RoWeeder: Unsupervised Weed Mapping through Crop-Row Detection

Pasquale De Marinis⬡, Gennaro Vessio⬡, and Giovanna Castellano⬡

University of Bari Aldo Moro, Bari, Italy
`{pasquale.demarinis,gennaro.vessio,giovanna.castellano}@uniba.it`

Abstract. Precision agriculture relies heavily on effective weed management to ensure robust crop yields. This study presents RoWeeder, an innovative framework for unsupervised weed mapping that combines crop-row detection with a noise-resilient deep learning model. By leveraging crop-row information to create a pseudo-ground truth, our method trains a lightweight deep learning model capable of distinguishing between crops and weeds, even in the presence of noisy data. Evaluated on the WeedMap dataset, RoWeeder achieves an F1 score of 75.3, outperforming several baselines. Comprehensive ablation studies further validated the model's performance. By integrating RoWeeder with drone technology, farmers can conduct real-time aerial surveys, enabling precise weed management across large fields. The code is available at: `https://github.com/pasqualedem/RoWeeder`. [1]

Keywords: Precision Agriculture, Crop-Row Detection, Weed Mapping, Deep Learning, UAVs

1 Introduction

Agriculture is essential for human sustenance, and advancements in farming machinery and techniques have improved crop yield. *Weed management* is crucial for removing unwanted plants that compete with crops. Effective weed management enhances crop productivity and promotes sustainable agriculture.

Drones, or unmanned aerial vehicles (UAVs), have proven invaluable assets in precision agriculture, offering both versatility and cost-effectiveness [34]. These devices can capture high-resolution imagery and data from farmlands, enabling farmers to monitor crop development, detect diseases and pests, and fine-tune irrigation strategies. By providing accurate and timely information, drones help reduce expenses, boost crop yields, and minimize the use of inputs such as water, fertilizers, and pesticides. Traditional methods like manual field inspections or satellite-based remote sensing cannot match the level of detail drones provide. UAVs can conduct real-time aerial surveys of crops, allowing farmers to make prompt, informed decisions. Moreover, drones can efficiently cover vast areas

---

[1] Presented at the Computer Vision for Plant Phenotyping and Agriculture (CVPPA) workshop at ECCV 2024.

in a matter of hours, a task that would typically require days or weeks using conventional approaches. This time-saving capability enables farmers to conserve resources and make more timely decisions.

Recent research has suggested that deep learning models can be used for semantic segmentation and weed identification in drone-captured or aerial imagery [8, 10, 28]. However, despite these significant advancements, automatically detecting weeds remains a complex challenge. Deep learning techniques are not yet widely adopted in agriculture, mainly due to the extensive manual annotation required for the large amount of data needed in the learning phase. This issue is particularly pronounced in agricultural datasets, where labeling plants in field images is time-consuming. Furthermore, these models often demand significant computational resources, posing challenges for drone implementation with constrained processing capabilities and limited power supply. The need for real-time performance, coupled with these limitations, underscores the importance of developing *lightweight* solutions for weed mapping applications.

In agricultural landscapes, crops are systematically planted in linear crop rows. Automatically detecting these crop rows can benefit various applications [32, 35]. Among them, it can be used to distinguish crops from weeds, as unwanted vegetation typically proliferates in the spaces between these rows. This spatial pattern has led to a significant body of research focused on developing and refining *crop-row* detection techniques [5, 6, 14, 16, 33]. This mechanism allows to detect inter-rows weeds. However, it may fail to detect intra-row weeds, as they are classified as crops. The generated detection can be used to create a pseudo-ground truth, which can be used to train a deep learning model. This approach has been explored in previous works [3, 4], but these methods primarily functioned as image classifiers rather than end-to-end segmentation tools.

This paper proposes a novel lightweight, fully automatic method for weed mapping that combines crop-row detection with a *noise-resilient* deep learning model. Our approach, RoWeeder (Fig. 1), leverages the crop-row information to create a pseudo-ground truth, which is then used to train a deep learning model. The model is based on the SegFormer encoder architecture [36], coupled with an ad-hoc decoder that fuses the features extracted by the encoder. Evaluation is conducted on the WeedMap dataset [29], which contains multispectral images of sugar beet fields captured by drones.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 details our proposed approach. Section 4 evaluates our model's performance. Section 5 concludes the paper with a summary of our findings and directions for future research.

