Crop and Weeds Classification for Precision Agriculture using Context-Independent Pixel-Wise Segmentation

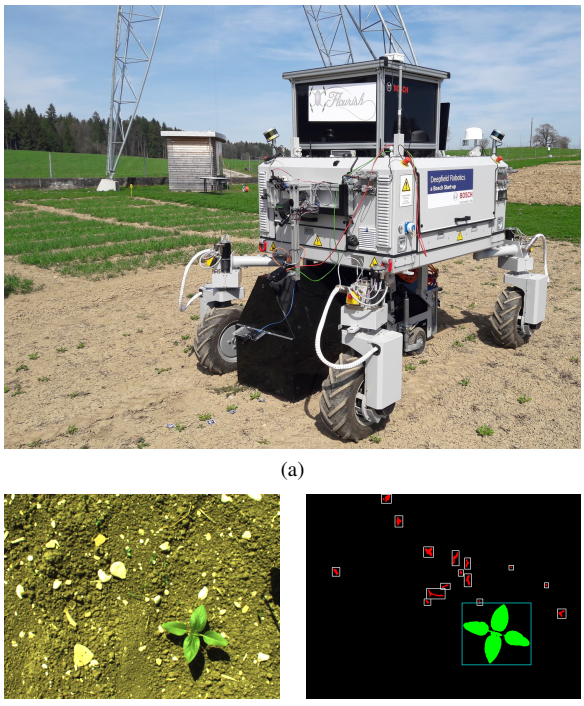
Mulham Fawakherji*, Ali Youssef*, Domenico D. Bloisi†, Alberto Pretto* and Daniele Nardi*
*Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Rome, Italy
Email: nardi@diag.uniroma1.it
†Department of Mathematics, Computer Science, and Economics, University of Basilicata, Potenza, Italy
Email: domenico.bloisi@unibas.it

***Abstract*—Precision agriculture is gaining increasing attention because of the possible reduction of agricultural inputs (e.g., fertilizers and pesticides) that can be obtained by using high-tech equipment, including robots. In this paper, we focus on an agricultural robotics system that addresses the weeding problem by means of selective spraying or mechanical removal of the detected weeds. In particular, we describe a deep learning based method to allow a robot to perform an accurate weed/crop classification using a sequence of two Convolutional Neural Networks (CNNs) applied to RGB images. The first network, based on an encoder-decoder segmentation architecture, performs a pixel-wise, plant-type agnostic, segmentation between vegetation and soil that enables to extract a set of connected blobs representing plant instances. We show that such network can be trained also using external, ready to use pixel-wise labeled data sets coming from different contexts. Each plant is hence classified between crop and weeds by using the second network. Quantitative experimental results, obtained on real world data, demonstrate that the proposed approach can achieve good classification results also on challenging images.**

I. Introduction

Autonomous robotics applications for precision agriculture represent a concrete solution towards a sustainable agriculture and chemical treatments reduction [1]. The term *crop* defines the cultivated plant, while the term *weeds* defines unwanted plants that grow spontaneously in the field. Precision weed control is a challenging task that aims to reduce the amount of herbicides without compromising the quality of crops. It can be achieved by selective spraying or accurate mechanical removal of weeds, while achieving that manually is time-consuming and expensive.

Autonomous robots equipped with automatic weed detection systems can be used to improve the efficiency of precision farming techniques on weed control by modulating herbicide spraying appropriately to the level of weeds infestation. However, the great variety of crop and weeds shapes, size and colors, together with the presence of overlapping between plants, makes automatic crop/weeds classification throughout images a challenging task for autonomous farming robots [2]. Nevertheless, the capability to generalize the trained models still remains an obstacle to employ farming robots in different farm conditions, e.g., caused by environmental changes, plants characteristics and types of soil, as highlighted in [3], [4].



(a) (b) (c)

Fig. 1: (a) The robot used to acquire the data sets used in the experiments. (b) An example of RGB image provided by the camera mounted on the robot. (c) Label mask with bounding boxes predicted by our approach for image (b); pixels that belong to crop are colored in green while pixels that belong to weeds are colored in red.

