Check for updates

A self-supervised overlapped multiple weed and crop leaf segmentation approach under complex light condition

Anand Muni Mishra[1] · Prabhjot Kaur[2] · Mukund Pratap Singh[3] · Santar Pal Singh[4]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

One of the most damaging obstacles to crop production is weeds; weeds pose a serious risk to agricultural output. Due to the homogenous morphological properties of weeds, farmers are unable to identify and classify the weed leaves.This study can aid farmers in identifying, categorizing, and quantifying the true extent of crop yield reduction. Computer vision is a sophisticated technique widely used for weed and crop leaf identification and detection in the agricultural field. This work has used three different datasets, such as 'Deep Weed', 'Crop Weed Filed Image Dataset (CWFID)', and Multi-view Image Dataset for Weed Detection in Wheat Field (MMIDDWF), and collected 5090 images for training the model. This work uses segmentation techniques for vegetation and semantics for weed object detection. Furthermore, the masked image is distributed as small tiles; often the patches are square tiles, as in 25×25 (px), 50×50 (px), and 100×100 (px). This work has proposed a Deep Learning segmentation model named 'Pyramid Scene Parsing Network-USegNet' (PSPUSegNet) for data classification and compared the accuracy of the data from existing segmentation models such as UNet, SegNet, and USegNet. The suggested model, PSPUSegNet, obtained 96.98% precision, 97.98% recall, and 98.96% data accuracy in the Deep Weed dataset. The proposed model has self-supervised in term of deep learning mechanism.Our findings demonstrate that the deep weed dataset has achieved greater data accuracy compared to the CWFID and MMIDDWF datasets. The findings support the effectiveness of the suggested approach for weed species recognition.

Keywords Segmentation · USegNet · PSPUSegNet · Semantic Segmentation · Weed Species

✉ Santar Pal Singh
  spsingh78@gmail.com

1   Chandigarh University, Mohali, Punjab, India

2   Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India

3   School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India

4   Department of Computer Science & Engineering, Rashtrakavi Ramdhari Singh Dinkar College of Engineering, Begusarai, India

1 Introduction

The weed is an unwanted plant in the crop fields. Weeds compete with crops for soil, nutrients, and sunshine, which cause crops to develop slowly and become smaller. This reduces agricultural production. Therefore, these nutrients are required for the growth of crop plants, but due to the presence of weed plants, crop growth is affected [1]. There are two important factors for crop yield loss: first, weed density and mix, and second, the similar morphological properties of weeds and crop plants. In the current situation, the farmers manually assess the weeds [2]. Another important factor is the overlap between weed plants. For these, the weed plant identification of overlapped weed plants, detection, coverage area, and growth stages are measured in the work. However, it has a tedious task for weed identification and classification, which has affected crop yield. So the automation of these tasks has been interesting to researchers in recent years [3]. Weed recognition is the focus of the computer vision system. The primary problem is changing the morphological properties of weeds and crop plants due to environmental conditions. So, the collection of the data field images is a tedious task, and another is the identification and classification of overlapped weed or crop leaves in computer vision. The main objective is to create effective models for the identification and classification of overlapping weed and crop leaves, uneven weed patch densities, varying sizes across multiple images, and discriminating the similar morphological properties of weeds and crop leaves [4]. In addition, the large extent of plantings or mixed crops weeds out computational time problems in image processing methods. The recent Deep Learning (DL) technique has proven to overcome the limitations of the classical image processing model [5].

Consequently, CNN has had great success classifying plant species. Crop disease detection, plant segmentation, and weed characterization, among other applications. However, CNN has some disadvantages [6]. The enormous number of manually annotated images needed to establish a model is one of them. Annotating the needed images by hand is a time-consuming and, in some cases, impossible operation. In 2016, the author proposed deep learning-based semantic segmentation for weed identification and detection of weed and crop leaves. They separate several crop species with a pixel accuracy of 79.59% [7]. However, their findings showed that employing CNNs for the task of identifying crops and weeds has enormous promise. Accurately classifying pixels as "corn" or "weed" in a two-class classification issue enhanced the approach to differentiate maize from seven distinct kinds of weed species in later trials. The author has attained 0.94 per-pixel accuracies, an F1-Score of 80% on the crop, and an Intersection-Over-Union (IoU) metric of 81% based on pixel-base data classification of weed and crop leaves [8].The major contributions to this work are:

   i. Computed the overlapped weed regions and density using vegetation segmentation and compared the three different datasets using the proposed model as a PSPUSegNet classifier.
   ii. This work has used a mixed approach of PSPNet and USegNet CNN models by replacing 7 Conv layers of UNet and 13 Conv layers of SegNet CNN models in downsampling to maintain the global feature of the data. This work has used the pooling indices (feature vector) from the encoder feature and transferred them for mapping to the corresponding upsampling layer.

iii. This work used pixel and tile wise data classification with different sizes of tiles as 25×25 (px), 50×50 (px), and 75×75 (px) and binary classification of images to achieve 9.7% IoU of data segmentation.

iv. Improve the model's scalability and generalizability by incorporating semantic and vegetation segmentation.

Since various plant species may only be identified by precise and nuanced taxonomic keys that may not even always be apparent in an image, segmenting multi-species overlapping weeds is a more challenging and difficult challenge than it has been previously [9]. This study uses a combined strategy to discuss multispecies overlapping segmentation. To eliminate the requirement for manual annotation, first provide a unique approach to integrating synthetic and single-species datasets. Then, suggest a novel architecture to carry out multispecies semantic segmentation effectively. Insufficient knowledge regarding weeds and crops has significantly contributed to the annual reduction in crop yield caused by weeds. This can provide additional support to the agricultural community in evaluating the precise crop quality, thereby promoting sustainable farming practices and their overall economic advancement.

The rest of the paper is structured as follows: Related work is illustrated in Section 2. Section 3 focused on the data description, and the methodology is discussed in Section 4. The performance analysis of the model and discussion are in Section 5. Finally, the conclusions drawn are presented in Section 6.

2 Literature study

According to recent research and studies in the field of agriculture, a variety of factors influence crop yield. Weeds are the foremost factor that could harm crop yield. Therefore, this is the most important task to identify and control weeds at an early stage of weed growth in the context of weed identification, detection, growth rate, and density estimation, which are reviewed in this literature, and include a comparison of different sources. This literature has included different deep-learning techniques for weed identification, detection, and classification.

Mishra et al. (2022) have discussed the different types of biennials and perennials, monocot and broad-leaved weed species, and weed control methods. It has also described the morphological and texture properties of common perennial weeds such as 'Paspalumdichotomum', 'Cynodondichotomum', 'Scirpusmaritimus', and 'Cyperusrotundus' in paddy crop agriculture. Furthermore, the author has also described weed control techniques such as biological, cultural, physical, and chemical methods. The authors use instance and semantic segmentation techniques for object detection, and the Gray Level Co-occurrences Matrix (GLCM), Hue, Saturation, and Value (HSV), are used for feature extraction. The author has applied different CNN techniques for image data classification and compared the techniques based on the performance of the model. There are a few performance parameters that have been discussed by the author in terms of precision, recall, F1-score, accuracy, Absolute Error (AE), and Mean Absolute Error (MAE) [10].

X Ma. et al. (2020) have discussed the RGB color photographs of seedling rice collected in a paddy field, and Ground Truth (GT) images were created by manually labeling the pixels in the RGB images with three distinct categories: rice seedlings, background, and weeds. The class weight coefficients were developed to address the issue of the

classification category numbers being unbalanced. 80% of the samples were chosen at random as the training dataset, while the remaining 20% were utilized as the test dataset. The suggested method has been compared against a traditional semantic segmentation model, specifically the FCN and UNet models. The SegNet method had an average accuracy rate of 92.7%, whereas the FCN and UNet methods had average accuracy rates of 89.5% and 70.8%, respectively [11].

Chechlinski et al. (2019) have suggested automated weeding called agro robotics. In this technique, weeds can be identified using robotic technology. The author has described the Internet of Things (IoT) and Deep Learning (DL)-based techniques that have automatically recognized weed identification and detection. The model has achieved 47–67% weed detection accuracy. It has been tested in four different plants in a stadium under medium lighting conditions. The robotics system has used the custom semantic segmentation CNN using UNet, DenseNet, and ResNet architectures. Out of this CNN architecture, the ResNet pre-trained model achieved better data accuracy (87%). The author suggested that weed images can easily be transferred to computer vision for another agro-robotic task [12].

Rasti et al. (2019) have discussed discriminating the weeds from the soya bean crop plant. The pre-trained DL models such as AlexNet, SqueezeNet, GoogLeNet, ResNet-50, SqueezeNet-MOD1, and SqueezeNet-MOD2 for training the model. Furthermore, 11,600 weed images have been collected from the Crop Weed Field Image Dataset (CWFID) and trained in the models. The ResNet-50 has achieved more than 92% data accuracy. AlexNet, SqueezeNet, GoogLeNet, SqueezeNet-MOD1, and SqueezeNet-MOD2 have achieved 94%, 91%, 87%, 90%, and 95% data accuracy consecutively. The author calculated the processing time of the pre-trained ResNet CNN model and achieved 40.73 s to process 11,600 images. However, the author suggested that it can also be implemented in biotic and abiotic leaf disease identification and detection [13].