2   Related Work

Advances in computer vision and remote sensing have revolutionized precision agriculture, addressing tasks such as disease and pest identification, abiotic stress assessment, growth monitoring, crop yield prediction, and weed mapping.

Weed mapping, a semantic segmentation task, assigns each pixel in an image to one of two classes, either *weed* or *crop*. Deep learning algorithms have demonstrated superior performance over traditional techniques in this area. Early work by Dos Santos et al. [10] highlighted the effectiveness of Convolutional Neural Networks (CNNs), such as AlexNet, over SVMs and Random Forests. Lottes et al. [22] further advanced the field by using a CNN with dual decoders for stem detection and plant segmentation, showing promising results on the BoniRob and UAV datasets.

Integrating multispectral images, which capture detailed information on plant health and species, enhances the accuracy of deep learning models compared to RGB-only models. For example, U-Net has successfully separated weeds from crops and soil [9]. WeedNet, based on SegNet and trained on the WeedMap dataset, is another example of successful application in this domain [28]. The WeedMap dataset, containing multispectral images from sugar beet fields in Germany and Switzerland, has become a benchmark for weed mapping studies [29]. Recently, we have explored the benefits of RGB pre-training and fine-tuning on multispectral images, as well as the application of knowledge distillation to improve the performance of lightweight models [7, 8].

Partially supervised methods, such as semi supervised and unsupervised learning, have also been explored for weed detection. Studies employing semi supervised approaches include [17, 23, 26, 31], while clustering methods for unsupervised weed detection are discussed in [2, 12, 25]. Crop-row detection methods, such as those using the Hough Transform, have been employed to distinguish between crop plants and weeds [24, 27, 30]. Bah et al. [3, 4] introduced deep learning techniques incorporating crop-row information, though these methods primarily functioned as image classifiers rather than end-to-end segmentation tools.

Building on these unsupervised methods, our work proposes a novel approach that combines crop-row detection with a noise-resilient end-to-end deep learning model. This lightweight model can detect weeds in real-time during inference, integrating crop-row information learned during training. To our knowledge, this is the first work that merges these two approaches.

3   RoWeeder

This research aims to develop a framework that can transfer knowledge derived from crop-row detection to a deep learning model for weed mapping. Our approach, RoWeeder, uses crop-row information to create a pseudo-ground truth, which is then used to train the deep learning model. The proposed framework consists of three main components: a plant detection module, a crop-row detection module, and a deep learning model for final segmentation. Figure 1 illustrates the workflow of the RoWeeder framework.

The plant detection module uses classical thresholding to detect plants in the image. When Near Infrared (NIR) images are available, the Normalized Difference Vegetation Index (NDVI) is calculated to improve plant detection.
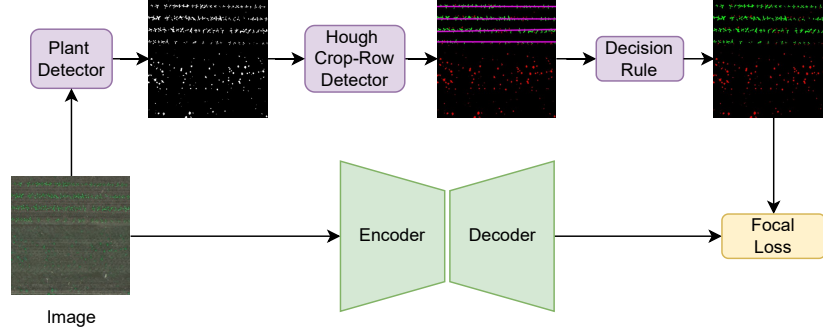
Fig. 1: Overview of the proposed framework, illustrating the training process of the semantic segmentation model. In the top branch, the image is fed into a plant detection algorithm, and the mask produced is used to detect crop rows. Every plant on a crop row is classified as a crop; otherwise, it is classified as a weed. This pseudo-ground truth is used to train a deep learning model.