In this paper, we present a novel approach for combining robust pixel-wise segmentation with a supervised image classification based on Convolutional Neural Networks (CNNs) applied to RGB images acquired by a farming robot on a sunflower field (see Fig. 1). In particular, we use a deep convolutional encoder-decoder architecture for robust semantic

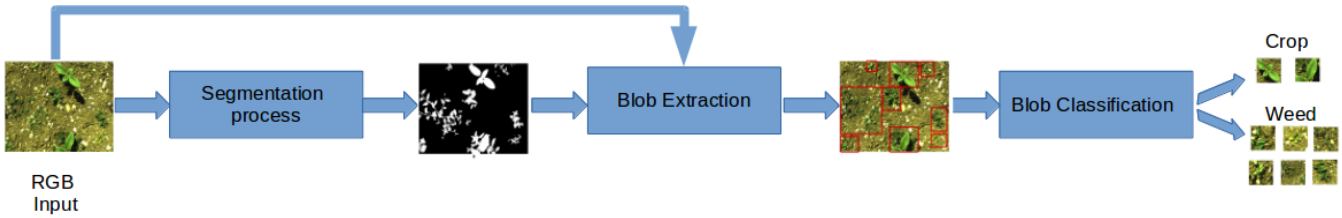


Fig. 2: The proposed three-steps approach. The first step is a binary pixel-wise segmentation (i.e., soil/plant) of the RGB input image. The second step concerns the extraction of the image patches to be classified. Crop/weed classification is carried out in the third step.

pixel-wise segmentation, background removal and the extraction of regions of interest (ROIs). The chosen network is based on the UNet architecture [5] with a modified VGG-16 encoder [6] followed by a binary pixel-wise classification layer. A coarse-to-fine classifier based on CNN is used to classify the extracted ROIs into crop and weeds.

The main contributions of this work are:

- A background removal method that uses a deep pixel-wise segmentation to distinguish between soil and plants.
- An accurate crop/weed classifier based on a deep CNN.

The aim of the proposed approach is to reduce the limitations of CNNs in generalizing when a limited amount of data with pixel-wise annotations is available: pixel-wise labeling is in fact the bottleneck for most crop/weeds classification methods. Our method relies on a robust binary segmentation that is agnostic to plant species, so easily trainable also by using external, ready to use pixel-wise labeled data sets that possibly do not includes the target crop. The classification between crop and weeds is then obtained feeding a classification CNN with image patches (i.e., bounding boxes) enclosing plant instances. The generation of specialized data sets for such a CNN is a simpler and faster operation compared with respect to the generation of pixel-wise annotated data sets.

The reminder of the paper is organized as follows. Section II contains a discussion of similar approaches present in the literature. Section III describes the details of the proposed method, while Section IV shows both qualitative and quantitative results obtained on publicly available data. Finally, conclusions are drawn in Section V.

II. Related Work

The problem of vision based crop and weeds classification has been addressed in different ways. Handcrafted features are used, among others, in [7], [8]. De Rainville *et al.* [7] present an unsupervised classification method based on morphological features extracted taking into account the spatial localization of vegetation in the field. Haug *et al.* [8] present a method to classify carrot plants and weeds from RGB and near-infrared (NIR) images that uses a background removal step based on the Normalized Difference Vegetation Index (NDVI) and a Random Forest classifier applied to features extracted at sparse pixel positions. This approach has been extended in [2], where a plant arrangement prior is added to the features list used for classification, and tailored for UAVs (Unmanned Aerial Vehicles) applications in [9].

The adoption of deep CNNs in overcoming the limitations of handcrafted features has been explored, among others, in [4], [10], [11]. Fine-tuned pre-trained CNN models are used for plant classification of 44 different species in [10]. Potena *et al.* [11] proposes an on-line perception system for weed-crop classification that uses a cascade of two different CNNs: a shallow CNN performs vegetation detection, while a second, slightly deeper CNN discriminates between weeds and crops. Encoder-decoder architectures such as the SegNet segmentation network [12] are used in [3], [13], [14]. In [13], a SegNet network is fed with 3 channels images that include the NIR channel, the red channel from the RGB image, and the NDVI map. A similar approach is exploited in [3], where 14 channels images that include several vegetation indices are used as input for a modified version of the SegNet network. Procedurally generated synthetic training data sets are used to train a SegNet network in [14], by randomizing the key features of the target environment (i.e., crop and weed species, type of soil, and light conditions). The fully convolutional network (FCN) proposed in [15] is employed in [4], [16]. In [4], authors exploits the crop arrangement as a further source of information, by analyzing image sequences that cover a portion of the field. Class-wise stem detection and pixel-wise crop/weeds semantic segmentation is jointly addressed in [16]. Model compression and mixtures of lightweight CNNs are exploited in [17] to learn from a very deep, pre-trained model a lighter model which allows real-time weed segmentation also for robots with limited computing power. Multi-spectral features and 3D surface features are exploited for plant classification in [18].

III. Methods

We propose to perform crop/weed classification with a three-step procedure (see Fig. 2).