Teimouri et al. (2018) have discussed 10 different types of weed species that grew in rabi and kharif crops. The author has explained the morphological and texture properties of weed leaves. Furthermore, the author described weed detection and classification techniques. There are 9649 weed and crop images collected for the standard data repository as the CWFID dataset. In this context, the author used three different classifiers, such as ResNet 150, Google Net, and the VGG-16 pre-trained CNN model, for data classification. Out of these, the VGG-16 model has achieved 96% data accuracy [14].

Kropff et al. (2021) have suggested a weed identification and detection technique based on four different steps: data collection, data segmentation, feature extraction, and finally data classification. Data has been collected from a multi-class deep weed dataset. After that, the data has been annotated as "Cynodon Dactylon,", "Convolvulus Arvensis", "Poa Annual", "Medicago Polymorpha,", and "Hypochaeris Radicata.". The unstructured RGB data has been resized to 256×256×3 and then implemented in the semantic segment for object detection. For the classification, we used SegNet, UNet, and ResNet151 CNN models and achieved 93.05%, 93%, and 92.78% data accuracy, respectively. The author has compared the proposed model in terms of accuracy and found that the SegNet CNN model provides better accuracy. The author also discussed the computation time of image processing in the CNN model. From the experimental results, it was found that the SegNet classifier consumed less time, i.e., 0.90 ms [15].

Zhao et al. (2017) have suggested the PSPNet model for pixel-wise data classification in the Line Mode-Occluded (LMO) dataset. This dataset has 33 classes of images and used 2,688 images for training the model. The author has used two benchmarks: PASCAL VOC 2012 and the Cityscapes benchmark. There is 85.4% mIoU and 80.2% data accuracy on PASCAL VOC 2012 and Cityscapes, respectively; using a single PSPNet data model [16].

In this literature, despite the use and usefulness of several CNNs for overlapped weed location, identification, detection, and density estimation in different crops using the pre-trained CNN technique, research has been challenging to detect multi-class weed species on target crops. Developing a hybrid DL technique that could quickly assess the condition of multi-class weeds in target crop fields would assist growers in determining target location, identification, and density estimation. This paper demonstrates an effective modified HDS-CNN model for weed location, identification and detection, and density estimation in soya bean crops on a large dataset.

3 Dataset description

In this study, the functional dataset was trained using instance and semantic segmentation. Three distinct datasets 'Deep Weed', 'CWFID' and 'MMIDDWF' are used in this study to annotate images. To expedite the manual annotation of real image datasets (dataset i) [13]. This work presents certain changes. Additionally, presents various techniques for creating datasets without the need for manual annotation: a) a technique for creating artificial datasets based on a single plant image (dataset ii) [17]; and b) a technique for creating actual field datasets made up of numerous plant images of a single weed species (dataset iii) [18]. The complete discussion of the dataset is given in the next subsection.

3.1 Deep weed dataset

To create an appropriate image collection for training and validation. Due to its potential to enhance agricultural output, research into robotic weed management has expanded recently. Deep Learning is the best method for identifying different weed species in challenging grassland habitats because of its unmatched accomplishments. This study provides the first sizable, public, multiclass image collection of weed species from Australian grasslands, enabling the development of reliable classification techniques to enable effective robotic weed treatment. This work has collected 1720 broad-leaf weed species such as 'Cerastiumvulgatum L.', 'Chenopodium album', and 'Amaranthusretroflexus' [19].

3.2 CWFID dataset

This dataset has a standard weed and crop image repository, and there are 2000 grass samples collected for training the model. Furthermore, 'Setariaverticillat' and 'Digitariasanguinalis' have collected 1200 and 800 weed images from grass weed species, which are available online (http://github.com/cwfid) [20]. For each image from the dataset, this work presents a Ground Truth vegetation segmentation mask and manual annotation of the plant category (crop vs. weed).

3.3 MMIDDWF dataset

The dataset intends to provide a public weed dataset to support the development of weed identification techniques in wheat fields and includes photos of wheat, broad-leaf weed, and grass weed in two modes and nine perspectives. This work has collected 1370 'Echinochloacrusgalli' broad-leaf grass weed images from the 'MMIDDWF' dataset [18]. This

work was developed to show the current status of leaf segmentation technology and the challenges of segmenting all leaves in a plant picture. The countability of the dataset is described in Fig. 1.

4 Methodology

The suggested method determines the weed-infested areas, weed leaf count, weed growth, and related weed density to treat the farmland under cultivation in a targeted manner. These four processes-processing, segmentation, feature extraction, and classificationwere used in this study to train the model.

4.1 Data enhancement and pre-processing of the image

This work has collected 5090 weed images from different sources. This dataset has been pre-processed and segments the particular object from the image. Therefore, it needs to enhance the quality of the image. This work has used the Contrast-Limited Adaptive Histogram Equalization (CLAHE) technique for data enhancement. These weed images are pre-processed using the 'CLAHE' technique, which improves image quality [21]. Data pre-processing, data segmentation, feature extraction, and data classification have all been assigned to the data flow. Furthermore, data segmentation has used semantic, vegetation, and background segmentation [22]. This algorithm is a method of computer image processing that boosts contrast in pictures. The adaptive method differs from typical histogram equalization in that it computes several histograms, each one corresponding to a distinct region of the image, and then uses them to distribute the brightness values of the image [23]. After that, each tile's transformation function is calculated. The pixels in the tile center are a good fit for the transformation functions [24]. All other pixels are given interpolated values and up to four transformation functions based on the center pixels of the tiles that are closest to them. The bulk of the image's pixels (shown in shaded blue) are interpolated bilinear; those near the edge (shown in shaded green) are interpolated linearly; and those near the corners (shown in shaded red) are converted using the corner tile's transformation function. This work has used segmentation techniques such as semantic segmentation, vegetation segmentation, and background segmentation for weed leaf detection
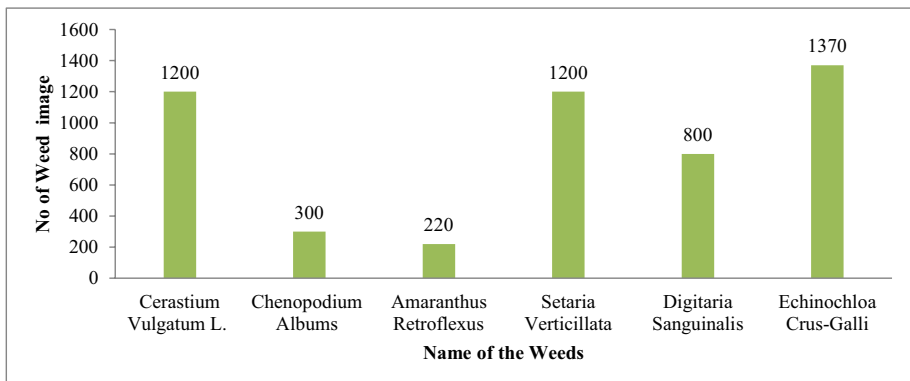

Fig. 1 Dataset description

and classification. To ensure that the output is continuous as the pixel gets closer to a tile center, the interpolation coefficients represent the locations of pixels between the nearest tile center pixels. The complete flow of data is given in Fig. 2.

This work has enhanced the quality of the image using hologram Eq. 1

$$q_n = \frac{\text{No of the pixel of the image with in tensity n}}{\text{total no of } a \text{ pixel of an image}} \quad (where\ n = 0, 1, 2, \ldots \ldots m - 1)$$

$$(1)$$

Let H_{RGB} be a given image, which can enhance the quality of the image based on q_n hologram equation. The 'q_n' has two parameters no of the pixel of the image with the intensity 'n' and another is the total no of the pixel of the image. The 'm' is the possible intensity value up to 0 to 255. Let 'q' be a normalized hologram of 'g'. This hologram equation can be defined as Eq. 2.

$$h_{ij,j} = floor((m - 1) \sum_{n=0}^{g_{i,j}} p_n))$$

$$(2)$$

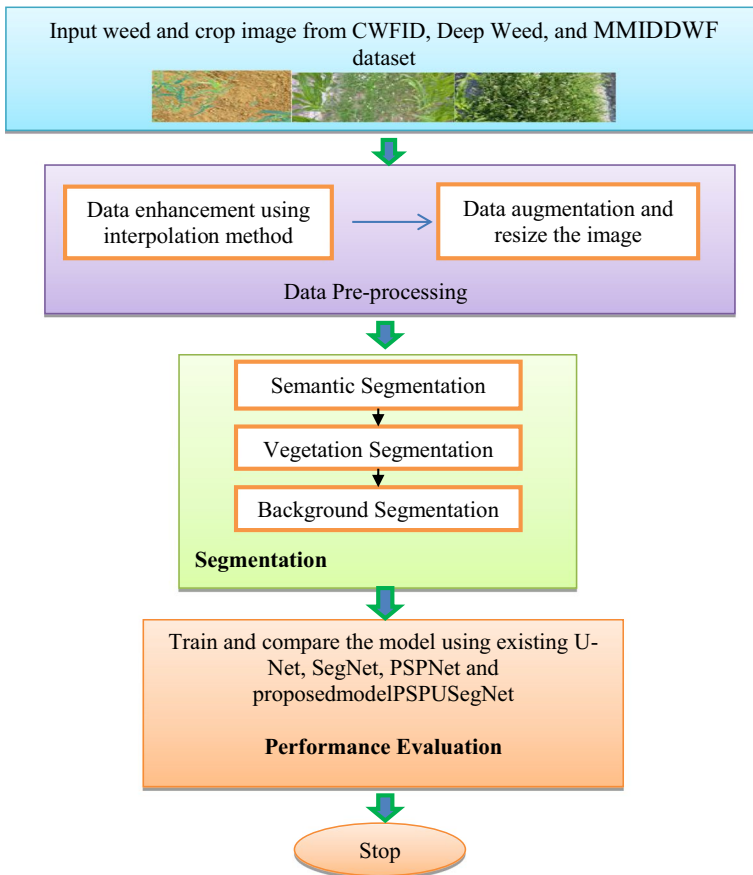**Fig. 2** Complete data flow of model

The floor function rounds down/up to the nearest integer to the transform Eq. 3.

