IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 12, 2021, accepted January 30, 2021, date of publication February 8, 2021, date of current version February 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3057912

Weed Density and Distribution Estimation for Precision Agriculture Using Semi-Supervised Learning

SHANTAM SHOREWALA[1], ARMAAN ASHFAQUE[2], R. SIDHARTH[3], AND UJJWAL VERMA[1], (Senior Member, IEEE)

[1]Department of Electronics & Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[2]Department of Information and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[3]Department of Computer Science Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: Ujjwal Verma (ujjwal.verma@manipal.edu)

ABSTRACT Uncontrolled growth of weeds can severely affect the crop yield and quality. Unrestricted use of herbicide for weed removal alters biodiversity and cause environmental pollution. Instead, identifying weed-infested regions can aid selective chemical treatment of these regions. Advances in analyzing farm images have resulted in solutions to identify weed plants. However, a majority of these approaches are based on supervised learning methods which requires huge amount of manually annotated images. As a result, these supervised approaches are economically infeasible for the individual farmer because of the wide variety of plant species being cultivated. In this paper, we propose a deep learning-based semi-supervised approach for robust estimation of weed density and distribution across farmlands using only limited color images acquired from autonomous robots. This weed density and distribution can be useful in a site-specific weed management system for selective treatment of infected areas using autonomous robots. In this work, the foreground vegetation pixels containing crops and weeds are first identified using a Convolutional Neural Network (CNN) based unsupervised segmentation. Subsequently, the weed infected regions are identified using a fine-tuned CNN, eliminating the need for designing hand-crafted features. The approach is validated on two datasets of different crop/weed species (1) Crop Weed Field Image Dataset (CWFID), which consists of carrot plant images and the (2) Sugar Beets dataset. The proposed method is able to localize weed-infested regions a maximum recall of 0.99 and estimate weed density with a maximum accuracy of 82.13%. Hence, the proposed approach is shown to generalize to different plant species without the need for extensive labeled data.

INDEX TERMS Artificial intelligence, artificial neural networks, computer vision, convolutional neural networks, deep learning, crops, weeds, machine learning, neural networks, precision agriculture, ResNet, segmentation, semi-supervised learning, unsupervised learning.

I. INTRODUCTION

Agriculture continues to be the most important industry across the world necessary for sustaining mankind. There have been significant improvements in the machinery operated by the farmers to cultivate their lands. One common aspect of farming is *weeding* - it refers to the removal or treatment of weed plants. Weeds are undesirable plants that compete with crop plants for natural resources such as sunlight, minerals, and water. Hence it becomes necessary to selectively remove these plants to ensure a healthy crop yield [1], [2]. However, the traditional practice of treating the entire farmland indiscriminately with agrochemicals for weed control, in addition to being expensive, adversely impacts the soil biodiversity, quality of freshwater available to humans as well as the human health [3]–[6]. An alternative to chemical weeding is to manually pick the weed plants (manual weeding). This approach, however, is time and labour intensive.

Precision Agriculture is a ''management strategy that takes account of temporal and spatial variability to improve the sustainability of agricultural production'' [7]. Common applications of precision agriculture include identification of weeds,

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Ji.

crop and soil health monitoring, site-specific management for tasks such as tillage, sawing, mechanical weeding and distribution of fertilizers, crop yield estimation, fruit/vegetable detection and picking [8]–[14].

Autonomous robots have been used for chemically weeding patches of weed plants [15]–[17]. These robots rely on systems, including machine vision, to identify and localize weed plants. A typical image processing based weed detection approach consists of four stages: pre-processing, segmentation, feature extraction, and classification. The pre-processing prepares the input image for segmentation and typically consists of various image enhancement methods such as color space transformation. Subsequently, the enhanced image is segmented into two regions: vegetation and background. The segmentation procedure can be grouped into two categories: Index-based and learning based. The index-based approach differentiates between vegetation and background by comparing each pixel's intensity value with a threshold parameter. This approach is usually not robust to varying lighting conditions and overlapping crop and weed plants [18]–[20]. The learning-based methods for vegetation segmentation have been shown to overcome this challenge and are the preferred approach to accurately identify the vegetation [2]. The segmentation procedure produces the vegetation mask that contains both crop and weed pixels in the same class. Therefore, a hand-crafted feature vector is computed based on biological morphology, spectral features, visual textures, and spatial contexts of the crop and weed plants. These feature vectors are then fed to a classifier to identify weeds from the segmented vegetation mask. The image processing based approach overcomes labour and time-intensive demands of manual weeding in addition to reducing the amount of chemical sprayed. However, the use of hand crafted features restricts the usage of these approaches to a particular crop/weed species. Recently, deep learning based approaches [21]–[23] have been proposed which eliminates the need for hand crafted features. However, majority of these approaches are supervised approaches which require a huge amount of training data, thus limiting its application to few crop/weeds. The major challenges to a reliable, scalable vision system for the autonomous robots are (1) varying lighting conditions, (2) overlapping and occluded weed and crop plants, (3) varying weed density, and (4) different species of crop and weed plants. In addition, the supervised learning-based approach depends on the availability of annotated data. It may be noted that there exist other sensors such as visible and near-infrared (Vis-NIR) spectroscopy, LiDAR, and sonar [2], [24], [25] for weed identification. However, this study focuses on an image-based system for weed identification.

There also exists image classification based approaches for weed detection. In this approach, the entire image is labelled as a particular weed species, based on the weed species present in the image [26]. This approach is able to identify the weed present in the field but would not be able to compute the weed density. In this work, we propose an alternative patch based approach, which eliminates the need for pixel wise annotation, and can compute weed density and distribution. The primary objective of our work is to evaluate a semi-supervised pipeline for weed localization and density estimation in order to minimize the amount of manually annotated data required to train the deep networks. By reducing the dependence on data-intensive segmentation networks, we can enhance the adoption rate for different species of crops/weeds and in different environments/settings.

The main contribution of our work is a semi-supervised decision support system for robust estimation of weed distribution and density from a single color image acquired using an autonomous robot. Instead of focusing on pixel-wise segmentation, we seek to address the more fundamental question of which regions should be selectively treated with agrochemicals. This decision can be on the basis of estimated *weed distribution or localization* and *weed density*. The proposed approach can

- Robustly identify weed infected regions
- Compute weed density in the infected regions
- Enhance scalability and generalizability as it does not require pixel-wise annotations unlike end-to-end deep learning segmentation networks

The proposed approach leverages unsupervised Convolutional Neural Network to cluster the pixels into vegetation and background class. It is worth noting that any foreign objects or non-soil, non-vegetation pixels are also classified as background in the proposed approach. The vegetation mask is then overlaid on the input color image which is divided into smaller tiles. Each tile with vegetation coverage is then passed through a classifier that labels it either as weed or crop. Algorithm 1 briefly describes the different stages in the proposed approach. Unlike existing image-based methods for weed classification, the proposed approach does not rely upon hand-crafted features. Moreover, the proposed approach does not require extensive segmentation labeling of crop and weed plant pixels as used in [21]–[23], [27].

Algorithm 1: Weed Distribution and Density Estimation

Input: Color image (I_{RGB}) of the field acquired from an autonomous robot;
Output: Weed density and distribution;
Given (I_{RGB}), Generate the vegetation mask (I_{veg}) using CNN based unsupervised segmentation;
Overlay I_{RGB} with I_{veg} to get I_{masked};
Divide the image I_{masked} into smaller regions I_{tile} (square tiles);
for *(I_{tile} in I_{masked})* **do**
    Classify I_{tile} into crop, weed or background;
    if I_{tile} *is weed* **then**
        Estimate weed density
    end
end

The rest of the paper is structured as follows: Section II discusses existing approaches to identify weed plants.

Section III describes individual steps of the proposed approach in detail. Datasets used and results are discussed in the following section. Finally, the conclusions drawn are presented in Section V.

II. RELATED WORK

This section summarizes existing traditional as well as deep learning-based approaches for image-based weed detection and classification. For a detailed discussion, reader is referred to [2]. Recent advances in the field of deep learning have been applied to precision agriculture to improve the limitations and inflexibility of traditional methods. A review of state-of-the-art deep learning approaches to disparate problems in agriculture, including identification of weeds, land cover classification, and fruit counting, among others, can be found in [28].

A. SUPERVISED LEARNING

In the last few years, deep learning methods have achieved state-of-the-art results on challenging datasets for various applications such as autonomous driving [29], [30]. However, they are generic, designed to handle a large number of object classes. For weed identification and mapping, a much smaller number of classes need to be handled. Multiple research works have previously proposed an end-to-end semantic segmentation network, built upon earlier works such as SegNet [29], that distinguish between crop and weed plants [21]–[23]. In [21], networks is trained on 465 multispectral images and achieves extremely high F1 scores (>0.95). While the number of training images is relatively small, it does rely on images captured from multispectral sensors which results in higher costs. Authors of [22], [23] obtained comparable performance (F1 score > 0.90) with networks trained on a set of 10,000 RGB images. These results establish the feasibility of training deep learning models to discriminate crop and weed plants. However, as with all supervised learning models, they require an extensively manually annotated dataset to train the network. This challenge is not as prominent in applications where models can generalize reasonably well to different settings without loss of performance (such as object detection for common items such as chairs, humans, etc). Authors of [22] also study the adaptability of their work to different plants by testing the trained network on a different set, achieving accurate results and demonstrating the need for adaptable networks. However, the datasets compared are similar in terms of background visual features of the vegetation. Instead, in our study, we propose an alternative approach that to pixel-wise segmentation models. In [31], authors utilise scatter transforms to produce feature vectors for an SVM to classify culture crops. The approach is trained on a synthetic dataset and achieves an accuracy of around 85% on culture crops. Another approach to supervised learning that has been used can be found in [32]. This approach uses synthetic markers for crops that are planted to accurately detect them via computer vision and achieve very high results of around 99.7%. On the other hand, our approach takes as input raw RGB images with no types of augmentations or physical markers placed on the field.

