improve performance for certain tasks. As shown in Fig. 1, the symbol X likely indicates a point where the two branches (FC, Relu and FC, Hard) are combined or merged. This could be done through an element-wise operation like addition or concatenation, allowing the network to leverage features from both branches. The "NL,|X|" component at the end likely represents a non-linear activation function applied to the merged features from the two branches (FC, Relu and FC, Hard).

In the context of weed detection using DeepLabV3+ with MobileNetV3 as the backbone, the input images would go as: the NL,Dvis layer extracts low-level features like edges, textures, and shapes from the input image using depth wise separable convolutions and non-linear activations. Then the 3x3 layer further processes these low-level features and captures spatial and structural information. The pooling layer down samples the feature maps, reducing their spatial dimensions while preserving essential information. The fully connected layers (FC, Relu and FC, Hard) capture higher-level semantic information crucial for tasks like object detection and segmentation. The merged features from the two branches are then fed into the main model which is chosen as per requirement. In our approach we use the main model as DeepLabV3+ to which the MobileNetV3 model acts as a backbone or foundation model to it.

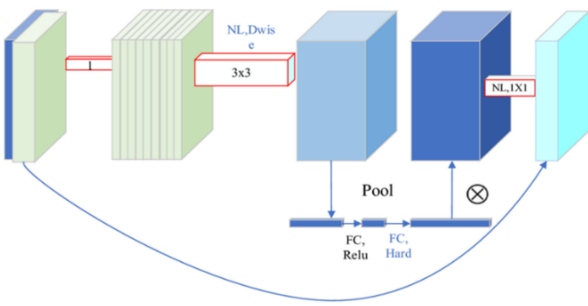

Fig. 1 Architecture of MobileNetV3

2.2. *DeepLabv3+*

From Fig. 2, we can see the architecture of DeepLabV3+, with a focus on the ASPP (Atrous Spatial Pyramid Pooling) layer. In DeepLabV3+ also we have number of layers which perform unique tasks at each level. The architecture of this DeepLabV3+ comprises of 2 main layers namely Encode and Decoder.

Encoder Layer

Encoder layer is one which again consists of 4 major levels namely 1x1 Conv, 3x3 with different rates of 4,8,12, Global Pooling Layer, ASPP (Atrous Spatial Pyramid Pooling) layer. The Encoder layer also consists of the main backbone model used for DeepLabV3+ model. In the Fig. 2 we can see that a dedicated "Deep Convolution Neural Network" block is present which contains the backbone model and, in our case, it is the MobilNetV3 model. The outputs of the backbone model are sent to the next 4 layers of DeepLabV3+ model. This all happens in the Encoder Layer itself. Now comes the 1x1 Conv Layer which applies 1x1 convolutional filters to the input segmented image, projecting the feature maps to a lower-dimensional space, which helps reduce computational complexity. Next comes the 3x3 layer each with a rate of 4,8,12 individually. These are parallel 3x3 dilated convolutional layers with different dilation rates (4, 8, and 12). Dilated convolutions introduce sparse sampling patterns, allowing the model to capture multi-scale information and expand the effective receptive field without increasing the number of parameters. Now the Global Pooling layer applies global average pooling or global max pooling to the feature maps, capturing global context information from the entire input image. This global context can help the model differentiate between different regions based on their spatial distribution and relationships. The last layer in the Encoder layer of DeepLabV3+ is the ASPP (Atrous Spatial Pyramid Pooling) layer. This is a crucial component of DeepLabV3+ that effectively combines multi-scale information. The ASPP module applies parallel dilated convolutions with different dilation rates (e.g., 6, 12, 18) and also includes an image-level feature pooled globally. By capturing information at multiple scales, ASPP enables the model to accurately segment objects of various sizes and shapes. The features from the parallel dilated convolutions and the global pooling are concatenated and processed by a 1x1 convolutional layer, effectively combining the multi-scale information. This multi-scale representation is essential for weed detection, as weeds can have varying sizes, shapes, and spatial distributions within the input image.