$$T(k) = floor((m-1)\sum_{n=0}^{k}P_n)) \qquad (3)$$

This equation has been imported from Eq. 4.

$$z = T(y) = (m-1)\int_{0}^{y} p_y(y)dy \qquad (4)$$

where p_y is the Probability Density Function (PDF) of, 'T' is the distribution function of y. Assume T is invertible and differentiable, y multiplied $(L-1)$ which is defined as in Eq. 5.

$$p_z(z) = \frac{1}{L-1} = 1 \qquad (5)$$

This equation is defined as a high-density pixel.

$$f(x,y) = T(f(x,y)+k) \qquad (6)$$

where $f(x,y)$ is thecoordination of x and y axis value, k is the constant value it will be 0 to 255.

The approximation of weed and crop image $pX(x)$ are illustrated transformation in Eqs. 1 and 2. Although the histograms produced by the discrete version won't be completely flat, this work will be flattened, which will improve the contrast of the image. The picture improvements took an average of 15 min. Enhancing the quality of the weed image technique is given in Fig. 3.

The weed image 'Chena podium Album L' has a blur; it has enhanced the quality of the image using a hologram transform equation. Additionally, the function $f(x,y)$ is the pixel coordination, which may improve the value from 0 to 255 in color vision and the pixels exhibited in the blue, green, and red shaded areas.The function $f(x,y)+k$, which increases the intensity of pixels in red, blue, and green shaded pixels, 'k' is used as a constant to set the value of color vision.

4.2 Overlapping plant leaves and density estimation of weeds

Generally, most of the different varieties of the plant germinate in the field. This study used a 'Vignamungo' plant field image with seven different classes of weed images. All
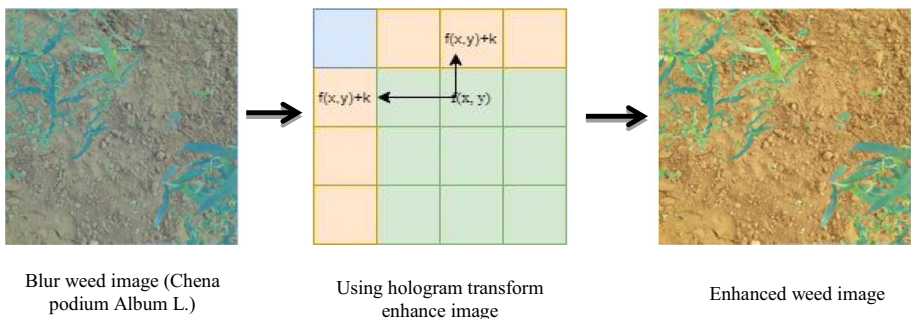

Blur weed image (Chena podium Album L.)

Using hologram transform enhance image

Enhanced weed image

Fig. 3 Enhance weed image

these classes have overlapped weed images. A sample of some overlapped weed plants is given in Fig. 4. Most weed leaves are overlapped, which has decreased the performance of the classifier. Tile classification is a sophisticated technique for identifying weeds and crop plants. This work uses $25 \times 25, 50 \times 50, 75 \times 75$, and 100×100 sizes of tile for calculating the overlapped weed image. The weed density is calculated based on weed-infested regions. The weed-infested region is identified by tile classification, which can be calculated by vegetation coverage in each region. In this work, the weed density has been calculated as Weed Cluster Rete (WCR) [24], as defined in Eq. 7.

$$\text{WCR/Weed density} = \frac{\text{Weed plant coverage in tile}}{\text{the total area covered in the region}} \tag{7}$$

This density estimate will help in selecting suitable areas for weeding and herbicides in the field. Some overlapped images are given in Fig. 4.

4.3 Weed/ crop image data segmentation

Enhanced 5090 images are used as input for the pipeline. For the segmentation, images are grouped into three clusters. First is semantic segmentation for homogeneous weed object; second is background segmentation for discrimination of object, and third is vegetation segmentation for foreground segmentation of object. The semantic segmentation creates homogeneous target object with the same pixel intensity. For the discrimination of object, there are two other segments, such as vegetation and background segmentation [25]. This segmentation technique creates the vegetation mask and mask object, which may be weed leaves or crop leaves. The complete process has been done using tile classification. The tile has been generated in the Region of Mask (RoM). The complete segmentation has over-used vegetation, semantic, and background segmentation techniques. The detailed descriptions are given in the next subsection.

4.3.1 Vegetation segmentation of the object

After the pre-processing of an image, image segmentation is the next specific task for discriminating weeds and crop plants from field image data. The vegetation segmentation is the foreground of the specific object. These object can discriminate between the overlapped weed image and the location estimation of the object. When the picture mask is applied, the only pixels that appear in the vegetation are those that are not zero. Following binary image segmentation, a particular plant or weed is displayed in different colors of the
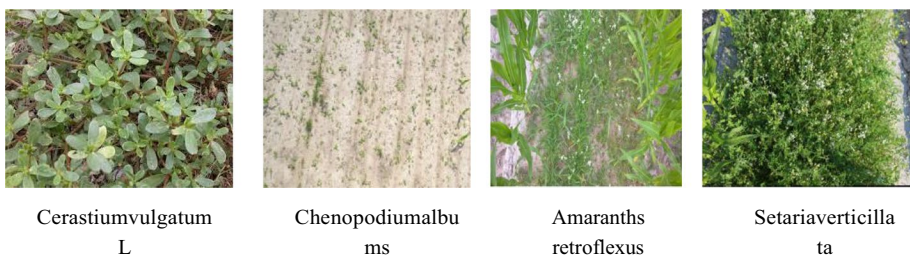

Cerastiumvulgatum L	Chenopodiumalbu ms	Amaranths retroflexus	Setariaverticilla ta

Fig. 4 Overlapped weed plant leaves

image, and individual plants should be segmented [26]. This particular task is challenging because weeds and crop plants grow together. Sometimes weeds and crop leaves overlap. The vegetation segmentation can also include information such as the growth stage of the weed or plant, leaf count; stem position, biomass amount, and others. Furthermore, it can also calculate the plant coverage ratio in the field, the interspacing of plants, and the count of plants in the field. Some weed vegetation segmentation is given in Fig. 5.

4.3.2 Background segmentation of object

The vegetation segmentation Foreground segmentation can discriminate a specific object. Our system's initial stage is foreground–background segmentation, which takes into account the difference between the actual picture and a background model. Foreground refers to areas where the observed picture and the backdrop model differ considerably. The background image has a different frequency of pixels; it may be a high- or low-density pixel. A collection of photos of the empty working space is usually used to create the backdrop model. Because the same model is used for consecutive photos, background removal only works for static backgrounds. It has a high-density pixel object [27]. The background segmentation includes high- and low-density pixels of the complete object.

4.3.3 Semantic segmentation of object

Semantic segmentation is the process of assigning a label to each pixel in an image. This contrasts with classification, which gives the entire image a single label. Semantic segmentation treats many object belonging to the same class as a single entity. These techniques
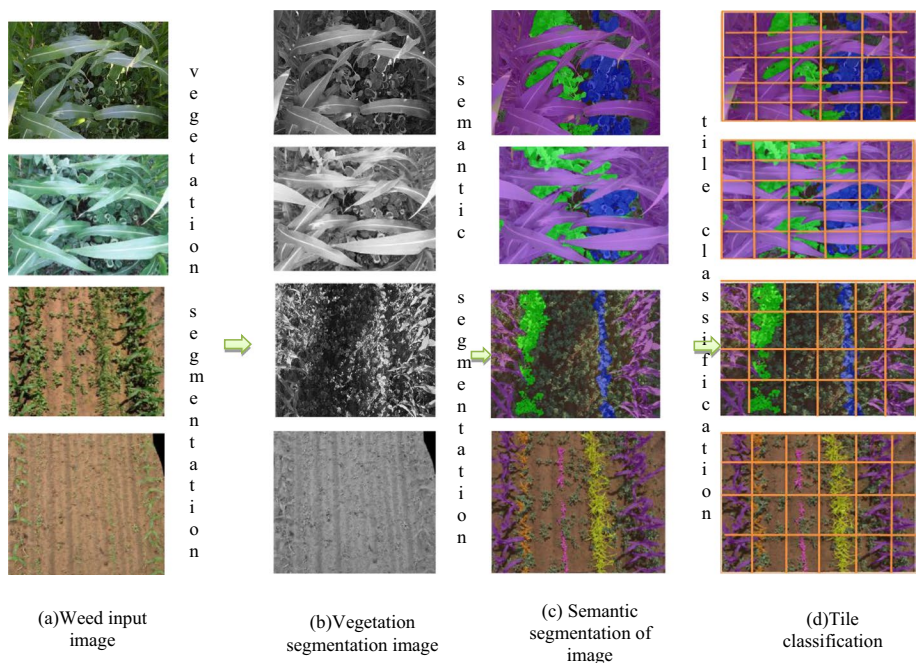

(a)Weed input image

(b)Vegetation segmentation image

(c) Semantic segmentation of image

(d)Tile classification

Fig. 5 Weed and crop plant segmentation

create an inhomogeneous color for the weed or crop object, which has helped to identify the weed or crop object. There are some weed and crop object given in Fig. 5.