Segmentation process. To remove the background (i.e., the soil), we firstly apply a robust pixel-wise soil/plant segmentation of the RGB image in input. We use a modified version of the UNet semantic segmentation network [5], which is composed by a contracting encoder along with a symmetric expanding decoder. In our implementation, the contracting path consists of a VGG-16 structure modified by removing the last fully connected layers and fine-tuning the other layers. The indices of spatial information in the pooling operations

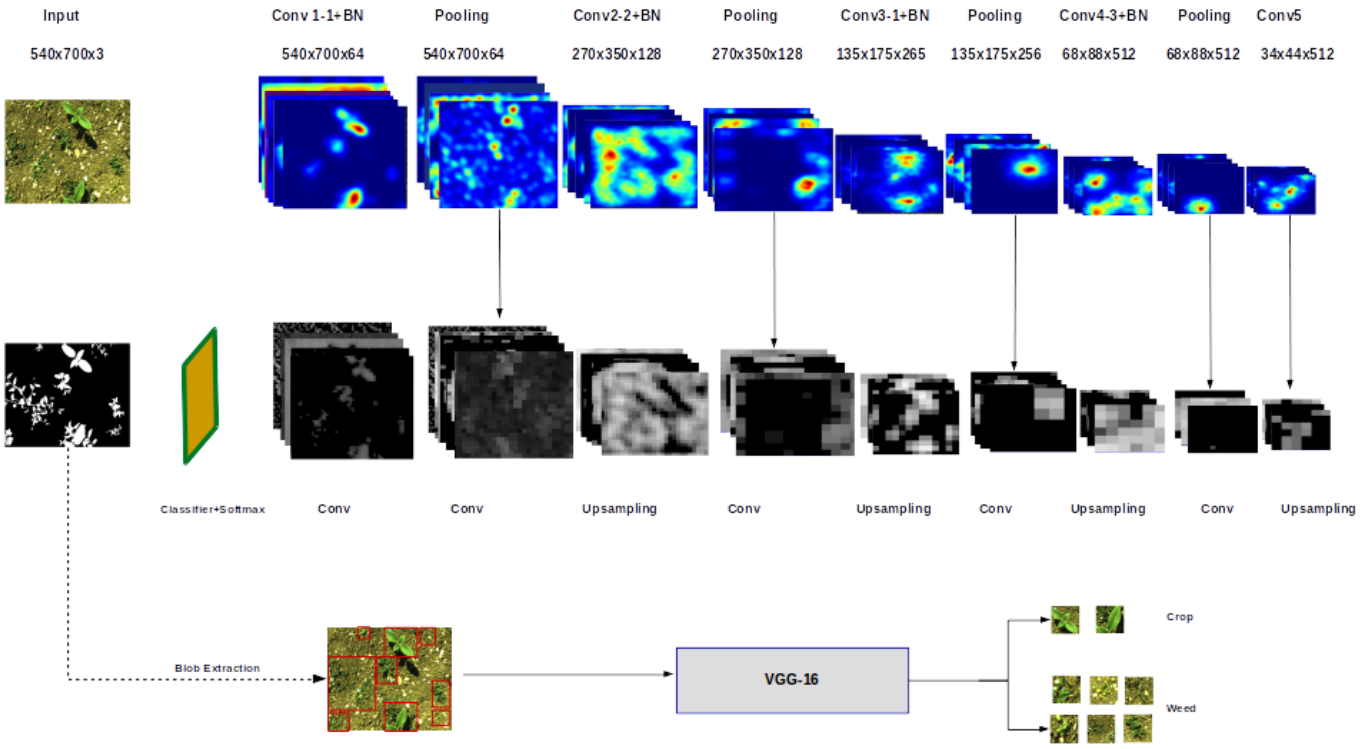


Fig. 3: The classification pipeline.

are spread through the expansive path, which contains a sequence of up-convolution operations of features encoded in the contracting path. The expanding decoder is designed with 4-convolutional layers, where each layer is composed of a batch normalization, 4-upsampling layers and a soft-max pixel-wise classifier. Between the contracting and expanding paths, there is a bottleneck consisting of two convolutional layers combined with batch normalization and a dropout activation function.

The lack of pixel-wise annotated data sets for each possible crop type and for different field conditions can lead to strong challenges in generalizing an end-to-end crop/weeds segmentation network. The goal of the first step is to obtain a robust binary segmentation mask that enables to generate blobs corresponding to vegetation pixels in the RGB image, so to simplify the following classification step. This is obtained by exploiting the similarities of various plants properties instead of differentiate between them. The idea is to train the first segmentation network by exploiting external, ready to use data sets coming from different contexts, containing different plants categories, types of fields, and captured under varying environmental conditions. This context-independent training possibly enables to avoid to pixel-wise annotate large amount of data acquired in the target filed, an operation that usually requires a lot of manual work.