An object detection based approach is proposed in [33] for weed identification. A deep neural network is trained to produce coverage maps and bounding boxes for localization of crops and weeds. While achieving accurate results, this is a very data-intensive approach that requires both covering maps and bounding boxes to be manually annotated. In a separate study [34], multispectral orthomosaic maps are generated by projecting a 3D point cloud onto the ground plane. They propose to overcome the challenge of scanning a large area while preserving the fine details of plant distribution. A modified SegNet model is then used to segment the weed pixels in these maps. Such an approach is data-intensive (the study used a dataset with more than 10,000 images), requiring sensors that can produce point clouds besides having to train an end-to-end segmentation model. Another study by [35] proposes a two-stage network that uses an end-to-end segmentation network to first create a binary vegetation mask. Vegetation blobs are then passed as patches to a deep VGG-16 network for classification. The two-stage pipeline is an useful technique but both the networks require training on the chosen types of crop fields. Our study builds upon the idea of a two-stages for identifying weed infestation by leveraging unsupervised learning for vegetation segmentation (which is the first stage). This leaves only tile labels to be generated for training the classifier. Thus, the use of these modules aids in reducing the data dependency and can be easily extended to any crop/weed combination.

B. TRANSFER LEARNING

The authors in [36] proposed a weed classifier which utilizes features extracted from a pre-trained sparse autoencoder [37], [38]. However, the algorithm makes two simplifications. Firstly, the example patches from the aerial images used are pre-selected hence making the pipeline semi-automated. Secondly, the dataset being used is balanced, which, in reality, is not the case with crop-weed datasets. Previous works such as [39] have tried to address the dependency on manually annotated extensive datasets. Using a pre-trained network [40], the authors trained a much smaller network compared to others. The results show that the network is able to generalize well to a small dataset without compromising on performance, achieving a best of 93.9% accuracy.

C. SEMI-SUPERVISED AND UNSUPERVISED LEARNING

Semi-supervised and unsupervised learning methods have also been studied to perform weed detection. For instance, a comparative study of two deep unsupervised learning algorithms JULE [41] and DeepCluster [42] is presented by [1], along with a deep network like VGG-16 [43] or ResNet-50 [44] to help classify and automatically label different classes of weeds. [45] use K-Means pre-training to adjust network weights before a LeNet-5 [46] model is used to classify the type of weeds. These approaches do not predict a dense map for weed or

weed pixels, only the class to which the image belongs. Hence, they cannot estimate the weed density, which is imperative as variable spraying of herbicide leads to increased application efficiency and reduced environmental impact [47]–[49]. Authors of [50] propose an unsupervised approach to cluster plants into different classes. They achieve competitive results under the assumption that none of the plants (either weed or crop) overlap each other. In practice, it is not an assumption that will hold true for varying plant species. Moreover, one of the challenging tasks in the unsupervised approach is determining the optimal number of clusters in which the image should be segmented [1]. In the proposed work, we alleviate this difficulty by utilizing the unsupervised approach for only segmenting the vegetation mask, thereby fixing the number of the cluster as two.

The method described by [35] is closest to the proposed approach. The authors utilize a deep learning-based method for weed identification. A two-stage network was used: first, CNN extracts the vegetation mask while the second CNN identifies weeds from crops. However, there is a significant difference when it comes to the components used in the proposed work. Compared to supervised learning networks adopted by [35], the proposed method requires significantly less training data (vegetation segmentation is unsupervised while the classifier is trained with a small number of region labels). The dataset used in [35] consisted of 2000 images while the proposed approach is tested on network trained on datasets with 90 and 500 images respectively (including augmented images); this highlights the significant reduction in the number of images. This contrast is further increased by the type of annotation required - proposed pipeline eliminates the need for pixel-wise annotations whereas the networks in [35] need pixel-wise annotated images for training. Only the classifier needs to be fine-tuned with images of new plant species and binary labels in the proposed work. Further, they mention that a large percentage of errors arise due to overlapping plants. The proposed approach is shown to be both robust to poor illumination, occlusions, and plant density as well as adaptable to varying plant species. Unlike other semi-supervised approaches, the proposed method can still robustly estimate both the weed density and distribution, from RGB images.

D. WEED DENSITY ESTIMATION

Weed density is an important parameter which helps in identifying the regions to be treated with chemicals [47]–[49]. In [51], [52], authors propose methods to estimate weed densities in row crops. Reference [51] describes the weed density with cluster rate (ratio of weed quantity to land area) and weed pressure (ratio of weed quantity to crop) parameters. Reference [52] extract the weed distribution using a positional histogram. The histogram is plotted by counting the number of white pixels in a binary vegetation mask along each column to obtain the lateral pixel distribution. Weed density (ratio of weed pixels in each interval to the image size) is obtained for fixed intervals. This approach is suitable only for only inter-row weed plants and does not account for weed-crop overlapping. Also, prior knowledge of crop row positions to estimate weed densities is assumed. Hence, it is not easily adaptable to different kinds of crop plantations.

III. METHODOLOGY

In order to selectively treat the farmland under cultivation, the proposed approach identifies the weed-infested regions and the corresponding weed density. A single RGB image is taken as the input for the pipeline. Image pixels are first clustered into two classes (vegetation and background) using an unsupervised deep learning-based segmentation network. This process generates a vegetation mask (foreground) and a background mask. The vegetation mask, I_{veg}, is then overlaid on the original RGB image to get the region of interest (RoI) denoted by I_{masked}. This is then divided into smaller regions or patches I_{tile} (square tiles). For each tile (I_{tile}), a corresponding feature vector is extracted that describes the properties of the vegetation pixels present in the tile. These vectors are then used to classify I_{tile} as either crop or weed plant using a binary classifier. In addition, the performance of a fine-tuned CNN (ResNet50) is also studied for classifying I_{tile}. The location of weed-infested regions can be inferred from the regions I_{tile}, which are classified as weed label. Weed density in the region can be estimated from the vegetation pixel density in the area. The ratio of the number of vegetation pixels in each region, classified as either crop or weed, to the region's total land area in pixels gives the corresponding density estimate. It may be noted that only a part of the proposed method is trained in a supervised manner, resulting in a scalable approach that can be adapted for different weed and crop plant species. Figure 1 provides an overview of the proposed pipeline. The rest of the section describes the individual steps in detail.

A. VEGETATION SEGMENTATION

The input image (I_{RGB}) is first resized to 500 × 500 sq. pixels using the bicubic interpolation method implemented by OpenCV library [53]. Each pixel in the image has to be clustered into one of the two classes - background or vegetation. For this purpose, we use the CNN based approach proposed in [54] for unsupervised segmentation. However, to make the paper self-contained, the work is described in brief. This iterative approach is solved in two steps: label prediction assuming fixed network parameter (forward pass of the network) and learning network parameters through back-propagation assuming fixed (predicted) labels. The approach proposes the following constraints for predicting the cluster or class to which each pixel might belong - (1) The first constraint is on feature similarity. Pixels that are similar to each other are clustered together. In order to achieve this, a response map for all the pixels is generated. Based on it, each pixel is assigned to a cluster to which it is closest according to the response map, (2) The second constraint is on spatial continuity. The authors use a number of superpixels extracted from the image and force the same cluster label for all the labels within. Superpixels can be defined as a group or cluster of pixels

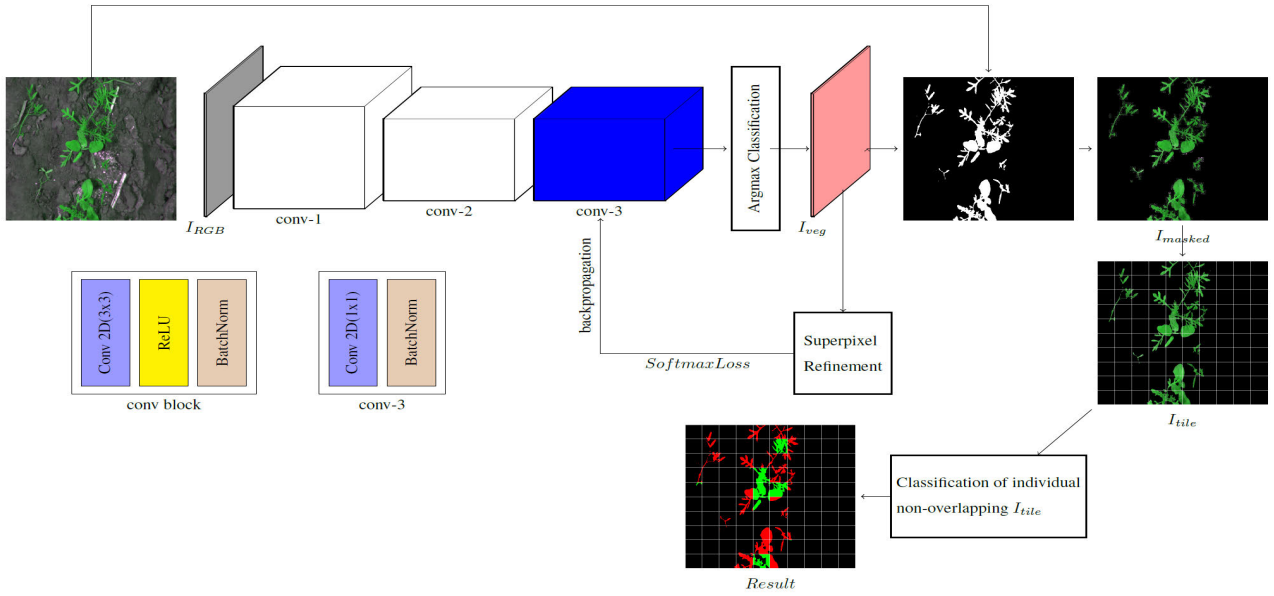


FIGURE 1. Overview of the proposed approach: An unsupervised CNN based binary segmentation is applied to the input color image to generate the vegetation mask. Subsequently, the masked image is sub-divided into non-overlapping tiles, which are then passed through a classifier. This classifier classifies the tiles as weed, crop, or background.

that exhibit common characteristics such as pixel intensity and proximity. The network extracts the superpixels using the Simple Linear Iterative Clustering (SLIC) algorithm [55] which operates in a five-dimensional space (three channels of CieLab colorspace and 2D image coordinates (x, y)), (3) Final constraint is placed on the number of unique clusters into which the image is segmented. Given a maximum number of clusters q, the preference is for a large number of classes to avoid under segmentation. The solution for this constraint is to perform intra-axis normalization on the response map before assigning the cluster labels. These constraints imposed on pixel-wise segmentation justified the choice for our pipeline. It assigns additional weight to spatially continuous pixels (each weed and crop plant is a closed-loop structure) and allows us to force the minimum number of clusters to two (background and vegetation). The pixel-wise segmentation is iterated until one of the following two conditions is met: (1) the majority of pixels are classified into two clusters or (2) maximum iterations are reached. This further avoids under or over-segmentation and places an upper bound on time taken to converge to the final segmentation result. Once the segmented image is generated, the cluster with the lower number of pixels is considered to be the vegetation mask. This is based on the assumption that the number of background pixels will be greater than the vegetation pixels. Validity of this assumption is discussed in the Section III-D.