Decoder Layer

The second layer in DeepLabV3+ is the Decoder layer which in turn has 5 important layers in it. Those are 1x1 Conv, Up sample by 4 layer, 3x3 Conv, Concatenate and again Up sample by 4 layers after which we get the final output image. Each layer has its own abilities and those are as follows: The 1x1 Conv layer combines and processes the multi-scale features from the Encoder outputs. Next the Up sample by 4 layer performs an operation which up samples the low-resolution feature maps from the Encoder, increasing their spatial resolution. This is necessary for producing a high-resolution segmentation map at the end of the process. After that the 3x3 Conv (Convolutional) layer processes the up sampled features which come from the previous layer refining the spatial information. Then the Concatenate layer performs concatenation of up sampled features from the Decoder with the corresponding low-level features from the Encoder. This helps the model leverage both high-level semantic information (from the Encoder) and low-level details (from the early layers of the Encoder) for accurate weed segmentation. At last, the Up sample by 4 again performs up sampling operation which further increases the spatial resolution of the feature maps, preparing them for the final segmentation map of the weeds.

The output segmentation map assigns a class label (weed or non-weed) to each pixel in the input image, enabling precise localization and identification of weeds. By using segmented images as input to the MobileNetV3 backbone, the model can leverage the pre-computed segmentation information to focus on extracting relevant features for weed detection. The multi-scale and global context information captured by the ASPP layer, combined with the low-level details from the early layers, enables the model to accurately segment and localize weeds, even when they appear at different scales and spatial distributions within the input image.

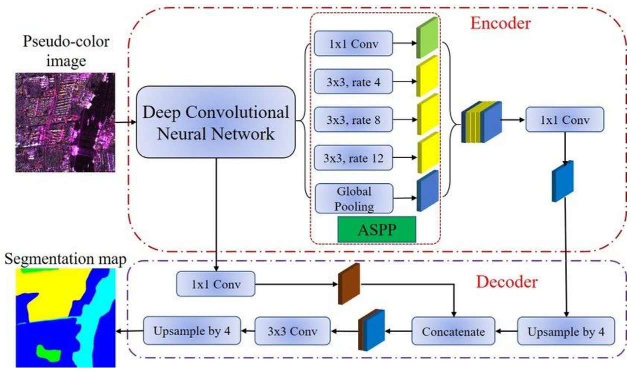

Fig. 2: Architecture of DeepLabV3+ model

3. Methodology

As shown in Fig. 3, initially we set up a camera which captures live aerial view of crops in the field with the help of OpenCV. Now the captured images are pre-processed and are sent to our trained model i.e. the DeepLabV3+ model with MobileNetV3 backbone. Inside the model, the input image is sent to the Encoder layer of the DeepLabV3+ model. In that Encoder layer, initially the images are sent to the MobileNetV3 model and the outputs from that model are then sent to other layers present in the Encoder layer. Now the remaining layers perform their operations and pass the results to the Decoder layer of DeepLabV3+ model. Sending into it, the sublayers in Decoder layer perform their operations and finally provides the high quality semantic segmented images of weeds present in the original images. Now convert the model's output to the binary mask and if needed for good visualization, apply color map to the binary segments. After that overly these color segmented maps with the original frame captured. Finally display the overlaid image at that instant on the screen. Now these segmented maps of weeds are further used for many applications in the field of precision agriculture.

Algorithm: Weed Detection using DeepLabV3+ with MobileNetV3
Input: Live Video from Camera
Output: Segmented Maps of Weeds
- *Step 1: Load necessary libraries (like OpenCV, TensorFlow)*
- *Step 2: Load the custom trained DeepLabV3+ model with MobileNetV3 backbone.*
- *Step 3: Continuously capture frames from the live video feed through OpenCV.*
- *Step 4: Preprocess the captured frame (formatting, resizing etc.,)*
- *Step 5: Pass the pre-processed frame to the loaded model*
- *Step 6: Obtain the segmentation map as output.*
- *Step 7: Convert the model's output to binary output And apply a color map to the mask to visualize the weed.*
- *Step 8: Overlay the segmentation mask with the original frame.*
- *Step 9: Display the frame with the overlaid weed detection mask.*
- *Step 10: Release the camera and close all OpenCV windows.*

DeepLabv3+ with MobileNetV3 is particularly well-suited for scalability in large-scale weed detection applications due to its lightweight architecture and efficient performance.
MobileNetV3 is designed to be lightweight, making it suitable for deployment on a wide range of devices, from high-end servers to edge devices like drones and field robots. This allows the model to be scaled across various platforms without significant modifications. The efficient computation and reduced resource requirements mean it can process large volumes of data quickly, which is essential for large-scale agricultural operations. Additionally, its compatibility with mobile and embedded systems ensures that real-time weed detection can be performed directly in the field, reducing the need for constant connectivity and enabling more autonomous operations.