Blob extraction. The second step concerns the extraction of the image ROIs to be classified. This is obtained by extracting the vegetation blobs contained in the binary mask generated during the segmentation process. In this stage, the input consists in the original RGB image plus the binary mask generated during the segmentation process. A dilation operator is applied to the binary mask to gradually increase the boundaries of the foreground regions (i.e., the areas containing vegetation pixels) to reduce the holes between those regions. Then, the connected blobs from the dilated mask are extracted, and a bounding box for each blob is determined. Finally, a set of patches from the original RGB image corresponding to the bounding boxes is generated.

Classification. A deep CNN for crop/weed classification is employed in the third step. The image patches identified in the previous step are fed to the CNN classifier, which is based on a fine-tuned model of VGG-16 exploiting the ability of deep CNN in object classification. The VGG-16 network architecture for object classification is used as encoder. The network consists of 13-convolutional layers with a kernel of 3×3. A max-pooling operation with a kernel of 2×2 with a stride of 2 for down-sampling. Batch normalization and a ReLU activation function are used too. This step just requires a training data set that includes labeled patches with positive and negative examples of the target crop. The annotation of such training data set just requires to specify a label for each image, that is a much faster operation than a pixel-wise annotation.

Fig. 3 shows the classification pipeline. To create the figure, we have used a fine-tuned VGG-16 encoder model at early step of the training, showing randomly picked up filters to illustrate the ability of the network to learn weights based on neurons

Algorithm 1 The proposed crop/weed classification algorithm

Input: RGB image I_{RGB}
Result: A set of classified blobs B_c
$M \leftarrow$ Segmentation of I_{RGB} using VGG-UNet
$C \leftarrow$ Contour-Extraction(M) ▷ set of contours belonging to connected regions
for i `in range len(`C`)` **do**
    $B_M[i] \leftarrow$ `BoundRect(`$C[i]$`)` ▷ $B_M[i]$ is the bounding box around the contour i
    $B_{RGB}[i] \leftarrow (I_{RGB} \cap B_M[i])$ ▷ $B_{RGB}[i]$ is the corresponding bounding box from RGB image
    $B_c[i] \leftarrow$ (classify $B_{RGB}[i]$ using VGG-16 into weed or crop)
end for

responses to image pixels (e.g., soil/plant pixel).

The full crop/weeds classification algorithm is reported in Algorithm 1.

IV. Experimental Results

For training our deep CNNs, we have used a data set recorded in a sunflower farm in Italy over a period of one month in spring 2016. To demonstrate the ability of the proposed approach in generalizing, in addition to the *sunflower* data set, we have considered also two other publicibly available data sets, which contains images about *carrots* [19] and *sugar beets* [20]. Carrots and sugar beets data sets were acquired in different field conditions, different species of plants, and at different stages of plant level with respect to the data set used for training.

The experiments at first aim to demonstrate the performance of different deep CNN architecture on pixel-wise segmentation in order to classify pixels in image into three classes, namely soil, crop, and weed. Then, the same network architectures are used to measure the performance on background removal (e.g., pixel classification into only two classes soil and plants). To this end, we use only RGB images as input to the recent state-of-the-art models SegNet [12] based on VGG-16 encoder, UNet, UNet based on VGG-16 decoder (VGG-UNet), BonNet [21] and fully connected network FCN8 [22]. In the same manner of [3], we have filtered the blobs that have less than 50 pixels with an input image size of 512×384.

To improve the segmentation performance, we increase the number of input channels by a set of vegetation indices: Excess Green (ExG), Excess Red (ExR), Color Index of Vegetation Extraction (CIVE), and Normalized Difference Index (NDI). Those indicators proved their ability to segment vegetation and they do not present high sensitivity to soil types or weather condition [3]. In addition to the previous inputs, we use the HSV (hue, saturation, value) representation of the input image, concatenating all those representations along with the input RGB image to form a multi-channel input volume.

A. Network Training

Segmentation. We trained the proposed VGG-UNet by initializing the encoder (VGG-16) with the weights taken from training the VGG-16 on the ImageNet data set, then we trained the whole network using Stochastic Gradient Descent (SGD) with a fixed learning rate of $1 \cdot e^{-4}$ and a momentum of 0.90. The parameters of the network are updated in a way that cross entropy loss is reduced. Mini-batches composed by one image were used for training.