In order to maximize the performance of the unsupervised segmentation, we tune the parameters of the network, purposing a randomly sampled subset ($\sim$ 30%) of both the datasets for validation. The parameters tuned for the network are learning rate, number, and compactness of superpixels. Only one parameter is varied at a time to determine the optimal values. Other parameters are kept constant during this time. The optimal values are selected for which a maximum mean intersection-over-union (mIOU) is obtained. The experimentally determined values of the parameters are: (1) learning rate - 0.1, (2) number of superpixels - 2500, (3) compactness of superpixels - 25. Using the optimal values, the vegetation masks, I_{veg} are generated for the images in the test split. As a benchmark and to compare the performance of the unsupervised approach with a supervised approach, we also trained U-Net [30] on the training split of the datasets. U-Net has been shown to be an effective supervised learning approach for pixel-wise segmentation in different use cases such as medical image segmentation and autonomous driving. The network uses an encoder-decoder structure first to contract (downsample) the image and then expand (upsample) to get the final prediction. At each ''upsampling'' step, the feature map from the corresponding contraction step is also concatenated. This concatenation helps the network to learn from the lost features during downsampling. The network was utilized to predict binary class labels, with the foreground pixels (white) representing the vegetation coverage.

B. TILE CLASSIFICATION

Once the vegetation mask I_{veg} is generated, the input image I_{RGB} is overlaid with I_{veg} resulting in the masked image I_{masked}. This masked image contains only the RGB pixels for the vegetation (crops and weeds), which ensures that classification is performed based on vegetation features alone. Further, the masked image I_{masked} is divided into multiple non-overlapping sub-images/tiles I_{tile} of size 50 $\times$ 50 sq. pixels. It is possible that many regions contain a very small number of vegetation pixels or even none. Therefore, I_{tiles} where vegetation coverage (number of vegetation pixels) is less than 10% of the total area of the region (in pixels) are

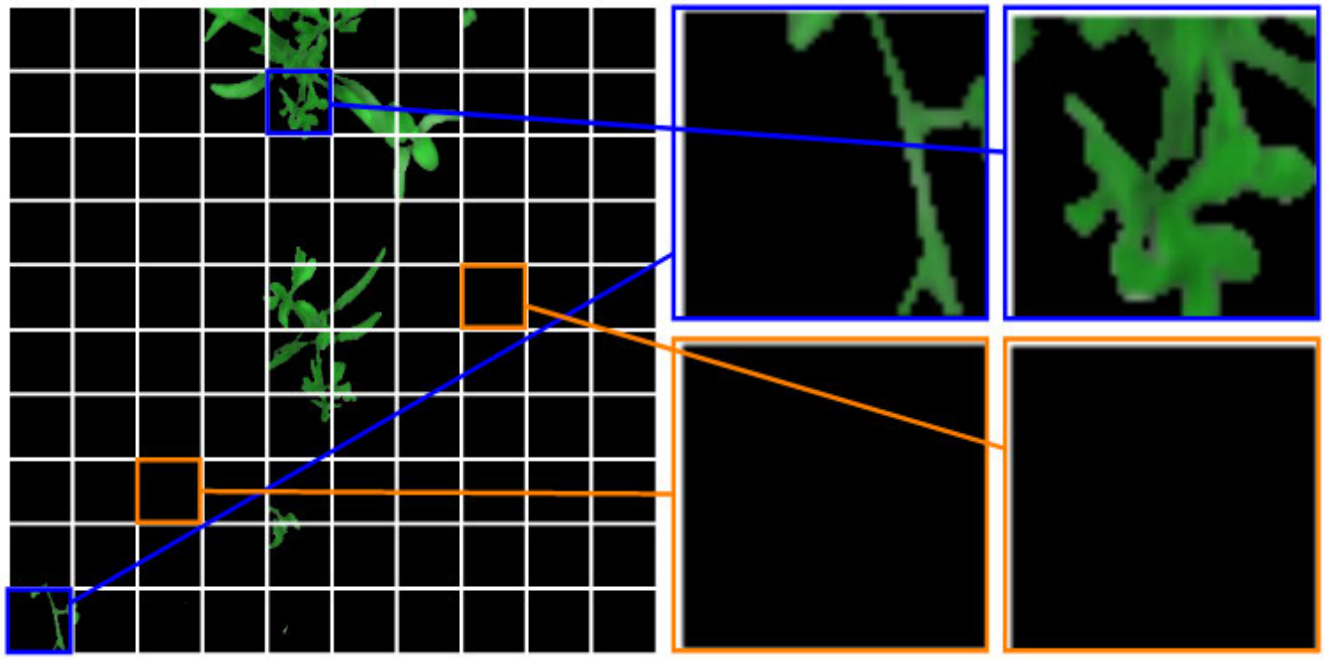


FIGURE 2. I_{masked} is divided into smaller tiles (I_{tile}) as shown. Note that I_{tile} is enlarged for better visualization. Also, the outline colors represents the classification for the region: Blue - Vegetation, Orange - Background/Soil.

considered to be not infested with weed plants and are ignored in the following steps. Figure 2 shows the image I_{masked} as well as the regions selected for training the classifier and the ones discarded because no vegetation pixels are present.

A variety of machine learning algorithms such as Support Vector Machines, Random Forest Classifiers, Gaussian naive Bayes, and multilayer perceptron networks [56]–[58] have commonly been used for classification. We compare the performance of these classifiers in terms of classifying I_{tile} as either weed or crop. This section first discusses classifiers which uses feature vectors for classifying I_{tile} as weed or crop. Also, an image based classifier is discussed which does not explicitly computes the feature vector but rather uses a fine-tuned CNN to classify I_{tile} as weed or crop.

1) FEATURE VECTOR BASED CLASSIFICATION

The feature vectors are computed from the filtered set of I_{tile}. However, instead of extracting features based on the biological morphology, physical appearance of the crop/weed, a pre-trained CNN is utilized. The use of pre-trained CNN as a feature extractor eliminates the need for designing hand-crafted features and can be extended to any crop/weed combination. ResNet50 [44] is utilized in our method to extract feature from I_{tile}. ResNet50 is a CNN based supervised learning approach for image classification. ResNet50 consists of multiple residual blocks stacked one after another. These residual blocks use skip connection wherein the activation of a $(l+2)^{th}$ layer is computed by the addition of activation of $(l+1)^{th}$ and l^{th} layers. This skip connection helps in designing a deeper network by alleviating the problem of vanishing gradient.

In this work, pre-trained weights learned on the ImageNet dataset [59] are used. Feature vectors, of length 16,384, are extracted from one of the intermediate layers in the network (the 3rd block from the end). This is based on the intuition that the network would learn generic features such as corners and edges in the earlier layers to extract useful information while using class-specific information like shape and color in the deeper layers to label images. Principal Component Analysis (PCA) [60] is further used to reduce the vector dimensionality to 2048. PCA uses an orthogonal transformation on the set of all feature vectors to convert any possibly correlated features to a set of linearly uncorrelated features. After the feature vector from all the regions I_{tile} have been extracted and reduced, they are then passed to the classification block.

The SVM classifier was trained with different kernels: radial basis function, linear kernel, sigmoid, and various degrees of polynomial curves. A small multi-layer perceptron was also trained for classification (Figure 3). It consisted of six hidden layers besides the input and output layer utilizing the Rectified Linear Unit (ReLU) activation ([61]). The activation $a^{[i]}$ for a neuron in i^{th} layer is defined as

$$a^{[i]} = g(w^{[i]}a^{[j]} + b^{[i]}) \tag{1}$$

$$g(x) = max(0, x) \tag{2}$$

where $w^{[i]}$ are the weights, $a^{[j]}$ are the activation of the previous (j^{th}) layer, and $g(x)$ is the non-linear ReLU activation. The input size of the network was 2048 (length of reduced feature vectors), and the output layer consisted of a single neuron that made binary predictions (0=crop, 1=weed).

Disproportionate occurrence of crop and weed plants in the dataset leads to class imbalance in the training set.

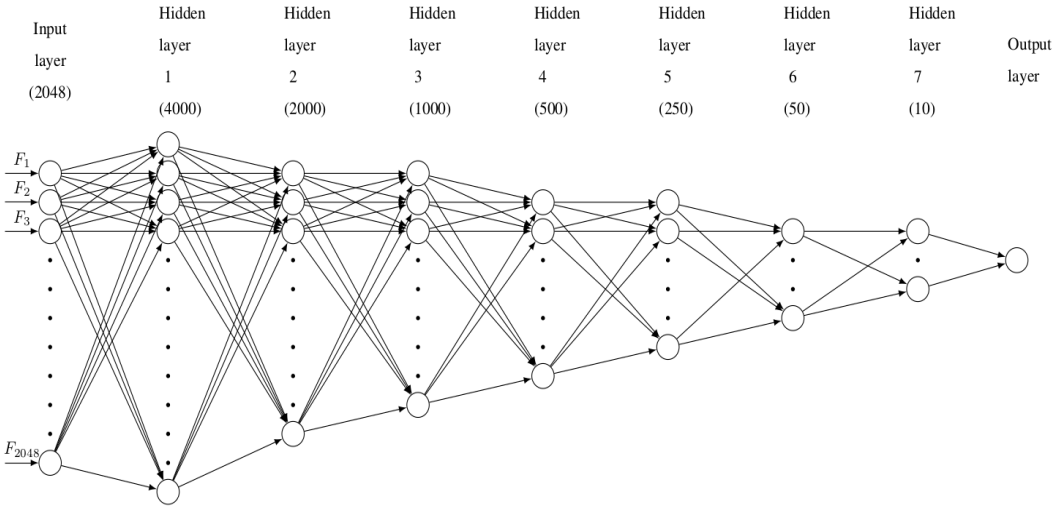


FIGURE 3. Multi-layer perceptron for classifying the sub-image I_{tile} as crop or weed. The input 2048 dimensional feature vectors are shown as F_i. Number of neurons for each layer are included in parentheses.