The crop-row detection module employs the Hough Transform [11] to detect crop rows in the image. The module takes the segmentation mask of the plants as input and outputs the crop-row mask. The crop-row mask is then used to create the pseudo-ground truth for the deep learning model. Since all the images are rotated by the same angle, they are first rotated to align the crop rows with the horizontal axis before crop-row detection. This consistent rotation ensures accurate detection by the Hough Transform. Moreover, the data may contain images filled with weeds (found at the edges of the fields), leading to the detection of false positive lines by the Hough Transform. In these images, the detected lines' angles (θ) are evenly distributed, while in images with crop rows, the θ angles are concentrated around a specific value. To tackle this issue, we analyze the distribution of θ angles to filter out false positive lines. We use the Kolmogorov-Smirnov test to check the uniformity of the θ angle distribution, and if the distribution is found to be uniform, we discard all the detected lines.

Crop and weed classification is performed using a simple rule-based method: if at least one pixel of a plant overlaps with a crop row, it is classified as a crop; otherwise, it is classified as a weed. Each plant instance can be detected using either the Simple Linear Iterative Clustering (SLIC) algorithm [1] or by calculating each connected component (CC) of the image. This method effectively detects inter-row weeds but may fail to detect intra-row weeds, as they are classified as crops. The hypothesis is that the deep learning model will learn to distinguish between crops and weeds despite the pseudo-ground truth noise. Examples of segmentations produced by the Hough Crop-Row Detector with the decision rule are shown in Fig. 2.

RoWeeder is built on the SegFormer architecture [36], a Transformer-based model that has shown state-of-the-art performance in semantic segmentation
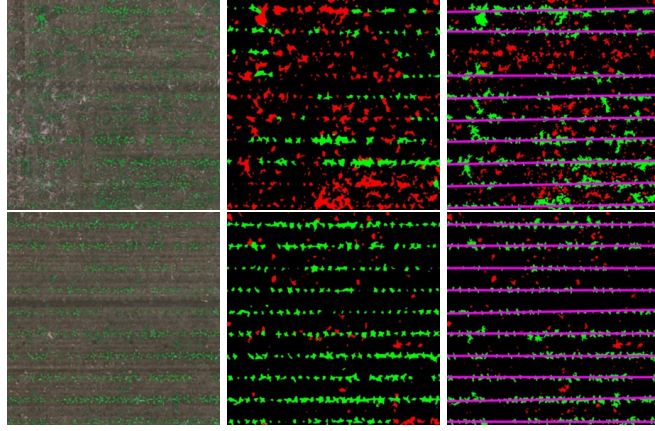
Fig. 2: Visual representation of the crop lines detected by the Hough Transform. From left to right: the original input image, the ground truth with crops in green and weeds in red, and the generated pseudo-ground truth with detected crop rows in purple.

tasks. SegFormer comes in six configurations, ranging from small to large models. We chose the smallest configuration, SegFormer-B0, which has only 3.7M parameters, to ensure that the model is light enough to be deployed on edge devices like NVIDIA Jetson TX2 and Jetson Nano [18]. The model is trained on the pseudo-ground truth created by the crop-row detection module.

SegFormer employs an all-MLP lightweight decoder. We designed different segmentation decoders to identify the most suitable for our task. The first is a pyramid-based decoder. Starting from the deepest feature map, at each stage, this map is upsampled to the next one, projected to the same dimension as the feature map with a point-wise convolution, and then fused. Fusion can be done by concatenation or addition, while upsampling can be done by bilinear interpolation or transposed convolution. Finally, the fused feature map is processed by a 3×3 convolution to handle spatial information, followed by a GELU activation function [15]. Given F_{deep} and $F_{shallow}$, the deepest and shallowest feature maps, a pyramid-based decoder block with addition as fusion can be defined as:

$$F_{out} = GELU\left(\text{Conv}_{3\times3}\left(\text{Conv}_{1\times1}\left(\text{Upsample}\left(F_{deep}\right)\right) + F_{shallow}\right)\right) \quad (1)$$
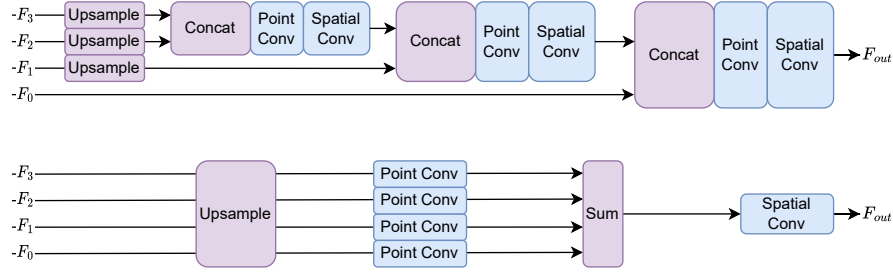
where Upsample can be bilinear interpolation or transposed convolution. A pyramid-based decoder block with concatenation as fusion can be defined as:

$$F_{out} = GELU\left(\text{Conv}_{3\times3}\left(\text{Conv}_{1\times1}\left(\text{Upsample}\left(F_{deep}\right) || F_{shallow}\right)\right)\right) \quad (2)$$

where $||$ denotes concatenation.

The second decoder is a "flat" decoder that projects all the feature maps in one step. The flat decoder with sum as fusion can be defined as:

$$F_{out} = GELU\left(\text{Conv}_{3\times3}\left(\sum_{n=1}^{N}(\text{Conv}_{1\times1}(\text{Upsample}(a_n)))\right)\right) \quad (3)$$

Fig. 3: Overview of the two decoders. On top is the pyramid decoder with concatenation as a fusion method. On the bottom is the flat decoder with sum as the fusion method. The input feature maps are ordered from the most shallow to the deeper ones. "Spatial Conv" is a 3×3 convolution, while "Point Conv" is a 1×1 convolution.

where a_n is the n-th feature map, N is the number of blocks, Upsample can be either bilinear interpolation or transposed convolution, and Fusion can be either concatenation or addition. The flat decoder with concatenation as fusion can be defined as:

$$F_{out} = GELU\left(\text{Conv}_{3\times3}(\text{Conv}_{1\times1}(\text{Upsample}(a_1)||\ldots||\text{Upsample}(a_N))))\right) \quad (4)$$

where $||$ denotes concatenation.

Each decoder is followed by a 1×1 convolution to reduce the number of channels to the number of classes and a softmax activation function to output the probability of each class. The two decoders are shown in Fig. 3.

4   Experimental Evaluation

4.1   Dataset and Validation

We assessed RoWeeder on the WeedMap dataset [29], which includes multispectral images from sugar beet fields in Germany (Rheinbach) and Switzerland (Eschikon). These images were captured using UAVs equipped with RedEdge-M and Sequoia cameras, respectively (see [29] for further details). The dataset comprises eight orthomosaic maps, labeled [000] through [007], with the first five belonging to the Rheinbach subset and the last three to the Eschikon subset. Multiple tiles of size 512×512 pixels were derived from each orthomosaic map by sliding a non-overlapping window over the maps.

We focused on the Rheinbach subset because the Eschikon subset lacks the blue channel necessary for our RGB-focused experiments. Experimentation with other channels is beyond the scope of this work. The dataset was split to perform a 5-fold cross-validation. Additionally, the training set was randomly divided into training and validation subsets, and the validation subset was used for model selection. We randomly divided the training set rather than basing it on specific fields to avoid relying solely on a single field for validation. The dataset contains

annotations for crops and weeds used only to evaluate the model's performance at test time.

We used the macro-averaged F1 score as a metric for our experiments, which can deal with the class imbalance typical of this task.

4.2   Setting

The NDVI-based plant detection system utilized a threshold value of 0.1 to detect significant pixels of plants. For the Hough transform, a threshold of 160 was used. The uniformity test employed a p-value of 0.1. In the SLIC algorithm, the number of clusters was calculated as $n = 0.005 \times H \times W$, where H and W are the image height and width. The compactness parameter was set to 20, and σ was set to 1. The mentioned parameters were selected through an empirical evaluation of the detected lines within a representative sample of the dataset.