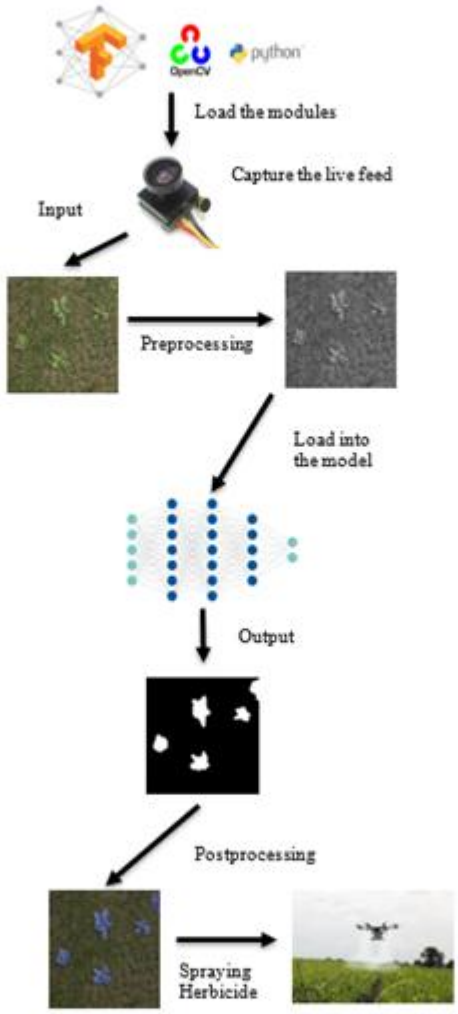

Fig 3: Pictorial representation of Algorithm

The segmented images of weeds from this model can be utilized in various applications across agriculture and related fields. In precision agriculture, these images can control robotic weeders for targeted weed removal and enable variable rate application of herbicides, thereby reducing chemical use and costs. In crop management, they can aid in yield prediction and field mapping to understand weed distribution and plan strategies accordingly. For research and development, the data can support weed ecology studies and crop breeding programs. Smart farming solutions can integrate these images into decision support systems and drone technologies for efficient weed monitoring. Additionally, they are valuable in environmental monitoring for biodiversity studies and managing invasive species. In education, they can create training materials and simulation models for agricultural training. The data also enhances machine learning models for better weed detection and segmentation accuracy and facilitates big data analysis to uncover patterns for future weed management practices. Commercially, these images can support agricultural consulting services and the development of customized weed management solutions,

leading to improved crop productivity and environmental sustainability.
Overall, the combination of DeepLabv3+ with MobileNetV3 offers a compelling mix of performance, efficiency, and scalability, making it superior to many other models for weed detection tasks. This makes it an ideal choice for extensive agricultural applications, ensuring high accuracy and real-time processing capabilities while being resource-efficient and versatile enough to be used across a broad spectrum of devices and platforms.

4. Experimental Results

The dataset used for this weed detection research consists of 1,400 high-resolution RGB images, captured from various agricultural fields to ensure a comprehensive representation of different weed species and environmental conditions. These images were sourced from both publicly available agricultural image repositories and direct field data collection using a drone-mounted camera. To facilitate the training, validation, and testing of the weed detection model, the dataset was divided into three subsets: 1,160 images were allocated for training, 196 images for validation, and the remaining 44 images for testing. This division ensures that the model is trained on a diverse range of images and can be rigorously evaluated to confirm its effectiveness in accurately detecting weeds in various scenarios. The outputs of test images are shown in the Fig. 4 and Fig. 5:

Fig. 4: Sample RGB images of test data

From Fig. 4, we can see the original RGB test samples which are randomly taken from the test dataset i.e. on which the model is applied, and Fig. 5 shows the predicted masks of the weed in binary format which are generated by the DeepLabV3+ model. Now these outputs can be further analyzed to get the coordinates or areas of the weed mask and send or store that information for further applications in precision agriculture.