It could severely hinder the classifier's ability to recognise the weed-infested regions. In order to address the issue, we implement and compare two different sampling techniques to increase the frequency of weed plant samples as for most imbalanced data sets, the application of sampling techniques does indeed aid in improved classifier accuracy [62]. Firstly, a combination of random oversampling (resample certain data points from the minority class) and undersampling (drops data points from the majority class) is used while training the model, thus increasing the ratio of weed tile samples in the dataset. The alternative approach is to implement Synthetic Minority Oversampling Technique (SMOTE) [63] which uses K-Nearest Neighbours algorithm to generate synthetic samples of the minority class by utilizing the existing minority class data points. Both these techniques expose the classifier to a greater number of weed plant tiles. This reduces the bias towards the majority class during the learning process.

2) IMAGE-BASED CLASSIFICATION

Instead of extracting feature vectors from the tiles, ResNet50 itself can be fine-tuned on the filtered set of tiles (I_{tile}) to make label predictions (crop/weed). It has been shown that ResNet50 trained on more than a million training images from ImageNet database for more than 1000 categories learns rich feature representations [64]. Same pre-trained weights described in the previous paragraph are used here as well. However, instead of using ResNet50 as a fixed feature extractor, the weights of the last layer of ResNet50 is fine-tuned. In this approach, weight of only the last layer of the pre-trained ResNet50 is updated via backpropogation, while the weights of all the other layers are fixed (freezed). In order to address the class imbalance problem, ResNet50 is tuned using the weighted binary cross-entropy loss function. The loss function for each class is defined as follows:

$$loss[x, c] = \sum_{n} -weight[c] * (x[c] + \log(\sum_{j} \exp(x[j]))) \tag{3}$$

Here, c denotes the class, $j \in [1, \text{number of class}]$, n is the number of images in the batch and x is the distance between the target and predicted label. The weights for crop and weed class are determined experimentally as 0.33 and 0.67 respectively. Besides, the performance of weighted cross-entropy loss is also compared with standard cross-entropy loss (i.e. with weights equal to 0.5). The model is trained for 250 epochs with the learning rate set to 0.001. It is also important to note that network is trained using the masked tile images instead of complete RGB images - this allows the network to make predictions based on vegetation coverage in the region instead of background/foreign objects (which are segmented out in the previous step).

C. WEED DENSITY ESTIMATION

Once the weed-infested regions (I_{tile} classified as a weed) have been identified, the weed density can be computed from the vegetation coverage in each individual region. In this paper, the cluster rate ([51]), denoted by CR, is used to quantify or model the weed density. It is defined as follows:

$$CR = \frac{\text{Weed plant coverage in the region (in pixels)}}{\text{Total land area of the region (in pixels)}} \tag{4}$$

The weed density estimate is crucial information in the site-specific weed management system [2]. This density estimate would assist in selecting the appropriate regions for weeding with herbicides in the field. This decision-making process would depend on a variety of factors, including but not limited to crop and weed plant species and plant spacing.

D. DATASET

The proposed approach is validated on two commonly used publicly available datasets: Crop/Weed Field Image dataset [65] and the Sugar Beets dataset [66]. Authors of both the datasets provide annotated images that mark crop and weed plant pixels distinctly.

1) CROP/WEED FIELD IMAGE DATASET (CWFID)

It contains images acquired by an autonomous field robot BoniRob from a carrot farm. This dataset includes 60 top-down field images with intra-row and close-to-crop weeds. In this work, CWFID was augmented using common techniques such as skewing, flipping, rotating, and zooming. This resulted in a total of 90 images, which are split into train and test set in the ratio of 2:1 (60 training images, 30 testing images).

2) SUGAR BEETS DATASET

It contains field images acquired by the same autonomous robot BoniRob from a sugar beet farm for over three months. While the entire dataset is quite extensive and includes data from multiple sensors, only a subset of the Sugar Beets dataset (500 images) is used in this study. Compared to CWFID, the dataset presents a variation in terms of plant species and the number of overlapping plants. Besides, unlike CWFID, Sugar Beets dataset suffers from poor contrast arising due to insufficient illumination. It is further split into the train and test set with a ratio of 7:3 (350 training images, 150 testing images).

3) CLASSIFICATION DATASET

Since the pixel wise annotated masks were provided with the dataset, the tile label for classification was deduced from this pixel level annotation. The full sized images were divided into 50×50 squares and a corresponding label was chosen for the tile based on the number of the crop/weed pixels in the tile image. If the tile image contained more crop pixels, it was labelled as crop, if it contained more weed pixels it was labelled as weed. Any tiles with less than 10% vegetation coverage were ignored. Table 1 shows the number of crop and weed tiles used by the authors to train the model from the dataset. It is important to note that in absence of pixel wise annotations, only a single binary label for each tile needs to be manually specified. This will reduce the time and effort required for the annotation process significantly.

TABLE 1. Tiles generated for classification.

Dataset	Crop		Weed	
	Train	Test	Train	Test
CWFID	1370	411	637	244
Sugar Beets	5585	2833	2555	947

Since both the datasets are captured from ground robots, the vegetation pixel density in the images is sparse compared to background or soil pixel density. The vegetation pixel density for CWFID and Sugar Beets dataset is 10.09% and 6.58%, respectively. Further, the maximum vegetation density in a single image for the dataset is 23.76% for CWFID and 19.53% for the Sugar Beets dataset. Hence, due to the steep difference in the vegetation pixel and soil pixel densities, the assumption made in Section III-A is reasonable (the vegetation cluster will always contain a lower number of pixels compared to the background).

The motivation behind selecting datasets captured using the same autonomous robot was to facilitate easier integration of the proposed method in the existing infrastructure. Even though the images have been acquired using the same autonomous robot, the two dataset contains two different crop/weed species (Sugar Beets and Carrots Crops) with varying overlap and image contrast (likely due to varying lighting/illumination). This results in a significant difference in the image content of the two datasets.

E. EVALUATION

Mean intersection-over-union (mIoU) is a popular metric to evaluate pixel-wise segmentation networks [29]. However, the focus of our approach is not a dense segmentation prediction. Instead, the accuracy of weed distribution and density estimation is evaluated as explained below. Besides, in order to justify the choice of individual components, we also evaluate the output from the intermediate step (unsupervised binary segmentation).

1) VEGETATION SEGMENTATION

The first step of the pipeline segments the vegetation pixels for the input RGB image. The vegetation pixels are denoted by white color in the binary image. Ground truth vegetation coverage ($coverage_{GT}$) and predicted vegetation coverage ($coverage_{pred}$) are compared using the mIoU metric defined below. It helps determine the effectiveness with which the vegetation pixels are identified in the image.

$$mIoU = \frac{\sum_i x_{ii}}{C\left(\sum_i \sum_j x_{ij} + \sum_j x_{ji} - x_{ii}\right)} \tag{5}$$

where C is the number of classes (two in this work), x_{ij} represents the number of pixels belonging to class i and predicted as class j. The maximum value of mIoU is 1.0 which signifies that the all the pixels are correctly labelled.

2) WEED DISTRIBUTION ESTIMATION

The predictions made by the classification model for all the regions (squares tiles) are compared against the ground truth labels. Classifier accuracy highlights the overall system performance but is also biased towards the majority class (the class with a significantly greater number of samples). Hence, the improvements in the classifier performance with respect to a specific class can be reflected by the recall and precision metrics (Equations 6, 7). Besides, the F1-score can be utilized

to study the overall performance of the classifier (Equation 8)

$$precision = \frac{TP}{TP + FP} \tag{6}$$

$$recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1\text{-score} = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{8}$$

Here, TP refers to true positives, FP refers to false positives, and FN refers to false negatives. For a particular class, true positives are all the regions correctly assigned to that class, false positives are the regions incorrectly assigned to that class, and false negatives are the regions incorrectly labeled as another class. While F1-score is an essential and reliable metric for measuring the classifier performance, recall is also afforded significant weightage. The reason for our choice is that we aim to minimize misclassification for weed-infested regions.

3) WEED DENSITY

In order to evaluate weed density, error in predicted density for each region correctly classified is computed. The weed density for each tile is measured by the cluster rate (Equation 4). The estimated cluster rate (CR_{est}) is compared against cluster rate in ground truth pixel-wise annotations (CR_{gt}). The following three metrics are computed to quantify the error in weed density estimation: (1) mean accuracy, (2) mean absolute error (MAE) and (3) root mean squared error (RMSE) (Equations 10, 11. 12).

$$\text{Absolute Error} = \mid CR_{gt} - CR_{est} \mid \tag{9}$$

$$\text{Mean Accuracy} = 1 - \sum_{i} \frac{\text{Absolute Error}/CR_{gt}}{N} \tag{10}$$

$$\text{MAE} = \sum_{i} \frac{\text{Absolute Error}}{\text{N}} \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_i \text{Absolute Error}^2}{N}} \tag{12}$$

where CR_i is the ratio of weed plant coverage in the given tile to the total land area of the region (both in pixels), i = Ground Truth (gt), Estimated (est) and N is the total number of regions/tiles.

F. SOFTWARE

The proposed approach is developed using the Python programming language. Most of the programs are developed from scratch by the authors while open-source implementations are also used. In addition, common libraries such as OpenCV [53] and Scikit-Learn [67] are also utilized. The program is developed using Ubuntu 16.04 LTS operating system. The system had an octa-core CPU (Intel i7-7700HQ) and a NVIDIA GeForce GTX 1050 Ti (4GB RAM) graphics card. The code will be made available publicly on GitHub at https://github.com/ShantamShorewala/weed-distribution-and-density-estimation.