Figure 5 includes some different categories of images, such as vegetation-segmented and semantic-segmented images. The vegetation segmentation image includes foreground object with high density, and these object have the same density pixel using semantic segmentation [28]. The object has been identified using tile classification. The tile includes high-density pixels. After that, pixels are put on a future vector for feature extraction.

4.3.4 Tile classification of the object

Further, more input weed image data has been taken from the 'Deep Weed', 'CWFID', and 'MMIDDWF' datasets and acquired as black gram field images. The concept of inputting any single weed image (H_{RGB}) has represented the image. The object has been identified by the vegetation mask in the 'Deep Weed', 'CWFID' and 'MMIDDWF' datasets and acquired as black gram field images. The concept of inputting any single weed image (H_{RGB}) has represented the image. The object has been identified by the vegetation mask (H_{veg}), which has been generated by and applied by It has achieved a Region of Concern (RoC), which is denoted as an object. Furthermore, the masked image (H_{masked}) is distributed as small tiles (H_{tile}), 'Deep Weed', 'CWFID', and 'MMIDDWF' datasets and acquired as black gram field images. The concept of inputting any single weed image (H_{RGB}) has represented the image. The object has been identified by the vegetation mask (H_{veg}), which has been generated by and applied by It has achieved a Region of Concern (RoC), which is denoted as an object. Furthermore, the masked image (H_{masked}) is distributed as small tiles (H_{tile}), and often the patches are square tiles. It may be 25×25 (px), 50×50 (px), or 75×75 (px). The term tile (H_{tile}) denotes the morphological characteristics of weeds taken from and in possession of the vegetation pixels at any given time in the image (H_{tile}). Additionally, the resulting scores are used to categorize plants as either weeds or crops. A binary classifier is used to categorize these plants (crops and weeds). Utilizing the vegetation segmentation approach for classification, weed, and crop density performance measurements have been completed [29]. There are a few abbreviations used in the algorithm (OWID) given in Table 1.

The steps of the proposed Overlapped Weed/Crop Image Data (OWID) algorithm are given in Algorithm 1 and Algorithm 2 and Fig. 6.

Applying segmentation based on CNN, create the vegetation mask (H_{veg}) from the picture (H_{RGB}), which is taken from a common data store. This segmentation is overlaid H_{RGB} with H_{veg} to get H_{masked} it has divided the image into smaller regions (square tiles). Furthermore, classify it into crop, weed, or background of the image. The high-density pixel is put on a feature vector with a threshold value of 2700 pixels and checked over-segmentation. Further, segment the object as used for calculation.

4.4 Data classification using the proposed model

This work has trained three existing CNN models, such as UNet, SegNet, USegNet, and the proposed model PSPUSegNet. The learning rates are slow in UNet, SegNet, and USegNet CNN due to the deeper intermediate layers. The proposed PSPUSegNetmodel has been ignored over the deeper intermediate layer. This work solves this problem by offering a global prior representation that is both effective and efficient, which is discussed in the next subsection.

Table 1 Abbreviations used in algorithm 1 (OWID)

Notation	Meaning
H_{RGB}	Input colored weed image of weed
H_{veg}	Vegetation mask
H_{masked}	Fragmented image
H_{tile}	Square tiles
VCRE	Vegetation Cluster Rate Estimation
S_h	The complete height of Weed
O_{wd}, F_V	Segmented weed object, Feature Vector
$M_p(indices)$	Max pool indices, exponent variable of softmax
CE, and BCE	Cross-entropy and Binary Cross-Entropy

4.4.1 PSPUSegNet(Pyramid Scene Parsing Network USegNet)

The proposed PSPUSegNet model has included the functionality of the PSPNet, UNet, and SegNet models. It has a total of 83 convo layers, which include 25 convo layers from PSP-Net, 16 convo upsampling layers from UNet, and the remaining 26 convo downsampling layers from the SegNet CNN model. The proposed model includes input, convolutional, softmax, up-sampling, and a max pool layer. Further, the three max pool layers out of the five layers by up-sampling the layers of the pyramid finally, softmax layers will generate the final result of image classification. This work has 83 Conv layers, 5 max pools layers, and 5 up-sampling layers applied in the hybrid SegNet CNN model [30]. After pre-processing the image ($w \times 3$), it has input for the proposed CNN model. The morphological feature map of weeds in an image has been achieved by the proposed model. The scale of the image feature map has been reduced using Max, the pooling layer, and the up-sampling process. The final result has been shown after processing the soft-max layer into pixel-wise data representations of each class and creating the pyramid. The proposed model shows a "U" shape [19]. Initially, UNet was invented for biological image segmentation, but it has also achieved high performance in other industries.

There are two main reasons for the use of this UNet and SegNetCNN model. Firstly, it can extract exhaustive features from local information through convolution layers. Secondly, it will provide the best accuracy for the limited number of samples. The classical UNet and SegNet models had a large consumption of calculation resources and a slow speed; therefore, the proposed model has simplified these factors. This work is very similar to the SegNet CNN model for image segmentation using the skip connection method. The skip connection method has been lost using up-sampling of the bottom layer in the SegNetmodel. The classical SegNet model is the skeleton of the proposed model. There is more time consumption for pooling in the basic SegNet model. Therefore, it's mandatory to reduce the number of pooling layers at first. This work has been performed by the Skip Connection Technique (SCT) in the SegNet CNN model. This technique arranged spatial information at the same level after using the up-sampling bottom layer. Batch normalization (BN) was added in the final stage of the convolutional layer to guarantee data stability [31].

This paper has proposed a PSPUSegNetmodel with a skip connection method and a unified kernel size (3, 3) for the convolution layer. This work has used kernel size, padding, and activation functions. There are 3 kernel sizes; for padding, use 0 in the outer ring of the
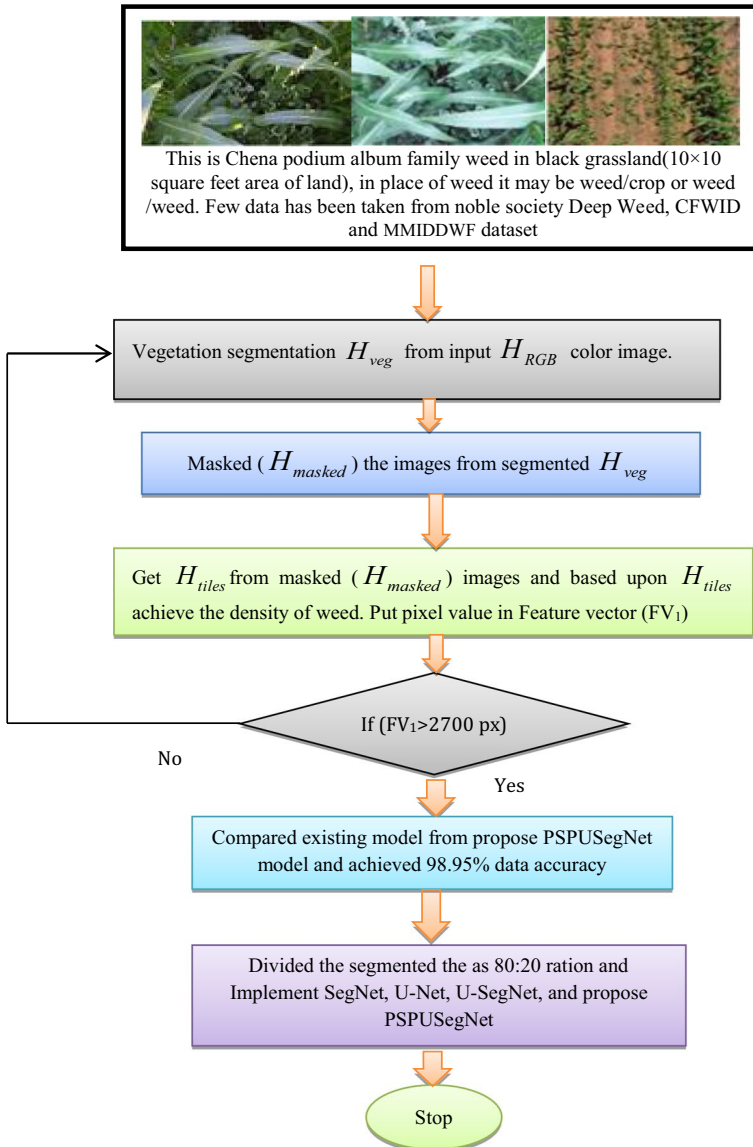
Fig. 6 Flow chart of the proposed model

image. The ReLu activation function used 0, Conv 64, and ConV128 masks, and finally, the kernel size of the outer layer is (1, 1). Furthermore, the sigmoid function handles the binary (0~1) image segmentation problem. The complete steps of the proposed model are described in Fig. 7.