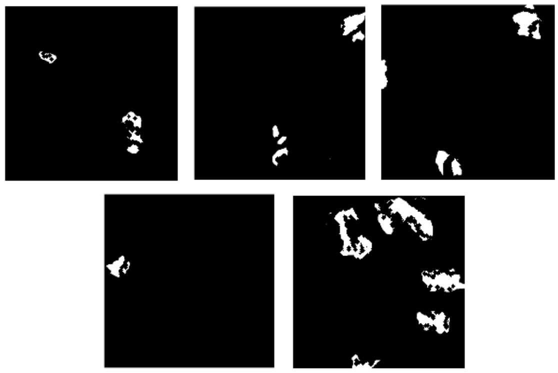

Fig. 5: Binary masks for the test images

Fig. 6 illustrates the relationship between the number of epochs and the corresponding accuracy achieved at each epoch during the training of the DeepLabV3+ model for weed detection. The training process of DeepLabV3+ involves optimizing the model's parameters to minimize the loss function, which measures the discrepancy between the predicted labels and the ground truth. As the number of epochs increases, the model iteratively adjusts its parameters to improve its performance on the training dataset. The overall accuracy achieved by the DeepLabV3+ model is around 84%, which is a significant performance metric for real-time weed detection applications. This high level of accuracy indicates that the model effectively distinguishes between weed and non-weed areas in the images. On other side Fig. 7 shows the graph between number of epochs vs the loss obtained at each point. The loss starts from 0.6 and gradually decreases to 0.06 which is considered to be minimal in real time. This overall metrics suggest that the model is performing well on weed images.

Weed detection using DeepLabv3+ with MobileNetV3 stands out when compared to other models due to its balance of accuracy, efficiency, and speed. While models like U-Net or traditional DeepLabv3 might offer high accuracy, they often require significant computational resources and processing time, making them less suitable for real-time applications in resource-constrained environments. In contrast, MobileNetV3 is specifically optimized for mobile and embedded devices, offering a lightweight yet powerful backbone that significantly reduces the model's computational requirements without sacrificing much accuracy. DeepLabv3+ enhances this further with its advanced segmentation capabilities, such as atrous spatial pyramid pooling and an improved decoder, which capture fine details and boundaries of weeds more effectively.
Compared to heavier models like ResNet or Inception used in segmentation tasks, DeepLabv3+ with MobileNetV3 is faster and more efficient, enabling deployment on edge devices in the field. This ensures quick processing and decision-making, which is critical

in precision agriculture. Additionally, the reduced computational overhead translates to lower energy consumption and cost, making it a more sustainable and practical choice for widespread agricultural use. Overall, the combination of DeepLabv3+ with MobileNetV3 offers a compelling mix of performance, efficiency, and scalability, making it superior to many other models for weed detection tasks.