IV. RESULTS AND DISCUSSION

A. VEGETATION SEGMENTATION

The performance of the CNN based unsupervised segmentation for vegetation segmentation using the method presented in Section III-A is described in this section.

1) QUALITATIVE EVALUATION

Figure 4 visualizes the vegetation segmentation results for a few input images from both the datasets. It can be inferred that the unsupervised segmentation network matches, and in some cases even outperforms supervised segmentation (U-Net) in terms of segmenting the vegetation pixels from the background. For instance, the unsupervised segmentation approach is able to segment the finer details of the vegetation, as shown in Figure 4 (Top right corner of the image in the third row). It is also interesting to note that U-Net also labels some individual or extremely small clusters of pixels as vegetation. The unsupervised approach, in general, does not repeat this trend. It can be attributed to the fact that it gives weightage to spatial continuity of the vegetation clusters, whereas U-Net tends to look at only the immediate surroundings of a pixel (downsampling using max pooling). This trend was much more prominent for the Sugar Beets dataset, which exhibits poor contrast in comparison with the CWFID.

2) QUANTITATIVE EVALUATION

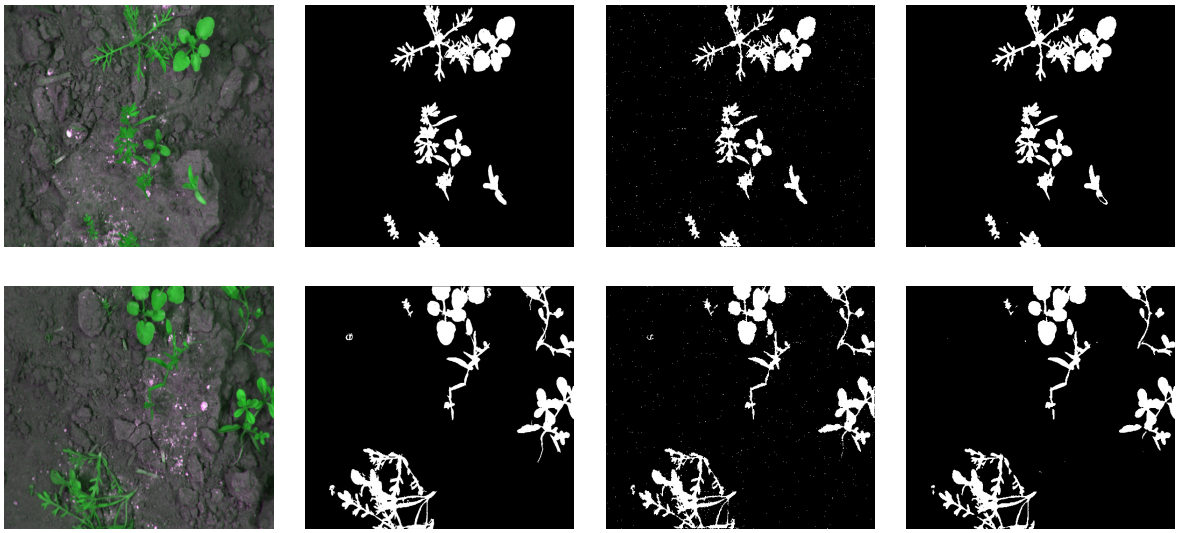
Mean intersection-over-union (mIoU) computed for both the test split of both the datasets is used to compare the performance of the two networks. The results are reported in Table 2. It is observed that the unsupervised network produced a slightly higher score for the CWFID and significantly outperformed U-Net for the Sugar Beets dataset. This observation can be attributed to the fact that, unlike U-Net, the unsupervised approach does not need to learn mapping from specific features to discriminate between the vegetation and background pixels. It should also be noted that both supervised and unsupervised approaches perform better for the CWFID compared to the Sugar Beets dataset. This observation can be attributed to the fact that the Sugar Beets dataset's images suffer from poor illumination and, hence, poorer contrast. These results justify the selection of the unsupervised network to segment the vegetation pixels from images captured for different plant species under varying conditions.

TABLE 2. Quantitative evaluation for vegetation segmentation.

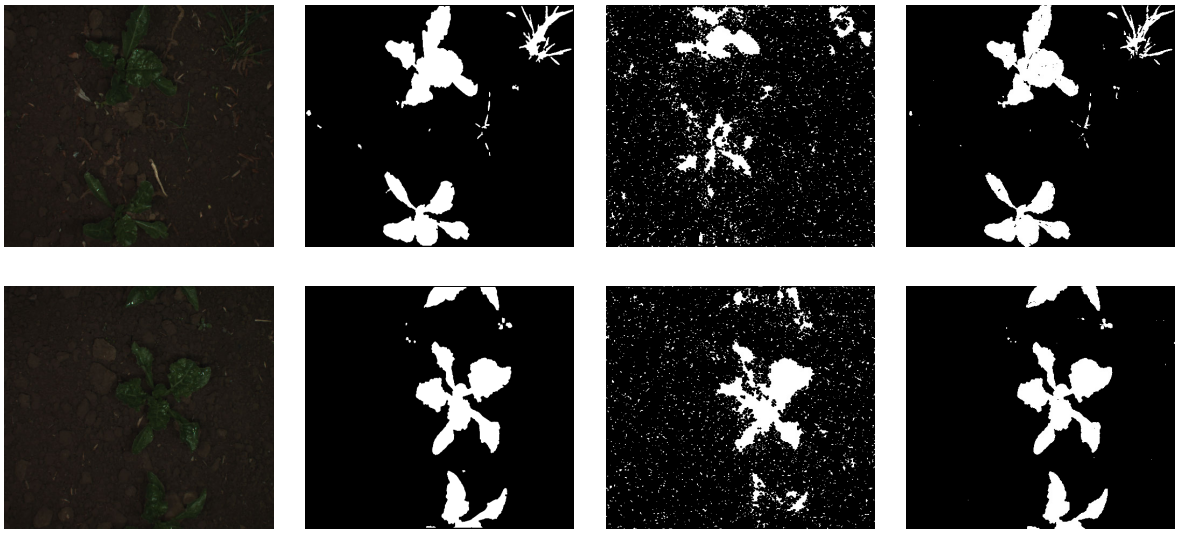
Model	Dataset	mIoU
Unsupervised Segmentation	CWFID	0.928
Unsupervised Segmentation	Sugar Beet Dataset	0.82
UNet	CWFID	0.913
UNet	Sugar Beet Dataset	0.76

B. FEATURE VECTOR BASED TILE CLASSIFICATION

As discussed earlier, the vegetation segmentation I_{veg} is utilized to identify the regions of vegetation containing crops and weed in the images. Subsequently, the input



(a) CWFID



(b) Sugar Beet Dataset

FIGURE 4. Vegetation masks (left to right): (I) Original Image, (II) Ground truth vegetation mask, (III) Vegetation mask predicted by UNet, (IV) Vegetation mask predicted by the unsupervised network.

image I_{RGB} is overlaid with I_{veg} resulting in masked image I_{masked}. This masked image is then divided into non-overlapping tiles (sub-images) I_{tile}. Subsequently, the features are extracted from each I_{tile} using a pre-trained ResNet50. The performance of different classifiers in classifying I_{tile} as weed or crop using these features is presented in Table 3.

Note the improvement in classifier performance due to weighted training using different approaches. (Table 3). This result substantiate previous finding [63] by demonstrating that sampling techniques (random sampling and SMOTE) helps in improving the classifier performance for an unbalanced dataset. The performance is measured using the computed precision and recall for the weed class on the test set. While the precision and recall values improve relatively due to sampling techniques that address class imbalance, the absolute values remain below the acceptable threshold. Random forest classifier achieved a recall of 1.0 but an extremely poor precision - all tiles were predicted as weed-infested. This demonstrates the inability of these classifiers to robustly discriminate between feature vectors generated from the proposed pipeline corresponding to crop and weed plants.

C. EFFECT OF TILE SIZE ON CLASSIFICATION ACCURACY

The choice of tile size (a square with side 50 pixels) was intuitively based on two observations: (1) it resulted in regions where pixels belonged largely to either one of crop or weed plants instead of both and (2) it avoided the formation of regions with nearly all pixels belonging to vegetation cluster. This would mean there would not be enough descriptive

TABLE 3. Comparison of classifier performance on the two datasets. (LK=Linear Kernel, PK = Polynomial Kernel, GNB = Gaussian Naive Bayes, NN = Neural Network, RF = Random Forest).

Dataset	Classifier	Vanilla			Random Sampling			SMOTE		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
CWFID	SVM with LK	0.26	0.12	0.16	0.3	0.23	0.26	0.31	0.22	0.26
	SVM with 2nd order PK	0.11	0.03	0.04	0.1	0.02	0.03	0.31	0.17	0.22
	SVM with 3rd order PK	0.0	0.0	-	0.33	0.01	0.02	0.28	0.15	0.2
	SVM with RBF Kernel	0.31	0.16	0.21	0.28	0.22	0.25	0.16	0.04	0.06
	SVM with Sigmoid Kernel	0.31	0.13	0.18	0.29	0.23	0.26	0.29	0.31	0.30
	GNB	0.16	1	0.28	0.16	0.1	0.28	0.16	1	0.26
	NN	0.32	0.30	0.31	0.31	0.31	0.31	**0.25**	**0.92**	**0.39**
	RF	0.29	0.07	0.11	0.35	0.14	0.2	0.22	0.06	0.0.09
Sugar Beets	SVM with LK	0.35	0.36	0.36	0.31	0.40	0.35	0.26	0.40	0.32
	SVM with 2nd order PK	0.32	0.50	0.39	0.32	0.37	0.34	0.31	0.52	0.39
	SVM with 3rd order PK	0.32	0.16	0.21	0.26	0.68	0.38	0.32	0.17	0.22
	SVM with RBF Kernel	0.44	0.13	0.20	0.28	0.47	0.35	0.34	0.46	0.39
	SVM with Sigmoid Kernel	0.34	0.42	0.38	0.29	0.59	0.39	0.29	0.61	0.39
	GNB	0.26	0.60	0.36	0.16	0.85	0.27	0.23	0.67	0.34
	NN	0.37	0.21	0.27	0.33	0.46	0.38	**0.28**	**0.70**	**0.40**
	RF	0.36	0.38	0.37	0.26	0.55	0.35	0.30	0.32	0.31

TABLE 4. Comparison of classifier performance for different patch sizes. (LK=Linear Kernel, PK = Polynomial Kernel, GNB = Gaussian Naive Bayes, NN = Neural Network, RF = Random Forest).