Successful non-trivial semantic segmentation object detection. In this work, the proposed model has changed the three max-pool layers out of 5 layers in a new framework of semantic segmentation. The max-pool information is proceeding before forward to

Algorithm 1 Estimation of Overlapped Weed/Crop Image Data (OWID)

Input: Take H_{RGB} the field acquired from the dataset.

Output: Overlapped Weed object (O_{wd}) and density $(W_{density})$

- Given (H_{RGB}), Generate the vegetation mask (H_{veg}) using CNN-based segmentation;
- overlay H_{RGB} With H_{veg} to get H_{masked};
- divide the image H_{masked} into smaller regions H_{tile} (square tiles);

  ❖ for (H_{tile} in H_{masked}) do

  Classify H_{tile} into crop, weed, or background of image;

Classify high-density pixels and put them on Feature Vector-1 (FV_1)

- Set the threshold value as 2700 pixels.

  ❖ *If (pixel value > 2700)*

Print →over-segmentation

else

- Compare each pixel from the object and segment the object as (O_{wd})
- Return→ O_{wd}

- estimate $W_{density}$ where $W_{density} = \dfrac{Weed\ plant\ coverage\ in\ region(O_{wd})}{Total\ land\ area}$

- Return→ $W_{density}$

- end

---

the next stage and finally third is before executing the semantic segmentation of the weed object to explore contextual information [32]. Overall changes are improving the flow and accurately achieving the object of the image. A detailed description is given below in Fig. 8.

Here 5 different max pool networks are closely related to Region Proposal Network (RPN) and CNN feature (G, T). The RPN can parallel predict the object in semantic and vegetation-segmented object. Here C1, C2, and C3 have predicted masks of object, and N1, N2, and N3 are bounding boxes. The bounding box and predicted mask have been shown below in Eqs. 8 and 9.

$$y_t^{box} = q(y, s_{t-1}), s_t = C_t(y_t^{box}) \tag{8}$$

$$y_t^{mask} = q(y, s_{t-1}), N_t = N_t(y_t^{mask}) \tag{9}$$

where y is the backbone feature of the CNN feature y_t^{box} and y_t^{mask} is donated as a bounding box and predicted mask feature. C_t, N_t is the box and mask head and t is a stage, and s_t, y_t is the predicted box and mask head.

Algorithm 2 Execute Overlap Weed Data (EOWD)

Input: Overlapped Weed object (O_{wd}) and density $(W_{density})$

Output: MAE, IoU, MIoU

The input O_{wd} object is divided into 80:20 ratios of data.

- The pre-processed image on a modified max-pool layer of PSPUSegNetand updated max-pool layer

Max pool, int l =256; Update last 3 max-pool layer of

- Max Pool -Layer PSPUSegNet

Input $C64(3\times3\times256)$ and remove one skip connection

Print➔ $C64(3\times3\times255)$

- Input O_{wd} and remove one layer skip connection and print: $C64(3\times3\times254)$

- Input $C64(3\times3\times254)$ and remove one layer skip connection $C64(3\times3\times256)$

- If (accuracy ofPSPUSegNet> accuracy of UNet, SegNet, and USegNet)
      Print➔ highest accuracy, IoU, MIoU.
            elseif implement UNet
            elseifSegNet

            elseUSegnet

print➔ highest accuracy, IoU, MIoU.

- Stop

4.4.2 Interleave execution of weed image

Processing of weed image object as two branches of bounding boxes in parallel execution in training stage (Eq. 1) and both two branches are not directly interacted within a stage. So it is mandatory to improve the architecture at N_{t-1} head. The interleaved execution and mask information flow is expressed as Eq. 10 and 11.

$$y^{box}_{\,t} = \eta(y_1 y_{t-1}), s_t = C_t(y_t^{\,box}) \tag{10}$$

$$y_t^{\,mask} = \eta(y, s_{t-1}), N_t = N_t(y_t^{mask}, N_{t-1}) \tag{11}$$

where N_{t-1} the intermediate object is a feature and $t-1$ is a stage of mask representation.

4.4.3 Object detection flow of weed image data

Weed object detected using Region of Interest (RoI) future and it has been implemented before the de-convolutional of data with the spatial size is 14×14. In stage' forwarded all the mask headswith the use of RoI_s and finally computed the masked object. Here 'F' is a function thatcombinesthefeatures of the current stage and here $N_t(F(y_t^{mask}, N_{t-1})$ is a feature transformation function with four 3×3 convolutional layers. Furthermore, N_1, N_2
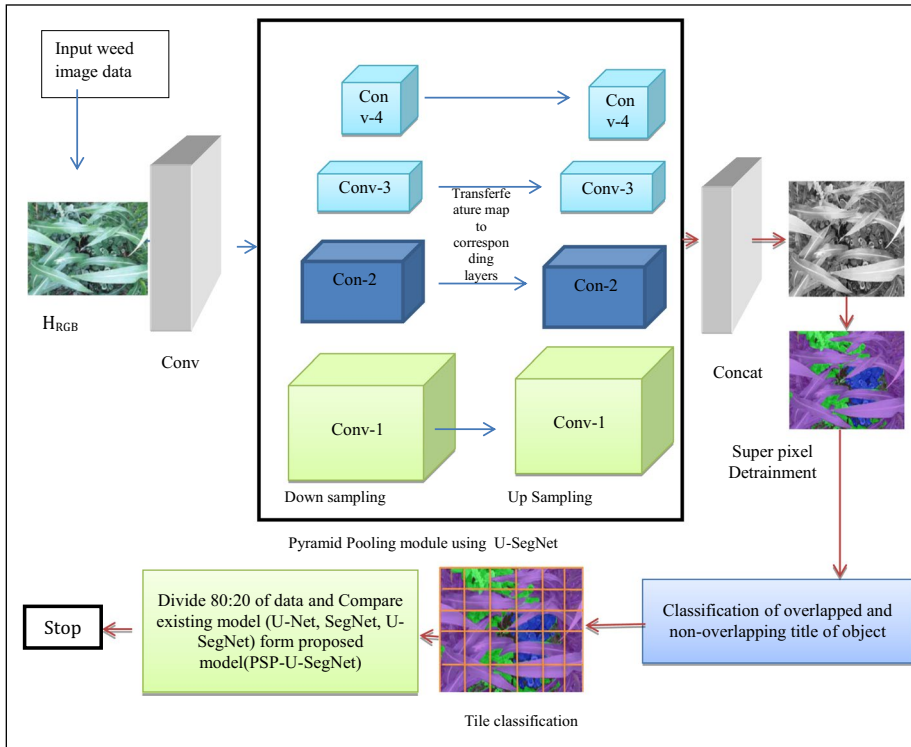
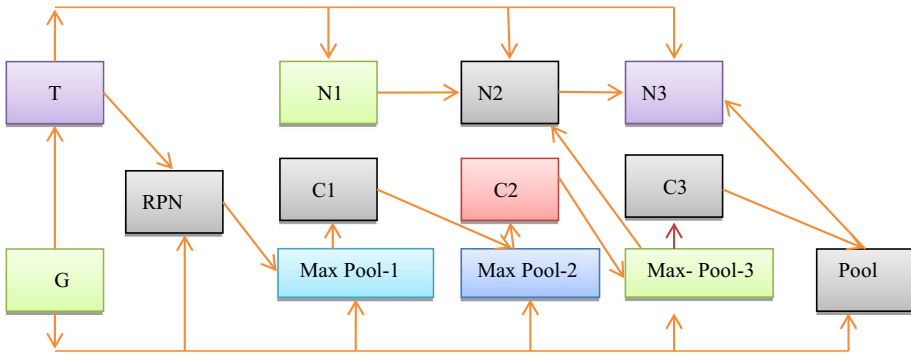**Fig. 7** Proposed PSP U-SegNet CNN model

Fig. 8 Modified Max pool network architecture of PSPUSegNet

, N_{t-1} are feature transformation with different mask such as y_t^{mask} and h_t is feature vector use for processing the binary classification of the data. Finally the mask object is computed as $N_t\left(F\left(y_t^{mask}, N_{t-2}\right)\right)$. Theobjection detection has been done through the backpropagation technique in Eq. 12.

$$y_t^{mask} = q(y, s_{t-1}), N_t = N_t(F(y_t^{mask}, N_{t-1})$$
$$F(y_t^{mask}, N_{t-1}) = y_t^{mask} + h_t(N_{t-1})$$
$$N_1 = N_1(y_t^{mask})$$
$$N_2 = N_2(F(y_t^{mask}, N_1)$$
$$\vdots$$
$$N_{t-1} = N_t(F(y_t^{mask}, N_{t-2}))$$

(12)

This work has been directly combined with Mask R-CNN and Cascade R-CNN, which is denoted as a hybrid cascade mask R-CNN.