Classification. We used the VGG-16 architecture, pre-trained on the ImageNet data set, from which we removed the fully connected layer from the top of the model and we used the rest of the model as feature extractor from our data set (bottleneck features). We then run this model on our training and validation data once, recording the output from the last activation maps before the fully-connected layers in two arrays. We trained a fully-connected model on top of the stored features. This allowed us to reach a validation accuracy of $0.93 - 0.94$ in two minutes using a single NVIDIA GTX 1070 GPU. This enables our classifier to adapt quickly and easily with new data sets. We used the RMSprop optimizer with learning rate of 0.001 and the binary cross entropy as loss function.

B. Training Data

Segmentation. The training data set is made of a set of 500 images acquired in a sunflower field by a custom-built agricultural field robot. The data set is labeled with three classes (i.e., soil, crop and weeds). Ground truth annotations consist in binary masks generated via manual segmentation. An intensity value of 1 in the binary masks corresponds to the segmented crop, 2 to segmented weeds, and pixels with 0 value correspond to the background soil. Data augmentation was performed using rotations, horizontal and vertical flipping, and zooming. The final data set was composed by 2000 images, divided into 1500 images for training, 350 images for validation, and 150 images for test.

Classification. The training data set for classification was generated from the same data set used for training the segmentation model. In particular, we extracted 1600 sunflower batches and 1500 weeds batches.

C. Testing Data

Segmentation. The data set used for the training procedure (Sec. IV-B) does not present all the challenges introduced when dealing with a real-world field, because it does not contain data captured with different field conditions and at different stages of plant level. For this reason, in order to properly evaluate our approach, we used for testing 100 images coming from the so-called sugar beets data set [20] and other 60 labelled images acquired on a commercial organic carrot farm [19] (the so-called carrots data set). It is important to note that carrots and sugar beets data sets were not included in fine tuning VGG-16 encoder and were used only to evaluate the capability of VGG-UNet to generalize.

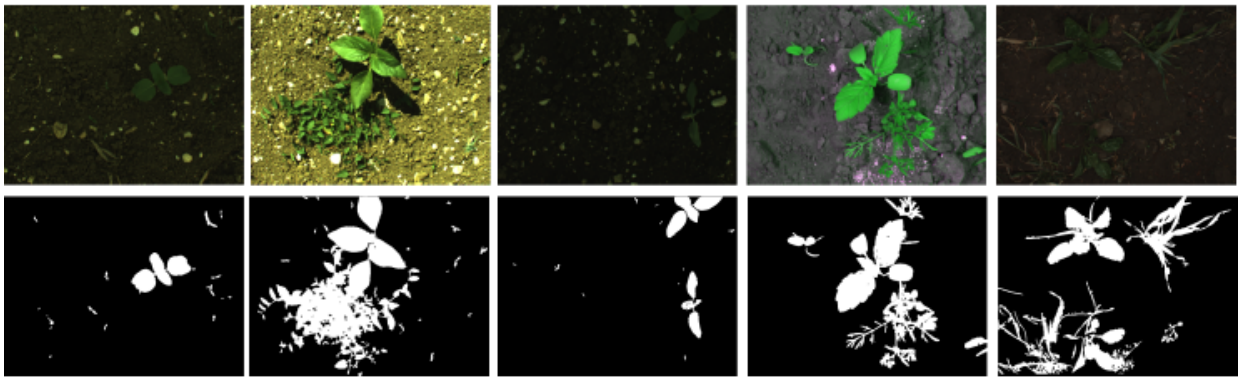


Fig. 4: Samples from the data set used for training and testing. The first row contains the RGB images in input, while the second row shows the ground truth masks. In the first, second, and third columns sunflower images taken under different lighting condition and different age of growth are shown. The fourth column refers to the organic carrots data set and the fifth column refers to the sugar beets data set.