Dataset	Classifier	50×50			100×100			25×25		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
CWFID	SVM with LK	0.26	0.12	0.16	0.39	0.32	0.35	0.35	0.48	0.40
	SVM with 2nd order PK	0.11	0.03	0.04	0.40	0.02	0.04	0.0	0.0	-
	SVM with 3rd order PL	0.0	0.0	-	0.22	0.07	0.11	0.0	0.0	-
	SVM with RBF Kernel	0.31	0.16	0.21	0.43	0.07	0.14	0.34	0.43	0.38
	SVM with Sigmoid Kernel	0.31	0.13	0.18	0.39	0.22	0.28	0.34	0.43	0.38
	GNB	0.16	1.0	0.28	0.0	0.0	-	0.35	0.24	0.29
	NN	0.32	0.30	0.31	0.30	0.52	0.38	0.42	0.21	0.28
	RF	0.29	0.07	0.11	0.75	0.07	0.13	0.36	0.16	0.22
Sugar Beets	SVM with LK	0.35	0.36	0.36	0.42	0.21	0.28	0.37	0.14	0.20
	SVM with 2nd order PK	0.32	0.50	39	0.27	0.42	0.33	0.31	0.29	0.30
	SVM with 3rd order PL	0.32	0.16	0.21	0.28	0.09	0.14	0	0	-
	SVM with RBF Kernel	0.44	0.13	0.20	0.34	0.39	0.36	0.29	0.67	0.41
	SVM with Sigmoid Kernel	0.34	0.42	0.38	0.31	0.44	0.36	0.28	0.7	0.40
	GNB	0.26	0.60	0.36	0	0	-	0.12	0.97	0.21
	NN	0.36	0.38	0.37	0.35	0.25	0.29	0.4	0.13	0.2
	RF	0.37	0.21	0.27	0.25	0.22	0.23	0.28	0.43	0.34

features for the classifier to distinguish crop and weed plants since they would look extremely similar. However, to validate the choice of tile size, results from regions of different sizes were compared. In this study, we retrained the classification models by both increasing and decreasing the side length (100 and 25 pixels, respectively). Table 4 reports precision and recall values for all the machine learning classifiers.

Considering both precision and recall values, classifiers trained with tiles of side length 25 and 50, on average, outperform those trained on 100×100 tiles. Further, Table 5 reports the computation time required to pass the image through the classification block (includes time to generate the tiles, extract feature vectors, and classify them). The computation time is similar for patch sizes with side 50 and 100 pixels but increases significantly for side length 25 pixels. The reason was that the number of tiles with vegetation pixel density greater than 10% remained similar for the first two but is much higher for patches with side of length 25 pixels. Hence, in order to choose between the tile side length between 25, 50, and 100 pixels, processing or computation time for a single image is also taken into account. Based on these considerations, choice of patch size with side of length 50 pixels is justified for both datasets. It may be noted that the tile size needs to be computed only once and would be fixed for the entire agricultural season for a particular crop/weed species assuming no significant changes in the acquisition system on-board the autonomous vehicle.

TABLE 5. Computation time for passing a single image.

Tile side length	100 px	50 px	25 px
Computation time	0.96 s	0.90 s	5.22 s

D. IMAGE-BASED CLASSIFICATION

Despite the poor performance of the classical machine learning classifiers as a whole, the comparison was useful in determining the ideal tile size and validating weighted training of the networks. For image-based classification, ResNet50 was fine-tuned to classify I_{tile} as crop or weed with the same tile size. The computed precision and recall values for weed/crop

TABLE 6. Precision and recall values for classification using ResNet50 on the two datasets. (Class 0: Crop, Class 1: Weed).

Dataset	Cross-Entropy Loss						Weighted Cross-Entropy Loss					
	P0	R0	F1	P1	R1	F1	P0	R0	F1	P1	R1	F1
CWFID	0.94	0.72	0.81	0.31	0.87	0.45	0.95	0.60	0.74	0.32	0.91	0.47
Sugar Beets	0.84	0.99	0.90	0.18	0.49	0.26	0.99	0.56	0.72	0.37	0.99	0.53

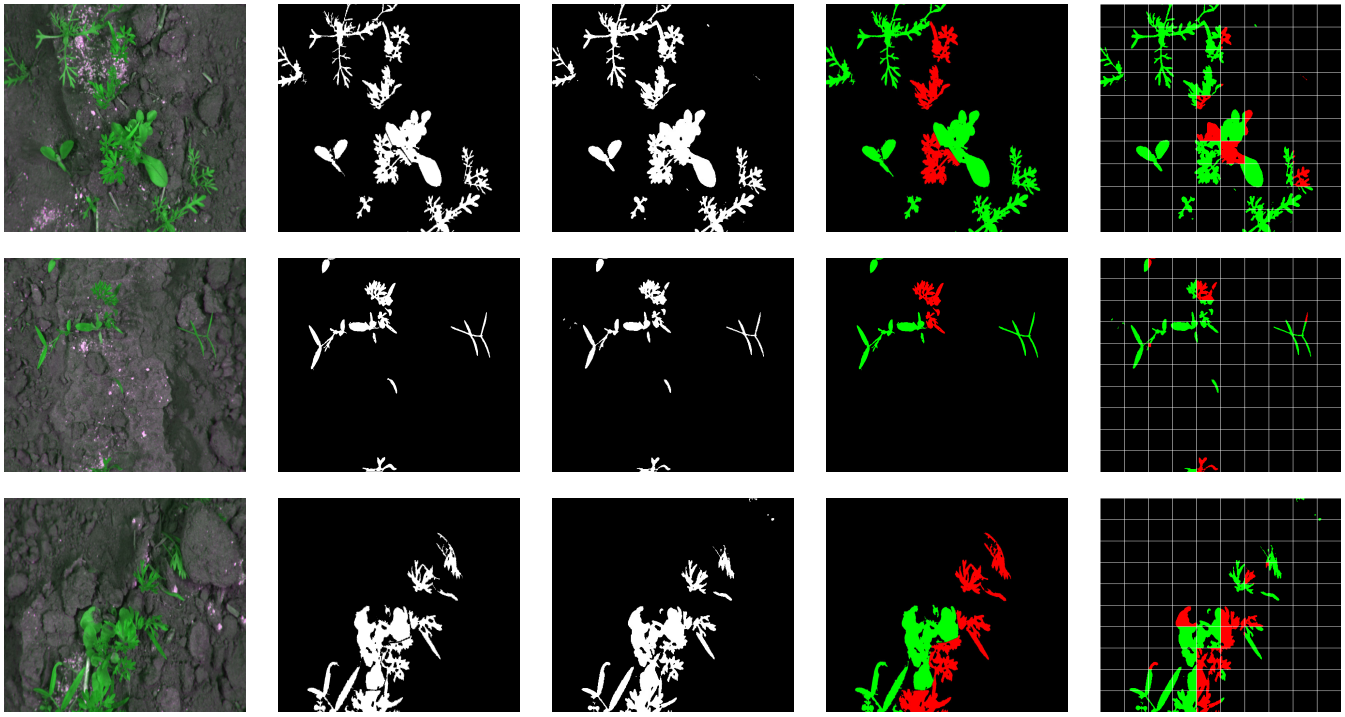


FIGURE 5. Result on the CWFID dataset: (left to right) (a) Input color image, (b) Ground truth vegetation mask, (c) Segmented vegetation mask, (d) Ground truth crop (red) and weed (green) pixels, (e) Predicted crop and weed pixels using fine-tuned ResNet50 as the classifier.

class are reported in Table 6. Weed precision is higher compared to feature-based classification with relatively higher recall value. On the other hand, the recall is lower for crop class but the precision is greater than 0.95 for weighted cross-entropy loss for both datasets. Besides, a recall of more than 0.91 is obtained for the weed class on both the datasets. This trend suggests that the network is likely to classify a region as weed-infested unless it is extremely confident that the region is free of any unwanted vegetation. This behaviour aligns with the objective to not overlook any weed-infested regions. It should be noted that given the larger number of crop tiles, this decision-making approach will correctly result in a large number of crop regions (with no weed plant coverage) not being treated with chemicals hence reducing their consumption significantly. Therefore, image-based classification using ResNet50 (with weighted cross entropy loss) is preferred over feature vector based classification. Figures 5 and 6 visualize the results for sample images from both the datasets.

E. WEED DENSITY ESTIMATION

Once the weed-infested regions are identified, the cluster rate for each tile can be computed from the segmented vegetation pixels. A comparison of the estimated cluster rate for the weed-infested regions with the ground truth values is presented in Table 7. The results show that the weed density can be estimated reasonably across both the datasets. The loss of weed density pixels can be attributed to four reasons: (1) discarding tiles or regions with vegetation cover less than 10% of the total area, (2) incorrect vegetation segmentation, (3) misclassification of weed-infested regions as crop plants and (4) presence of overlapping plants in a given tile. The first error source arises as the regions with low vegetation are ignored. The threshold, fixed at 10% in this study, can be varied for different crop and weed plant requirements. The results reported for vegetation segmentation and weed distribution show that the proposed approach results in a mean absolute error of 5% on CWFID and 1% on Sugar Beets datasets. This indicates that the proposed method can handle the error arising due to above listed sources reasonably. Moreover, the RMSE of less than 8% on two datasets of two different crop/weed species demonstrates that the proposed approach is scalable and can be adapted to any crop/weed species. Once the weed plant distribution and density have been estimated, it is possible to make a decision about which regions should be selectively treated with agrochemicals.