4.4.4 Learning the weed object using the proposed model

This work presented the PSPUSegNetfor semantic segmentation of weed and crop pictures. Figure 8 shows the different boxes and masks that have to interact with different branches. This work uses RoI align, such as 7×7 and 14×14 feature maps. Each stage is predicted by the box head, and the entire mask head has been predicted as the pixel-wise mask. The loss function takes the form of multi-task learning given in Eqs. 13, 14, 15, and 16.

$$M = \sum_{t=1}^{T} \beta(M_{cbox}^t + M_{mask}^t) + \lambda M_{seg}$$

(13)

$$M_{cbox}^t(d_i, s_t, \hat{d_t}, \hat{s_t}) = M_{cls}(d_t, \hat{d_t}) + M_{reg}(s_t, \hat{s_t})$$

(14)

$$M_{mask}^t(n_t, \hat{n_t}) = BCE(n_t, \hat{n_t})$$

(15)

$$M_{seg} = CE(t, \hat{t})$$

(16)

Here M_{cbox}^t cover the loss of the bounding box which has been predicted as the stage of t, and it has to combine as M_{cls} and M_{reg} which is defined as weed classification and bounding box regression. M_{mask}^t is denoted as a prediction mask in any stage of 't' which is called the Binary Cross Entropy (BCE). M_{seg} is used to balance the phases and tasks of segmentation. It is designated as semantic segmentation loss in the concept of cross-entropy. This work is used by default $\beta = [1, 0.4, 0.24]$, $\lambda = 1$ and $t = 2$.

5 Result and discussion

This work has taken 5090 pieces of data from various datasets, such as 'Deep Weed', 'CWFID', and 'MMIDDWF' datasets, distributed in 80 : 20 ratios. The complete distribution of the dataset is given in Table 2.

5.1 Qualitative performance of vegetation segmentation of the model

As a result of vegetation segmentation using a few input images from three different datasets, it can be observed that PSPUSegNetoutperformed the other model. After the discrimination of object using vegetation segmentation, semantic segmentation prepares the

Table 2 Dataset distribution

Name of Dataset	Total	Training image	Testing Image
Deep Weed	5090	1600	120
CWFID		1800	200
MMIDDWF		1200	170

homogeneous color object model with the same pixel intensity as the object. In observation, the 'Deep Weed' dataset has provided finer object detection. The background and vegetation segmentation can provide finer detail on the vegetation of the object.

It is also interesting that UNet can identify tiny groupings of vegetation objects. Further, classify it as a single pixel of the object. This is because it prioritizes the spatial continuity of vegetation clusters, whereas UNet tends to focus on a pixel's immediate surroundings. The CWFID dataset, which has weak contrast when compared to the MMIDDWF, showed a considerably stronger trend. The "Deep Weed" has a more prominent dataset using the PSPUSegNetclassifier. The quantitative evolution has evaluated using UNet, SegNet, USegNet and PSPUSegNet. It has given in Table 3.

The proposed model PSPUSegNet has provided 0.961% discrimination from vegetation and background segmentation of object from an image. The other existing classifier, USegNet, provides 0.92%, and SegNet has 0.91% for the MMIDDWF dataset.

5.2 Feature vector-based tile classification and effect of tile

As previously mentioned, the vegetation segmentation H_{veg} is used to detect the areas of vegetation in the pictures that contain crops and weeds. The output is a masked picture created by overlaying the input image H_{RGB} with H_{veg} then, non-overlapping tiles (sub-images) and titles are separated from this masked picture. A pre-trained UNet classifier is then used to retrieve the characteristics of each title. Table 4 shows how well various classifiers perform when identifying terms such as "weed" or "crop" using these attributes.

Take note of the enhancement in classifier performance brought on by weighted training utilizing various methods. By showing how sampling strategies (random sampling) aid in enhancing the classifier's performance for an imbalanced dataset, this study supports prior

Table 3 Quantitative evolution of data

Model Name	Dataset	mIoU
UNet	Deep Weed	0.911
	CWFID	0.872
	MMIDDWF	0.902
SegNet	Deep Weed	0.922
	CWFID	0.772
	MMIDDWF	0.913
USegNet	Deep Weed	0.934
	CWFID	0.741
	MMIDDWF	0.921
PSPUSegNet	Deep Weed	0.961
	CWFID	0.802
	MMIDDWF	0.943

Table 4 Pixel-wise data segmentation

Dataset	Classifier	25 × 25 (px)			50 × 50 (px)			75 × 75 (px)		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Deep Weed	UNet	0.25	0.13	0.15	0.41	0.33	0.02	0.36	0.49	0.55
	SegNet	0.11	0.03	0.04	0.43	0.02	0.04	0.34	0.33	0.35
	USegNet	0.34	0.33	0.35	0.22	0.07	0.11	0.34	0.33	0.35
	PSPUSegNet	0.31	0.16	0.21	0.43	0.07	0.28	0.34	0.43	0.38
CWFID	UNet	0.31	0.13	0.18	0.39	0.22	0.28	0.34	0.43	0.38
	SegNet	0.16	1	0.28	0	0	-	0.35	0.24	0.29
	USegNet	0.32	0.3	0.31	0.3	0.52	0.38	0.42	0.21	0.28
	PSPUSegNet	0.29	0.07	0.11	0.75	0.07	0.36	0.36	0.16	0.22
MMIDDWF	UNet	0.35	0.36	0.36	0.42	0.21	0.28	0.37	0.14	0.2
	SegNet	0.32	0.5	39	0.27	0.42	0.33	0.31	0.29	0.3
	USegNet	0.32	0.16	0.21	0.28	0.09	0.14	0	0	-
	PSPUSegNet	0.44	0.13	0.2	0.34	0.39	0.36	0.29	0.67	0.41

findings. The accuracy and recall computed for the weed class on the test set are used to gauge performance. While sampling methods that account for class imbalance result in a relative improvement in the accuracy and recall values, the absolute values still fall below the acceptable cut-off. As shown in Table 8, the suggested model PSPUSegNetclassifier obtained an accuracy of 98.96% and a recall of 97.98% using the Deep Weed dataset.

Every tile was expected to be covered in weeds. This highlights how these classifiers are unable to reliably distinguish between feature vectors produced by the suggested pipeline that correspond to agricultural and weed plants. Two observations served as the basis for the intuitive choice of tile size (a square with a side of 50 pixels) as primarily for either weeds or agricultural plants rather than both, and (2) it prevented the creation of zones where virtually all of the pixels belonged to a cluster of vegetation. Due to how similar crop and weed plants would seem, there would not be sufficient descriptive information for the classifier to differentiate between them.

Nevertheless, the outcomes from regions of varied sizes were examined to justify the choice of tile size. This study used both side length increases and decreases (75 (px) and 25 (px), respectively) to retrain the classification models. Classifiers trained using tiles of side lengths 25 (px) and 50 (px) perform better than those trained on tiles on average, taking into account both accuracy and recall values. Further, Table 5 shows the computation time by passing the tile processing. For patch sizes with sides of 50 and 100 pixels, computation time is comparable, while side lengths of 25 pixels result in a considerable increase in computation time. The explanation was that patches with sides longer than 25 pixels had a significantly higher percentage of tiles with vegetation pixel density than 10% compared to the previous two. Figure 9 has shown the vegetation mask of weed image.

5.3 Comparison of pixel-wise dense predictions

The patch-wise predictions may be utilized to provide accurate pixel-wise weed and crop segmentation, even though that is not the suggested method's main goal. Therefore, compare the anticipated ground coverage's accuracy using the F1 score measure (Eq. 8). End-to-end segmentation networks were suggested by the authors of [21] and [22] for predicting dense crop/weed maps on the Deep Weed, CWFID, and MMIDDWF datasets. The maximum-minimum value for the class of weeds is (0.41, 0.43) in the deep weed dataset in tile classification. The CWFID and MMIDDWF weed classes have 0.39 and 0.75 and 0.42 and 0.34 precision values, respectively. In observation, the CWFID dataset is more accurate than the other dataset. Another parameter, the F1-score, has been reported as a maximum of 0.28, 0.36, and 0.28 for the Deep Weed, CWFID, and MMIDDWF datasets, respectively. Our method falls short in terms of pixel-level precision in comparison (the maximum F1 value for the weed class is 0.36) in the CWFID dataset. The complete pixel data segmentation is given in Table 4.

However, a method to choose particular regions must be added to the segmentation networks to selectively treat specified parts. There will inevitably be an overlap of weed and crop pixels for the majority of the tiles if they are separated into sections like square tiles. The dominant label for such tiles will be used to determine how to handle a certain area.