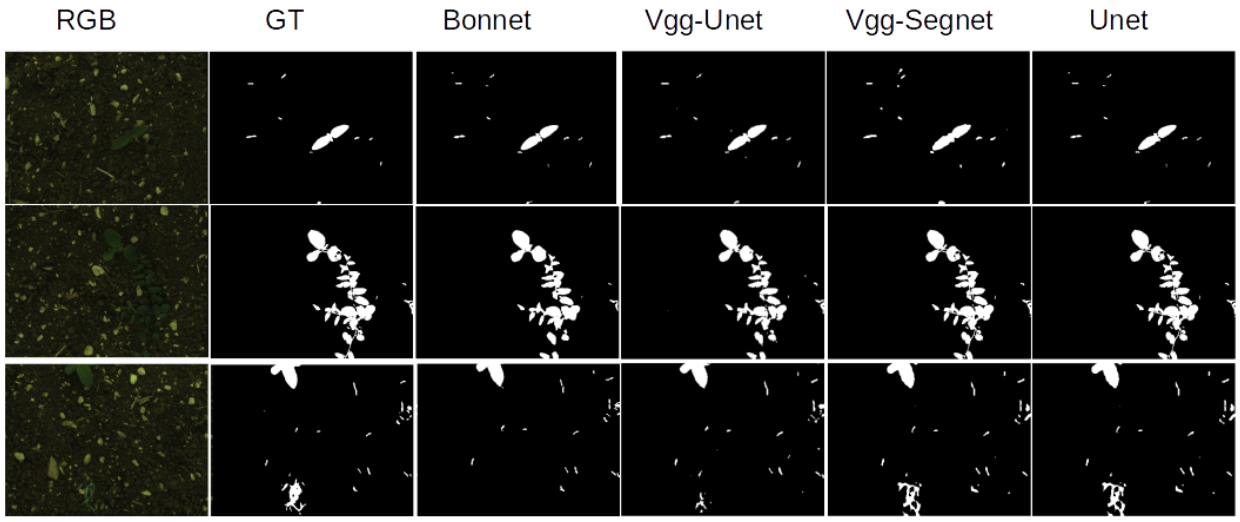


Fig. 5: Qualitative results achieved by different CNN structures. First column RGB images, second column ground truth mask, third, forth, fifth, and sixth prediction from Bonnet, VGG-UNet,VGG-Segnet, and UNet.

TABLE I: Quantitative results illustrating $mIOU$ obtained by different networks architectures on the sunflowers data set

Architecture	3-classes	2-classes
VGG-SegNet	0.68	0.90
UNet	0.62	0.90
BonnNet	**0.80**	0.90
FCN8	0.31	0.45
VGG-UNet	0.64	**0.91**

Fig. 4 shows images from the three different data sets and their masks.

Classification. We tested the fine-tuned VGG-16 binary classifier on a data set consisting of 150 sunflower batches and 150 weed batches. The batches were of various size and we resized them to 64×64 pixels.

D. Evaluation

Qualitative results of using different network architectures for pixel-wise segmentation (background removal) on three different input images are shown in Fig. 5.

The metric used in the evaluation procedure is the Mean Intersection-Over-Union (mIOU), which is a common metric used in evaluating image segmentation performance [21]. IOU is computed as follows:

$$mIOU = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_j}{TP_j + FP_j + FN_j} \quad (1)$$

where TP stands for True Positive, FP for False Positive, FN for False Negative, and C is the total number of classes.

Table I shows the results obtained by using different network architectures on the sunflower data set. When consid-

TABLE II: mIOU obtained using a multi-input for the segmentation model

Architecture	sugar beets	carrots	sunflowers
VGG-UNet	0.75	0.51	0.91
Multi-input VGG-UNet	**0.80**	**0.86**	**0.92**

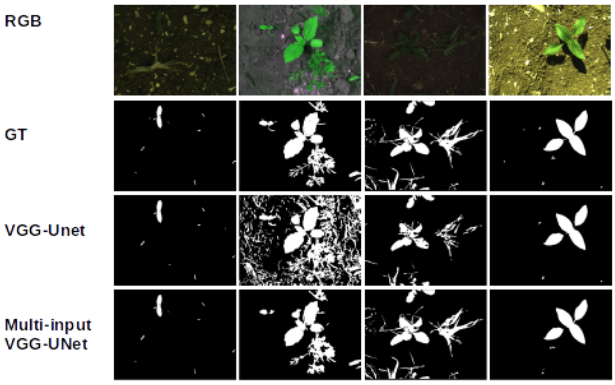


Fig. 6: Qualitative results obtained by VGG-UNet and multi-input VGG-UNet.

ering only two classes, namely soil/plants, the VGG-UNet approach outperforms the other tested approaches.

Augmenting the number of input channels to VGG-UNet (multi-input VGG-UNet) by using additional vegetation indices improves the segmentation performance and the ability of the VGG-UNet to handle data sets under various farm conditions, and plant and soil types. Quantitative results are given in Table II, showing how the segmentation performance of VGG-UNet decreased on the carrot data set, where images were acquired under different lightning conditions and different type of soil, and the improvement obtained by using a multi-input scheme. Fig. 6 shows the qualitative results obtained by applying multi-input channels to VGG-UNet.