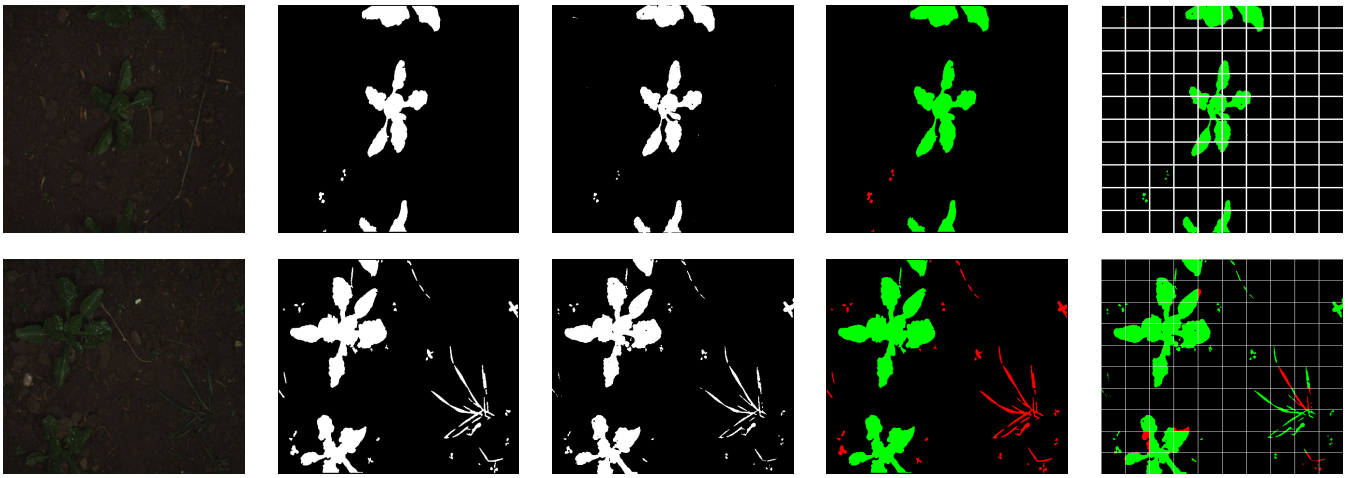


FIGURE 6. Result on the Sugar Beet Dataset: (left to right) (a) Input color image, (b) Ground truth vegetation mask, (c) Segmented vegetation mask, (d) Ground truth crop (red) and weed (green) pixels, (e) Predicted crop and weed pixels using fine-tuned ResNet50 as the classifier.

TABLE 7. Accuracy of weed density estimation. MAE and RMSE are for a region of 2500 pixels.

Dataset	Mean Accuracy (%)	MAE (%)	RMSE (%)
CWFID	75.24	5.02	7.85
Sugar Beets	82.13	1.62	3.06

F. COMPARISON OF PIXEL-WISE DENSE PREDICTIONS

Although the proposed method's focus is not to predict an accurate pixel-wise weed/crop segmentation, the patch wise predictions can be used to generate the same. Hence, we compare the accuracy of the predicted ground coverage using the F1 score metric (Equation 8).

The authors in [22], [23] proposed end-to-end segmentation networks for predicting dense crop/weed maps on Sugar Beets dataset. These methods report class-wise F1 scores where the maximum-minimum value for crop class is (0.9113, 0.9074), and for weed class is (0.8247, 0.7388). In comparison, our approach lags in terms of pixel-wise accuracy (maximum F1 value for crop class is 0.7153, and weed class is 0.3676). This can primarily be attributed to the reason that the end-to-end segmentation network can classify each pixel individually based on the features of its local area. However, in our approach, all the pixels belonging to a tile are classified as either crop or weed pixels regardless of the individual characteristics. In addition, we generate the vegetation segmentation, which contributes to error since few weed or crop pixels will be classified as background (soil) and vice-versa. On the other hand, segmentation networks have a single source of error as they segment and classify the pixels together.

However, for the purpose of selectively treating particular regions, the segmentation networks need to be augmented with an algorithm to select specific regions. If it is divided into regions such as square tiles, there is bound to be an overlap of weed and crop pixels for most of the tiles. The decision to treat a particular region will be taken from the dominant label for such tiles. Hence, the pixels which are correctly classified but are in the minority for a given tile do not influence the selective treatment. We argue that in the proposed approach, the focus is not on correctly identifying such pixels but correctly identifying the *regions* to be treated (which can be robustly estimated as shown previously). Besides, the volume of data required for the proposed method is significantly lower than that of an end-to-end segmentation network, enhancing generalizability and scalability. In addition, the proposed approach can be extended to any crop-weed combination as it eliminates the need to design hand-crafted features based on biological morphology and visual textures of the crop and weed. It may be noted that, to the best knowledge of the authors, there is no existing study on designing an end-to-end pixel-wise supervised CNN based segmentation approach on the CWFID dataset. One possible reason could be a limited number of images available in the CWFID dataset. However, encouraging results have been obtained on the CWFID dataset using the proposed tile-based semi-supervised approach.

It may also be noted that the proposed tile-based system can cover the entire area of the original image by assigning a label to each tile. Hence, eventually, the total area being analyzed is the same, whether it is pixel-wise or tile-wise classification.

V. CONCLUSION

Precision agriculture is described as a farmland management approach to maximize productivity and profits in a sustainable manner. Agrochemicals, such as weedicides, are an expensive input for farming in addition to being detrimental to the environment. Leveraging a computer vision system to identify regions for selective chemical treatment holds the potential to reduce their consumption significantly. In this paper, a semi-supervised approach to robustly estimate the

weed density and distribution to aid precision agriculture is presented. The proposed approach relies only on color images as input. The first step is to generate a binary vegetation mask by removing all the background pixels. An unsupervised network is used to cluster the pixels into either background or vegetation. The second step is to overlay the mask on the input color image and divide it into smaller regions (square tiles of side 50 pixels). These smaller regions are then classified as weed or crop. In this work, the performance of two types of classifiers are studied: a) classifiers such as SVM, Gaussian Naive Bayes, Neural Network, and Random Forest which uses a pre-trained ResNet50 as a feature extractor and b) a fine-tuned ResNet50. The proposed approach is validated on two datasets consisting of different crop/weed species - Crop/Weed Field Image [65] and Sugar Beets [66]. Weed infested regions are identified with a maximum recall of 0.99 and weed density in these regions is estimated with an accuracy of 82.13%.

One of the primary objectives of our work is to reduce the dependency on extensively annotated datasets. The use of unsupervised segmentation and pre-trained ResNet50 in the proposed work eliminates the need for designing a hand-crafted features for weed identification. Compared to previous approaches, it is shown that it is possible to estimate both the weed distribution and density without training an end-to-end *pixel-wise* segmentation network. Indeed, identification of weed-infected *regions* could also aid in design of a robust site-specific weed management system. The proposed pipeline is robust to varying plant species, overlapping plants, and images with poor contrast. This approach should help agricultural companies who are looking for low cost implementations as it requires very little training data and fine tuning. There is no need to invest in any extra sensors besides a regular RGB camera as long as there is a platform set up to collect top views of the plants.

One of the limitations of our work is the iterative nature of generating vegetation masks. Future work should aim to reduce the average number of iterations required by the unsupervised network to generate the vegetation mask. This would improve the time needed to process a single color image on-board an autonomous robot. Another future direction of this work is to extend the two-stage detection and localization approach to medical imaging for identifying diseases or lesions. Such an approach can also be taken for identifying crop diseases to further expand the scope of precision agriculture.

ACKNOWLEDGMENT

(Armaan Ashfaque and R. Sidharth contributed equally to this work.) The authors would like to thank Mars Rover Manipal for providing access to the computing resources to validate the proposed approach.

REFERENCES

[1] A. dos Santos Ferreira, D. M. Freitas, G. G. da Silva, H. Pistori, and M. T. Folhes, "Unsupervised deep learning and semi-automatic data labeling in weed discrimination," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104963.

[2] A. Wang, W. Zhang, and X. Wei, "A review on weed detection using ground-based machine vision and image processing techniques," *Comput. Electron. Agricult.*, vol. 158, pp. 226–240, Mar. 2019.

[3] P. Dureja, D. Patra, S. Johnson, and S. S. Tomar, "Effect of agrochemicals on earthworms," *Toxicol. Environ. Chem.*, vol. 71, nos. 3–4, pp. 397–404, Aug. 1999.

[4] M. A. Rodrigo, N. Oturan, and M. A. Oturan, "Electrochemically assisted remediation of pesticides in soils and water: A review," *Chem. Rev.*, vol. 114, no. 17, pp. 8720–8745, Sep. 2014.

[5] P. A. Adeoye, S. K. Abubakar, and A. R. Adeolu, "Effect of agrochemicals on groundwater quality: A review," *Scientia Agriculturae*, vol. 1, no. 1, pp. 1–7, 2013.

[6] N. Sharma and R. Singhvi, "Effects of chemical fertilizers and pesticides on human health and environment: A review," *Int. J. Agricult., Environ. Biotechnol.*, vol. 10, pp. 675–679, Dec. 2017.

[7] International Society of Precision Agriculture, Monticello, IL, USA. *Precision Ag Definition*. Accessed: Feb 9, 2021. [Online]. Available: https://www.ispag.org/about/definition

[8] R. Gebbers and V. I. Adamchuk, "Precision agriculture and food security," *Science*, vol. 327, no. 5967, pp. 828–831, Feb. 2010.

[9] N. Zhang, M. Wang, and N. Wang, "Precision agriculture—A worldwide overview," *Comput. Electron. Agricult.*, vol. 36, no. 2, pp. 113–132, 2002.

[10] Y. Lan, S. J. Thomson, Y. Huang, W. C. Hoffmann, and H. Zhang, "Current status and future directions of precision aerial application for site-specific crop management in the USA," *Comput. Electron. Agricult.*, vol. 74, no. 1, pp. 34–38, Oct. 2010.

[11] Y. Xiong, C. Peng, L. Grimstad, P. J. From, and V. Isler, "Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper," *Comput. Electron. Agricult.*, vol. 157, pp. 392–402, Feb. 2019.

[12] U. Verma, F. Rossant, and I. Bloch, "Segmentation and size estimation of tomatoes from sequences of paired images," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, pp. 1–23, Dec. 2015.

[13] U. Verma, F. Rossant, I. Bloch, J. Orensanz, and D. Boisgontier, "Shape-based segmentation of tomatoes for agriculture monitoring," in *Proc. 3rd Int. Conf. Pattern Recognit. Appl. Methods (ICPRAM)*, 2014, pp. 402–411.