Table 5 The computation time of tile processing	Tile size length	75×75 (px)	5×50 (px)	(25×25) px
	computations time	0.94	0.9	6.54

(a) Deep Weed

(b) CWFID

(C) MMIDDWF

| (I) Original weed Image | (II) Ground truth invegetation mask | (III) Vegetation mask predicted by U-SegNet | (IV) Vegetation mask predicted by the PSP U-SegNet |

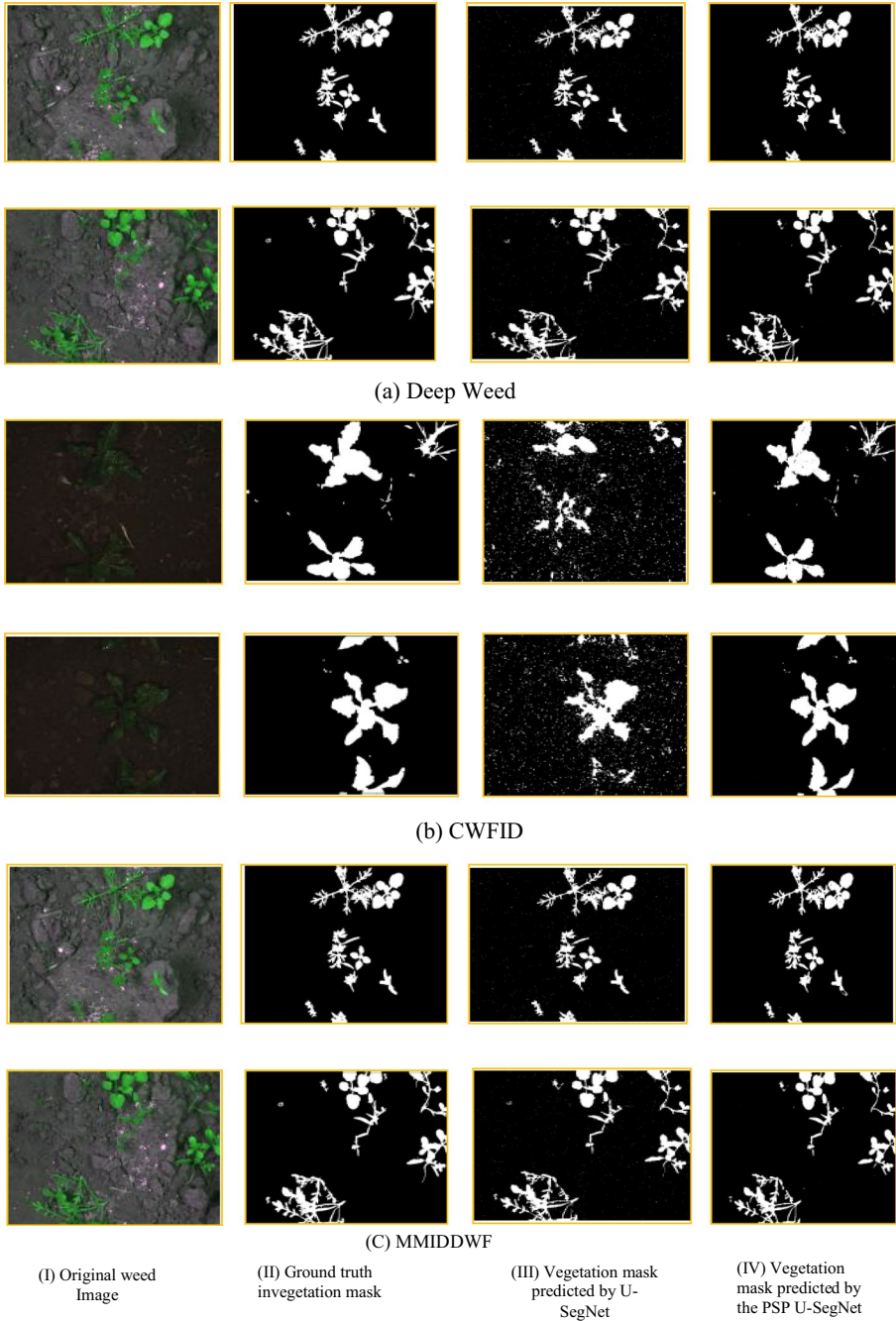**Fig. 9** Vegetation mask of weed image (left to right)

Table 6 Value loss using Cross-Entropy

Dataset	Binary Cross Entropy Loss						Weight cross Entropy Loss					
	PR0	RE0	F1 S0	PR1	RE1	F1 S1	PR0	RE0	F1 S0	PR1	RE1	F1 S1
DEEP EWWD	0.5	0.6	0.79	0.2	0.7	0.44	0.8	0.5	0.74	0.31	0.91	0.46
CWFID	0.7	0.6	0.80	0.3	0.8	0.44	0.9	0.6	0.76	0.32	0.92	0.47
MMIDDWF	0.9	0.8	0.82	0.4	0.9	0.46	1	0.7	0.77	0.34	0.93	0.49

As a result, the selective treatment is unaffected by correctly recognized pixels that are in the minority for a specific tile. The computation time for tile processing is given in Table 5.

This work contends that the suggested method places more emphasis on accurately identifying the treatment regions than it does on correctly identifying such pixels. Additionally, the enormous data needs of the suggested technique are far lower than those of an end-to-end segmentation network, which enhances generalization and scalability. The suggested method may also be applied to any crop-weed combination because it does not require the creation of custom features [31]. The value loss via cross-entropy is displayed in Table 6.
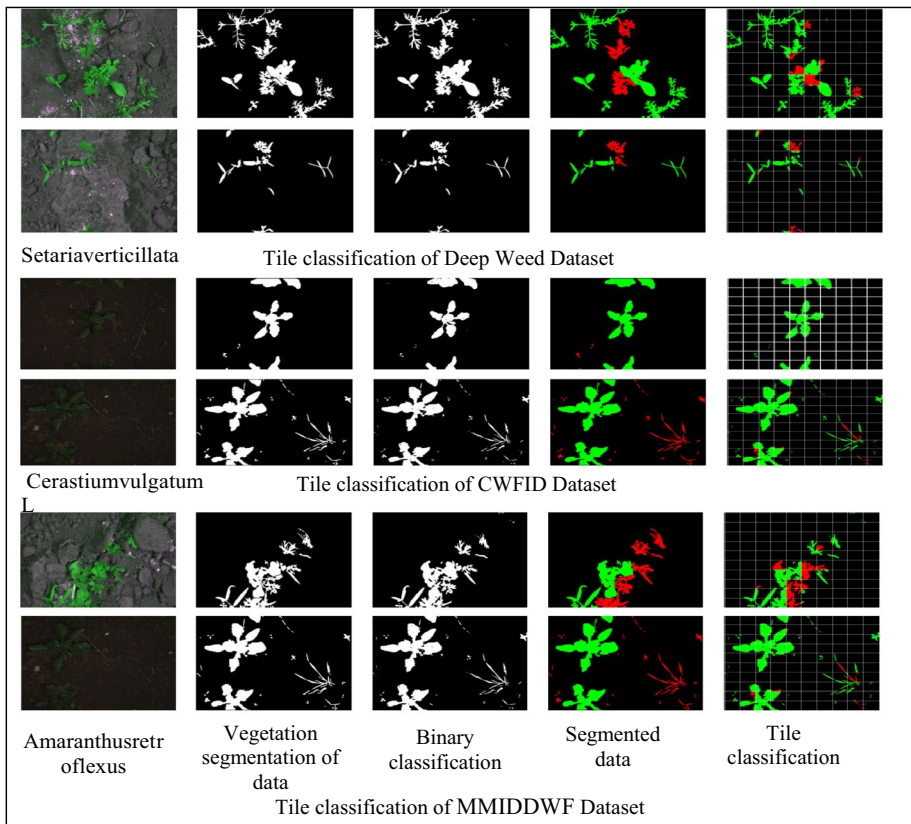

Fig. 10 Vegetation mask of data

Table 7 The error rate estimation of image data

Dataset	Mean Accuracy	MAE	RMSE
Deep Weed	82.13	1.62	2.06
CWFID	75.24	5.02	7.85
MMIDDWF	82.13	1.62	3.06

The soft-max layer of the proposed PSPUSegNetmodel has checked the cross-entropy and weight-cross entropy loss of images. This work used three datasets as Deep Weed, CWFID, and MMIDDWF. Out of these, Deep Weed has a precision of 0.5, and in the case of weight cross-entropy, the minimum precision is 0.8. The CWFID and MMID-DWF have maximum value losses. The vegetation mask of three different dataset has shown in Fig. 10.

This work has estimated the weed object based on tile classification of 'Amaranthusret-roflexus' weed image data. After analyzing the vegetation segmentation and binary classification, the data has been classified as a gray-scale image. For keen observation of the object, it has been segmented using tile classification. The classified object may be overlapped, and the partial or full object may be detected. The detected object is estimated by the error rate, which is given in Table 7.

Table 7 summarizes the error rate based on MA, MAE, and Root Mean Square Error (RMSE) of the Deep Weed, CWFID, and MMIDDWF datasets. After observation, the Deep Weed dataset has a lower MA, which is 82.13 and 1.62, and 2.06 error rates for MAE and RMSE. The performance of the model is given in Table 8.

In the Deep Weed dataset, the proposed model has achieved 96.98%, 97.98%, and 98.96% precision, recall, and data accuracy, respectively, using the proposed model PSPUSegNet. The existing model UNet classifier has achieved 89.93%, 90.90%, and 84.23% data accuracy. Another existing CNN model (UNet, SegNet, and USegNet) has achieved 90.98%, 93.87%, and 85.45% data accuracy, which is less accurate than the proposed model.