We have evaluated the binary-classifier using four metrics, namely *accuracy*, *sensitivity*, *specificity* and *precision*. The target classes are sunflower and weed. Fig. 7 illustrates the classification performance of the fine-tuned VGG-16 model, where red rectangles indicate wrong classified batches and green ones are the correct classified. Table III shows the quantitative results for sunflower/weed classification.

Full Pipeline. Examples of the application of the full pipeline (i.e., segmentation plus classification) can be seen in Fig. 8[1].

The evaluation of the full pipeline based on object-wise classification accuracy is given in the form of the confusion matrix shown in Fig. 9. The result for correctly detected crops was 87%, while the 13% of crop detected as weed is mainly due to the overlapping problem between weeds and crop. The 32% of soil detected as weeds is due to inaccuracies in the binary masks coming from the segmentation process: the

[1]The results on the complete set of test images can be seen in a video available at https://youtu.be/cBYtL6Zp5bQ

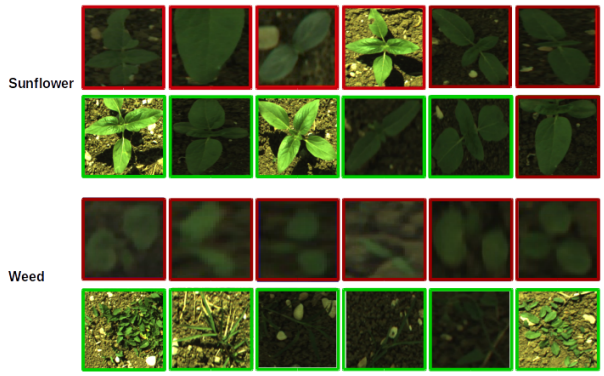


Fig. 7: Qualitative results for the classification process.

TABLE III: Quantitative results of the CNN based classifier

Class	Accuracy	Sensitivity	Specificity	Precision
Weed	0.90	0.94	0.87	0.88
Sunflower	0.90	0.87	0.94	0.92

dilation operation carried out to increase the boundaries of the foreground regions during the blob detection process increases the number of soil pixels included in weeds blobs.

V. Conclusions

In this paper, we present a crop/weeds classification approach based on a three-steps procedure. The first step is a robust pixel-wise segmentation (i.e., soil/plant) and image patches containing plants are extracted in the second step. In the third step, a deep CNN for crop/weed classification is used. The extracted blobs in the masked image containing plants information are fed to a CNN classifier based on a fine-tuned model of VGG-16 exploiting the ability of deep CNN in object classification.

Our goal is to reduce the limitations of CNNs in generalizing when a limited amount of data is available. In fact, our segmentation method is not based on plant types, but instead it possibly can use images coming from different types of crops and soils, which are usually easier to obtain in large quantities. The classification step can then be specialized to the types of plants needed by the application scenario.

The proposed approach has been tested on real-world images coming from three different data sets. In particular, we have quantitatively evaluated the segmentation and the classification processes separately. Then, we have evaluated the complete pipeline, including the first background removal phase and the subsequent classification stage. Experimental results demonstrate that the proposed approach can achieve good classification results on challenging data.

As future directions, we aim to insert between the segmentation and classification steps an automatic alignment process to improve the classification accuracy of our pipeline.

References

[1] T. Duckett, S. Pearson, S. Blackmore, B. Grieve, P. Wilson, H. Gill, A. Hunter, and I. Georgilas, *Agricultural Robotics: The Future of*