[14] U. Verma, F. Rossant, I. Bloch, J. Orensanz, and D. Boisgontier, "Segmentation of tomatoes in open field images with shape and temporal constraints," in *Pattern Recognition Applications and Methods*, A. Fred, M. De Marsico, and A. Tabbone, Eds. Cham, Switzerland: Springer, 2015, pp. 162–178.

[15] W. McAllister, D. Osipychev, A. Davis, and G. Chowdhary, "Agbots: Weeding a field with a team of autonomous robots," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104827.

[16] B. S. Faiçal, H. Freitas, P. H. Gomes, L. Y. Mano, G. Pessin, A. C. P. L. F. de Carvalho, B. Krishnamachari, and J. Ueyama, "An adaptive approach for UAV-based pesticide spraying in dynamic environments," *Comput. Electron. Agricult.*, vol. 138, pp. 210–223, Jun. 2017.

[17] M. N. Abd. Kharim, A. Wayayok, A. R. M. Shariff, A. F. Abdullah, and E. M. Husin, "Droplet deposition density of organic liquid fertilizer at low altitude UAV aerial spraying in rice cultivation," *Comput. Electron. Agricult.*, vol. 167, Dec. 2019, Art. no. 105045.

[18] M. Hassanein, Z. Lari, and N. El-Sheimy, "A new vegetation segmentation approach for cropped fields based on threshold detection from hue histograms," *Sensors*, vol. 18, no. 4, p. 1253, Apr. 2018.

[19] E. Hamuda, M. Glavin, and E. Jones, "A survey of image processing techniques for plant extraction and segmentation in the field," *Comput. Electron. Agricult.*, vol. 125, pp. 184–199, Jul. 2016.

[20] S. Sabzi, Y. Abbaspour-Gilandeh, and G. García-Mateos, "A fast and accurate expert system for weed identification in potato crops using metaheuristic algorithms," *Comput. Ind.*, vol. 98, pp. 80–89, Jun. 2018.

[21] I. Sa, Z. Chen, M. Popovic, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "WeedNet: Dense semantic weed classification using multispectral images and MAV for smart farming," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 588–595, Jan. 2017.

[22] P. Lottes, J. Behley, A. Milioto, and C. Stachniss, "Fully convolutional networks with sequential information for robust crop and weed detection in precision farming," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2870–2877, Oct. 2018.

[23] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2229–2235.

[24] D. Andújar, V. Rueda-Ayala, H. Moreno, J. Rosell-Polo, C. Valero, R. Gerhards, C. Fernández-Quintanilla, J. Dorado, and H.-W. Griepentrog, "Discriminating crop, weeds and soil surface with a terrestrial lidar sensor," *Sensors*, vol. 13, no. 11, p. 14662–14675, Oct. 2013.

[25] D. Reiser, J. Martín-López, E. Memic, M. Vázquez-Arellano, S. Brandner, and H. Griepentrog, "3D imaging with a sonar sensor and an automated 3-axes frame for selective spraying in controlled conditions," *J. Imag.*, vol. 3, no. 1, p. 9, Feb. 2017.

[26] A. Olsen, D. A. Konovalov, B. Philippa, P. Ridd, J. C. Wood, J. Johns, W. Banks, B. Girgenti, O. Kenny, J. Whinney, B. Calvert, M. R. Azghadi, and R. D. White, "DeepWeeds: A multiclass weed species image dataset for deep learning," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 2058.

[27] P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss, "Robust joint stem detection and crop-weed classification using image sequences for plant-specific treatment in precision farming," *J. Field Robot.*, vol. 37, no. 1, pp. 20–34, Jan. 2020.

[28] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.

[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[31] P. Rasti, A. Ahmad, S. Samiei, E. Belin, and D. Rousseau, "Supervised image classification by scattering transform with application to weed detection in culture crops of high density," *Remote Sens.*, vol. 11, no. 3, p. 249, Jan. 2019.

[32] R. Raja, D. C. Slaughter, S. A. Fennimore, T. T. Nguyen, V. L. Vuong, N. Sinha, L. Tourte, R. F. Smith, and M. C. Siemens, "Crop signalling: A novel crop recognition technique for robotic weed control," *Biosyst. Eng.*, vol. 187, pp. 278–291, Nov. 2019.

[33] M. Dyrmann, R. N. Jørgensen, and H. S. Midtiby, "RoboWeedSupport–detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network," *Adv. Animal Biosci.*, vol. 8, no. 2, pp. 842–847, Jul. 2017.

[34] I. Sa, M. Popović, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, and R. Siegwart, "WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming," *Remote Sens.*, vol. 10, no. 9, p. 1423, Sep. 2018.

[35] M. Fawakherji, A. Youssef, D. Bloisi, A. Pretto, and D. Nardi, "Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation," in *Proc. 3rd IEEE Int. Conf. Robotic Comput. (IRC)*, Feb. 2019, pp. 146–152.

[36] C. Hung, Z. Xu, and S. Sukkarieh, "Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV," *Remote Sens.*, vol. 6, no. 12, pp. 12037–12054, Dec. 2014.

[37] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," in *Proc. 19th Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Cambridge, MA, USA: MIT Press, 2006, pp. 1137–1144.

[38] J. I. Goodfellow, V. Q. Le, M. A. Saxe, H. Lee, and Y. A. Ng, "Measuring invariances in deep networks," in *Proc. 22nd Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Red Hook, NY, USA: Curran Associates, 2009, pp. 646–654

[39] C. McCool, T. Perez, and B. Upcroft, "Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1344–1351, Jul. 2017.

[40] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[41] J. Yang, D. Parikh, and D. Batra, "Joint unsupervised learning of deep representations and image clusters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5147–5156.

[42] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proc. ECCV*, 2018, pp. 132–149.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[45] J. Tang, D. Wang, Z. Zhang, L. He, J. Xin, and Y. Xu, "Weed identification based on K-means feature learning combined with convolutional neural network," *Comput. Electron. Agricult.*, vol. 135, pp. 63–70, Apr. 2017.

[46] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[47] F. C. Rocha, A. M. O. Neto, E. L. Bottega, N. Guerra, R. P. Rocha, and C. C. Vilar, "Weed mapping using techniques of precision agriculture," *Planta Daninha*, vol. 33, no. 1, pp. 157–164, Mar. 2015.

[48] L. S. Shiratsuchi, P. J. Christoffoleti, and J. R. A. Fontes, "Aplicação localizada de herbicidas," in *Embrapa Cerrados-Documentos (INFOTECA-E)*. Brasilia, Brazil: Brazilian Agriculture Research Corporation, 2003.

[49] J. R. A. Fontes and L. S. Shiratsuchi, "Levantamento florístico de plantas daninhas em lavoura de milho cultivada no cerrado de Goiás," in *Embrapa Cerrados-Boletim de Pesquisa e Desenvolvimento (INFOTECA-E)*. Brasilia, Brazil: Brazilian Agriculture Research Corporation, 2005.

[50] D. Hall, F. Dayoub, J. Kulk, and C. McCool, "Towards unsupervised weed scouting for agricultural robotics," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 5223–5230.

[51] Y. Xu, R. He, Z. Gao, C. Li, Y. Zhai, and Y. Jiao, "Weed density detection method based on absolute feature corner points in field," *Agronomy*, vol. 10, no. 1, p. 113, Jan. 2020.

[52] Y. Xu, Z. Gao, L. Khot, X. Meng, and Q. Zhang, "A real-time weed mapping and precision herbicide spraying system for row crops," *Sensors*, vol. 18, no. 12, p. 4245, Dec. 2018.

[53] G. Bradski, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, 2000.

[54] A. Kanezaki, "Unsupervised image segmentation by backpropagation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1543–1547.

[55] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[56] A. M. Hearst, "Support vector machines," *IEEE Intell. Syst.*, vol. 13, no. 4, pp. 18–28, Jul. 1998.

[57] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[58] P. Norvig and S. Russell, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.

[59] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[60] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.

[61] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: https://arxiv.org/abs/1803.08375

[62] M. Feng, L. Wan, Z. Li, L. Qing, and X. Qi, "Fetal weight estimation via ultrasound using machine learning," *IEEE Access*, vol. 7, pp. 87783–87791, 2019.

[63] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[64] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 2. Cambridge, MA, USA: MIT Press, 2014, pp. 3320–3328.

[65] S. Haug and J. Ostermann, "A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks," in *Computer Vision—ECCV Workshops*, L. Agapito, M. M. Bronstein, C. Rother, Eds. Cham, Switzerland: Springer, 2015, pp. 105–116.

[66] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, "Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields," *Int. J. Robot. Res.*, vol. 36, no. 10, pp. 1045–1052, Sep. 2017.

[67] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

SHANTAM SHOREWALA received the B.Tech. degree in electronics and communication engineering from the Manipal Institute of Technology, Manipal Academy of Higher Education, India. His research interests include deep learning methods for image segmentation, 3D reconstruction, and object pose estimation.

ARMAAN ASHFAQUE is currently pursuing the B.Tech. degree in information technology with the Manipal Institute of Technology, Manipal Academy of Higher Education, India. His research interests include deep learning for computer vision and robotics.

R. SIDHARTH is currently pursuing the B.Tech. degree in computer science engineering with the Manipal Institute of Technology, Manipal Academy of Higher Education, India. His research interests include deep learning for computer vision, image processing, and robotics.

UJJWAL VERMA (Senior Member, IEEE) received the M.S. (Research) degree in signal and image processing from IMT Atlantique, France, and the Ph.D. degree in image processing from the Télécom ParisTech, University of Paris-Saclay, Paris, France. He is currently an Associate Professor with the Department of Electronics and Communication Engineering, Manipal Institute of Technology, India. His research interests include variational methods in image segmentation, action recognition, and deep learning methods for scene understanding. He was a recipient of ''ISCA Young Scientist Award 2017–2018'' by Indian Science Congress Association (ISCA), a professional body under the Department of Science and Technology, Government of India. He was the Joint Secretary, IEEE Mangalore Sub-Section for the year 2019. He was also a recipient of ''Young Professional Volunteer Award 2020'' by IEEE Mangalore Sub-Section in recognition of his outstanding contribution to IEEE activities.