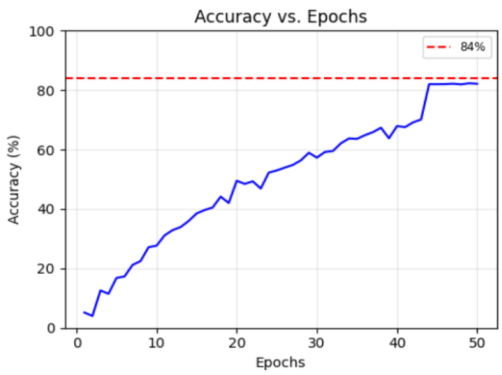

Fig. 6: Graphical analysis between Accuracy and Number of Epochs

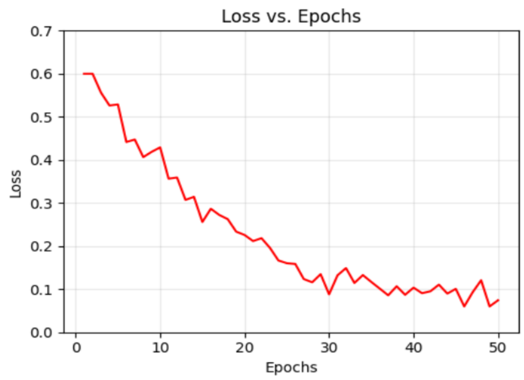

Fig. 7: Graphical analysis between Loss obtained and Number of Epochs

5. Conclusion

This research presents an innovative deep learning approach for weed detection using the DeepLabv3+ model with a MobileNetv3 backbone, offering a significant advancement in precision agriculture. The integration of these architectures provides a robust, efficient, and lightweight solution for real-time weed detection, addressing the challenges posed by traditional weed management methods. The DeepLabv3+ model's strength in semantic segmentation, combined with MobileNetv3's efficiency, enables accurate and fast weed identification, crucial for optimizing agricultural productivity and sustainability. Our methodology effectively leverages real-time image processing to distinguish between crops and weeds, facilitating targeted weed control interventions. This precision reduces the reliance on chemical herbicides, thereby mitigating their environmental impact and promoting sustainable farming practices. By harnessing advanced deep learning techniques, our approach enhances the efficiency of weed management strategies, contributing to better resource utilization and higher crop yields. The segmented weed images produced by our model can be utilized in various applications within the agricultural sector. These include guiding robotic weeders for precise weed removal, optimizing herbicide application rates, aiding in yield prediction, and supporting field mapping. Additionally, the data generated can be valuable for research in weed ecology, crop management, and machine learning model enhancement. Overall, the study underscores the potential of integrating advanced deep learning models into agricultural practices. The proposed DeepLabv3+ with MobileNetv3 backbone model not only provides an effective solution for weed detection with good accuracy but also paves the way for future innovations in smart farming. This approach exemplifies the move towards more sustainable and efficient agricultural practices, ensuring that modern farming can meet the growing demands for food production while preserving environmental integrity.

References

1. P. Khaire, V. Attar and S. Kalamkar, "A Comprehensive Survey of Weed Detection and Classification Datasets for Precision Agriculture," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-5, doi:10.1109/ICCCNT56998.2023.10306880.
2. A. J. Irías Tejeda and R. Castro Castro, "Algorithm of Weed Detection in Crops by Computational Vision," 2019 International Conference on Electronics,Communications and Computers (CONIELECOMP), Cholula, Mexico, 2019, pp. 124- 128, doi: 10.1109/CONIELECOMP.2019.8673182.
3. S. Tummapudi, S. S. Sadhu, S. N. Simhadri, S. N. T. Damarla and M. Bhukya, "Deep Learning Based Weed Detection and Elimination in Agriculture," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 147-151, doi: 10.1109/ICICT57646.2023.10134186.
4. R. Aravind, M. Daman and B. S. Kariyappa, "Design and development of automatic weed detection and smart herbicide sprayer robot," 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, India, 2015, pp. 257-261, doi: 10.1109/RAICS.2015.7488424.
5. T. B. M. I, J. J. Paul, B. Beulah and J. Joanna, "IoT Based Weed Detection and Removal in Precision Agriculture," 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2023, pp. 1-4, doi:10.1109/ICAECA56562.2023.10200751.
6. J. Prabavadhi and K. S, "Mobile Based Deep Learning Application for Weed and Medicinal Plant Detection Using YOLOV5," 2023 International Conference on System, Computation, Automation and Networking

(ICSCAN), Puducherry, India, 2023, pp. 1-5, doi:10.1109/ICSCAN58655.2023.10395553.
7. N. Khalid, H. Elkhiri, E. Oumaima, N. ElFahsi, Z. F. Zahra and K. Abdellatif, "Revolutionizing Weed Detection in Agriculture through the Integration of IoT, Big Data, and Deep Learning with Robotic Technology," 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 2023, pp. 1-6, doi: 10.1109/ICECCME57830.2023.10252598.
8. Z. Yu et al., "Weed detection based on improved Yolov5," 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2023, pp. 1936-1940, doi: 10.1109/ICSP58490.2023.10248865.
9. Y. Dandekar, K. Shinde, J. Gangan, S. Firdausi and S.Bharne, "Weed Plant Detection from Agricultural Field Images using YOLOv3 Algorithm," 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA, Pune, India, 2022, pp. 1-4, doi: 10.1109/ICCUBEA54992.2022.10011010.
10. S. N, M. Sundaram, R. Ranjan and Abhishek, "Weedspedia: Deep Learning-Based Approach for Weed Detection using R-CNN, YoloV3 and Centernet," 2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS), Kottayam, India, 2023, pp. 1-5, doi: 10.1109/iQ-CCHESS56596.2023.10391389.

Authors Introduction

Dr. Renuka Devi Rajagopal

Dr. R. Renuka Devi is an Associate Professor in the School of Computer Science and Engineering, VIT Chennai, India. Her research interests include Cyber-Physical Systems, Block Chain Technology, Data Mining and Machine learning in the field of Health Care.

Mr. Manthena Rishit Varma

He is currently pursuing Bachelor's Degree in Technology in Computer Science and Engineering with specialisation in Artificial Intelligence and Machine Learning at Vellore Institute of Technology in India.

Dr. Heshalini Rajagopal

She received her PhD and Master's degree from the Department of Electrical Engineering, University of Malaya, Malaysia in 2021 and 2016, respectively. She is an Assistant Professor in the School of Engineering and Computing, Mila University. Her research interests include image processing, artificial intelligence and machine learning.