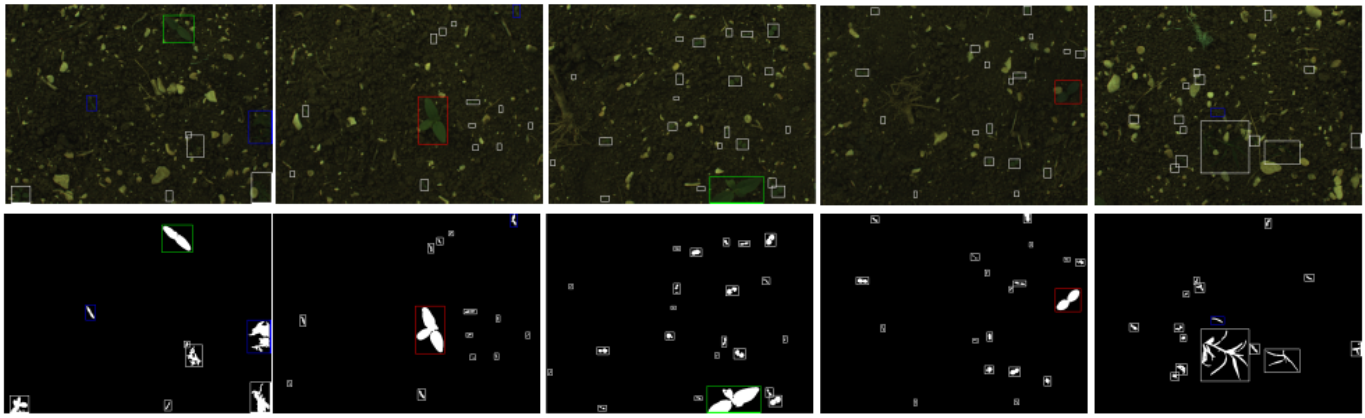


Fig. 8: Qualitative results achieved by the full pipeline (i.e., segmentation plus classification). White boxes denote true positive samples (weed), green boxes true negative samples (crop), blue boxes false positive samples, and red false negative samples.

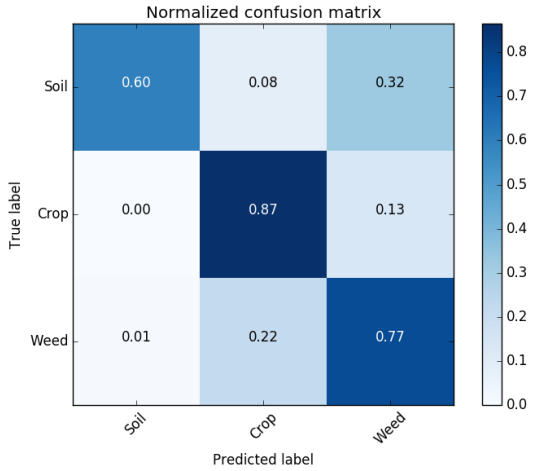


Fig. 9: Confusion matrix obtained by evaluating the full pipeline (i.e., segmentation plus classification).

Robotic Agriculture, ser. UK-RAS White Papers. UK-RAS Network, 2018.

[2] P. Lottes, M. Hoeferlin, S. Sander, M. Müter, P. Schulze, and L. C. Stachniss, "An effective classification system for separating sugar beets and weeds for precision farming applications," in *ICRA*, 2016, pp. 5157–5163.

[3] 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 *ICRA*, 2018, pp. 2229–2235.

[4] 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 Robotics and Automation Letters (RA-L)*, vol. 3, pp. 3097–3104, 2018.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[7] F.-M. De Rainville, A. Durand, F.-A. Fortin, K. Tanguy, X. Maldague, B. Panneton, and M.-J. Simard, "Bayesian classification and unsupervised learning for isolating weeds in row crops," *Pattern Analysis and Applications*, vol. 17, no. 2, pp. 401–414, 2014.

[8] S. Haug, A. Michaels, P. Biber, and J. Ostermann, "Plant classification system for crop/weed discrimination without segmentation," in *WACV*. IEEE, 2014, pp. 1142–1149.

[9] P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss, "Uav-based crop and weed classification for smart farming," in *ICRA*, 2017, pp. 3024–3031.

[10] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in *ICIP*. IEEE, 2015, pp. 452–456.

[11] C. Potena, D. Nardi, and A. Pretto, "Fast and accurate crop and weed identification with summarized train sets for precision agriculture," in *International Conference on Intelligent Autonomous Systems*, 2016, pp. 105–121.

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[13] I. Sa, Z. Chen, M. Popović, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "weednet: Dense semantic weed classification using multispectral images and mav for smart farming," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 588–595, 2018.

[14] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto, "Automatic model based dataset generation for fast and accurate crop and weeds detection," in *IROS*, 2017, pp. 5188–5195.

[15] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[16] P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss, "Joint stem detection and crop-weed classification for plant-specific treatment in precision farming," *arXiv preprint arXiv:1806.03413*, 2018.

[17] C. McCool, T. Perez, and B. Upcroft, "Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1344–1351, 2017.

[18] W. Strothmann, A. Ruckelshausen, J. Hertzberg, C. Scholz, and F. Langsenkamp, "Plant classification with in-field-labeling for crop/weed discrimination using spectral features and 3d surface features from a multi-wavelength laser line profile system," *Computers and Electronics in Agriculture*, vol. 134, pp. 79–93, 2017.

[19] 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 2014 Workshops*, 2015, pp. 105–116.

[20] 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," *The International Journal of Robotics Research*, 2017.

[21] A. Milioto and C. Stachniss, "Bonnet: An open-source training and deployment framework for semantic segmentation in robotics using cnns," *arXiv preprint arXiv:1802.08960*, 2018.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.