Table 8 Performance of Existing model to the proposed model

Dataset	Classifier	Precision	Recall	F1-Score	Accuracy
Deep Weed	UNet	86.97	87.45	87.97	89.93
	SegNet	87.65	86.54	88.90	90.98
	USegNet	89.76	90.09	90.56	91.90
	PSPUSegNet	96.98	97.98	97.88	98.96
CWFID	UNet	84.97	85.45	85.97	90.90
	SegNet	88.65	87.54	87.90	93.87
	USegNet	91.76	89.99	90.56	95.97
	PSPUSegNet	96.98	95.98	95.88	97.99
MMIDDWF	UNet	88.97	87.45	85.97	84.23
	SegNet	89.65	88.54	89.90	85.45
	USegNet	89.79	90.09	90.56	93.67
	PSPUSegNet	90.98	92.98	91.88	93.98

6 Conclusion

In the environment, agrochemicals like weedicides are an expensive input for farming. It may be possible to drastically lower their usage by using a computer vision system to locate areas that need specific chemical treatment. To support precision agriculture, a PSPUSegNettechnique to robustly predict weed density and dispersion is provided. The suggested method only accepts color images as input. The first step is to construct a binary vegetation mask by removing every background pixel. Precision agriculture is an approach to agricultural management that tries to gradually increase yield and revenue. In addition to being harmful to the environment, agrochemicals like weedicides are an expensive input for farming. It might be possible to drastically reduce their use by using a computer vision system to locate areas that need specific chemical treatment.

A PSPUSegNetapproach to accurately estimating weed density and dispersion is offered to enhance precision agriculture. The recommended approach only takes input from color photographs. The self-supervised approach has used as proposed method, in term of segmentation mechanism.This work has used a mixed approach of PSPNet and USegNet CNN models by replacing 7 Conv layers of UNet and 13 Conv layers of SegNet CNN models in downsampling to maintain the global feature of the data. The pooling indices (feature vector) from the encoder feature are transferred for mapping to the corresponding upsampling layer. Making a binary vegetation mask in the first stage entails erasing every backdrop pixel. A maximum recall of 97.98% is used to identify weed-infested areas in the Deep Weed dataset, with an accuracy of 98.96% used to assess their weed density. Reducing reliance on heavily annotated datasets is one of the main goals of our research. The ongoing process of creating vegetation masks is one of our work's constraints. Future research should aim to identify the mix crop weed species and also reduce the average number of iterations required by the unsupervised network to build the vegetation mask.

Data availability Data and source codes are available from the authors upon reasonable request.

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

1. Kazmi W, Garcia-Ruiz FJ, Nielsen J, Rasmussen J, Jørgen Andersen H (2015) Detecting creeping thistle in sugar beet fields using vegetation indices. Comput Electron Agric 112:10–19. https://doi.org/10.1016/j.compag.2015.01.008
2. Lu Y, Young S (2020) A survey of public datasets for computer vision tasks in precision agriculture. Comput Electron Agric 178:105760. https://doi.org/10.1016/j.compag.2020.105760
3. Yasrab R, Zhang J, Smyth P, Pound MP (2021) Predicting plant growth from time-series data using deep learning. Remote Sens 13(3):331. https://doi.org/10.3390/rs13030331
4. Kamath R, Balachandra M, Prabhu S (2020) Crop and weed discrimination using laws' texture masks. Int J Agric Biol Eng 13(1):191–197. https://doi.org/10.25165/j.ijabe.20201301.4920

5. Sharpe SM, Schumann AW, Boyd NS (2019) Detection of Carolina geranium (Geranium carolinianum) growing in competition with strawberry using convolutional neural networks. Weed Sci 67(2):239–245. https://doi.org/10.1017/wsc.2018.66

6. Lottes P, Behley J, Milioto A, Stachniss C (2018) Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. IEEE Robot Autom Lett 3(4):2870–2877. https://doi.org/10.1109/LRA.2018.2846289

7. Qian M, McLaughlin I, Quo W, Dai L (2017) Mismatched training data enhancement for automatic recognition of children's speech using DNN-HMM. https://doi.org/10.1109/ISCSLP.2016.7918386

8. Gao J, French AP, Pound MP, He Y, Pridmore TP, Pieters JG (2020) Deep convolutional neural networks for image-based Convolvulus sepium detection in sugar beet fields. Plant Methods 16(1). https://doi.org/10.1186/s13007-020-00570-z

9. Mishra AM, Gautam V (2021) Weed Species Identification in Different Crops using Precision Weed Management: A Review. Available: https://niti.gov.in/national-strategy. Accessed 17 Apr 2021

10. Muni Mishra A et al (2022) A Deep Learning-Based Novel Approach for Weed Growth Estimation. Intell Autom Soft Comput 31(2):1157–1173. https://doi.org/10.32604/iasc.2022.020174

11. Ma X et al (2019) Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. PLoS One 14(4). https://doi.org/10.1371/journal.pone.0215676

12. Chechliński Ł, Siemiątkowska B, Majewski M (2019) A system for weeds and crops identification—reaching over 10 fps on raspberry pi with the usage of mobilenets, densenet and custom modifications. Sensors 19(17). https://doi.org/10.3390/s19173787 (Switzerland)

13. Rasti P, Ahmad A, Samiei S, Belin E, Rousseau D (2019) Supervised image classification by scattering transform with application toweed detection in culture crops of high density. Remote Sens 11(3). https://doi.org/10.3390/rs11030249

14. Teimouri N, Dyrmann M, Nielsen PR, Mathiassen SK, Somerville GJ, Jørgensen RN (2018) Weed growth stage estimator using deep convolutional neural networks. Sensors 18(5):1–13. https://doi.org/10.3390/s18051580. (Switzerland)

15. Kropff MJ, Lotz LAP, Weaver SE, Bos HJ, Wallinga J, Migo T (1995) A two parameter model for prediction of crop loss by weed competition from early observations of relative leaf area of the weeds. Ann Appl Biol 126(2):329–346. https://doi.org/10.1111/j.1744-7348.1995.tb05370.x

16. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017, 2017:6230–6239. https://doi.org/10.1109/CVPR.2017.660

17. Haug S, Ostermann J (2015) A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. Lect Notes Comput Sci (including Subser. Lect Notes Artif Intell Lect Notes Bioinformatics) 8928:105–116. https://doi.org/10.1007/978-3-319-16220-1_8

18. Xu K, Jiang Z, Liu Q, Xie Q, Zhu Y, Cao W, Ni J (2022) Multi-modal and multi-view image dataset for weeds detection in wheat field. Front Plant Sci 13:936748. https://doi.org/10.3389/fpls.2022.936748

19. Kamath R, Balachandra M, Prabhu S (2019) Raspberry Pi as Visual Sensor Nodes in Precision Agriculture: A Study. IEEE Access 7:45110–45122. https://doi.org/10.1109/ACCESS.2019.2908846

20. Ahmad J et al (2018) Visual features based boosted classification of weeds for real-time selective herbicide sprayer systems. Comput Ind 98:23–33. https://doi.org/10.1016/j.compind.2018.02.005

21. Yasrab R, Zhang J, Smyth P, Pound MP (2021) Predicting plant growth from time-series data using deep learning. Remote Sens 13(3):1–17. https://doi.org/10.3390/rs13030331

22. Hasan ASMM, Sohel F, Diepeveen D, Laga H, Jones MGK (2021) A survey of deep learning techniques for weed detection from images. Comput Electron Agric 184. https://doi.org/10.1016/j.compag.2021.106067

23. Kaur P, Gautam V (n.d.) Plant biotic disease identification and classification based on leaf image: A review. https://doi.org/10.1007/978-981-15-9712-1_51

24. Mahmudul Hasan ASM, Sohel F, Diepeveen D, Laga H, Jones MGK (2022) Weed recognition using deep learning techniques on class-imbalanced imagery. Crop Pasture Sci. https://doi.org/10.1071/CP21626

25. Marwat SK et al (2013) Weeds of wheat crop and their control strategies in Dera Ismail Khan district, Khyber Pakhtun Khwa, Pakistan. Am J Plant Sci 04(01):66–76. https://doi.org/10.4236/ajps.2013.41011

26. Potena C, Nardi D, Pretto A (2017) Fast and accurate crop and weed identification with summarized train sets for precision agriculture. Adv Intell Syst Comput 531:105–121. https://doi.org/10.1007/978-3-319-48036-7_9

27. Shorewala S, Ashfaque A, Sidharth R, Verma U (2021) Weed density and distribution estimation for precision agriculture using semi-supervised learning. IEEE Access 9:27971–27986. https://doi.org/10.1109/ACCESS.2021.3057912

28. Vayssade JA, Jones G, Gée C, Paoli JN (2022) Pixelwise instance segmentation of leaves in dense foliage. Comput Electron Agric 195:106797. https://doi.org/10.1016/J.COMPAG.2022.106797
29. Nachiketh RV, Krishnan A, Krishnan KV, Harikrishnan P, Sasinas Alias Haritha ZA (2021) Southern Pea / Weed Field Image Dataset for Semantic Segmentation and Crop / Weed Classification using an Encoder-Decoder Network. SSRN Electron J. https://doi.org/10.2139/ssrn.3781351
30. Kaur P, Harnal S, Tiwari R, Upadhyay S, Bhatia S, Mashat A (2022) Recognition of Leaf Disease Using Hybrid Convolutional Network by Applying Feature Reduction. Sensors 22(2):575. https://doi.org/10.3390/s22020575
31. Mishra AM, Gautam V (2021) Weed species identification in different crops using precision weed management: A review. CEUR Workshop Proc 2786(February):180–194
32. Sa I et al (2018) WeedNet: Dense semantic weed classification using multispectral images and MAV for smart farming. IEEE Robot Autom Lett 3(1):588–595. https://doi.org/10.1109/LRA.2017.2774979