Our training strategy optimized the focal loss [19], which addresses class imbalance by focusing on difficult-to-classify examples. The loss for a single training instance was computed as:

$$\mathcal{L} = \frac{1}{N+1} \sum_{n}^{N+1} \left[w_n \cdot (1 - e^{-l_{ce}(\hat{y}_n, y_n)})^\gamma \cdot l_{ce}(\hat{y}_n, y_n) \right], \tag{5}$$

where $N+1$ is the number of classes, w_n are class-specific weights, l_{ce} is the cross-entropy loss, and γ is the focusing parameter that adjusts the emphasis on hard examples. We used the AdamW optimizer [21] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, complemented by a linear learning rate warmup [13] for 1000 iterations, followed by a step-wise cosine learning rate decay schedule [20]. The initial learning rate after warmup was 1e−5. The model was trained for 20 epochs, with the iteration count per epoch equivalent to the size of the training set.

Training harnessed the resources of the Leonardo cluster, which provided 512GB of RAM and four NVIDIA A100-64GB GPUs. This setup ensured the high-performance processing capabilities necessary to support our experiments' computational demands.

4.3   Results

We conducted a 5-fold cross-validation on the Rheinbach subset, with each fold corresponding to a different orthomosaic map. The pseudo-ground truth was used for the training folds, and the original ground truth was used for the testing folds. The results in Table 1 show that our method achieved a mean F1 score of 75.3, outperforming the baseline methods Hough+CC and Hough+SLIC. The results demonstrate that the model can learn to distinguish between crops and weeds even with noise in the pseudo-ground truth. This is particularly evident in fold 004, where the pseudo-ground truth generated by the Hough+CC method yielded an F1 score of 63.0, while our method achieved an F1 score of 74.3, and the Hough+CC+ResNet50 method adapted from [3] achieved an F1 score of

Method	Field					Mean
	000	001	002	003	004	
Hough+SLIC+ResNet50 [3]	73.8	**79.3**	72.7	**77.5**	52.9	71.2 ± 4.2
Hough+CC	68.1	72.0	**76.1**	72.7	63.0	70.4 ± 2.0
Hough+SLIC	68.1	72.0	**76.1**	72.5	63.1	70.4 ± 2.0
RoWeeder (SegFormer)	70.3	76.4	75.5	77.1	72.8	74.4 ± 1.1
RoWeeder (Pyramid)	71.6	74.2	74.7	76.8	72.0	73.9 ± 0.8
RoWeeder (Flat)	**74.5**	74.9	75.6	77.2	**74.3**	**75.3 ± 0.5**

Table 1: F1 scores from 5-fold cross-validation on the Rheinbach subset. RoWeeder, in the three different settings, outperforms the baselines and the competitors.

Method	Full			Inter-row			Intra-row		
	BG	Crop	Weed	BG	Crop	Weed	BG	Crop	Weed
Hough+SLIC+ResNet50 [3]	**98.5**	62.6	52.6	**99.0**	25.4	54.6	**98.8**	64.5	35.1
Hough+CC	**98.5**	65.8	46.9	**99.0**	01.9	55.6	**98.8**	69.4	00.1
RoWeeder (SegFormer)	98.4	69.5	55.4	98.8	40.9	59.5	98.6	70.3	38.2
RoWeeder (Pyramid)	98.4	69.9	53.3	98.9	41.0	56.6	98.7	71.3	36.9
RoWeeder (Flat)	98.4	**70.9**	**56.6**	98.9	**42.4**	**59.8**	98.7	**71.9**	**40.3**

Table 2: Per-class F1 score averaged over 5-fold cross-validation. *Full* represents the F1 score calculated across all pixels in the image. *Inter-row* refers to the F1 score for pixels belonging to plants on the crop rows, while *inter-rows* refers to the F1 score for pixels belonging to plants outside the crop rows. *BG* stands for background.

52.9. Fold 004 contains some low-quality, blurred images that may have compromised the Hough algorithm's predictions and effectiveness. However, the end-to-end model successfully overcame this issue, maintaining its accuracy. Note that RoWeeder learns from the output of Hough+CC, so achieving a higher score than its ground truth indicates its resilience to noise. The results also indicate that the SLIC superpixels do not enhance the model's performance, as the Hough+SLIC method achieved the same F1 score as the Hough+CC method.

The noise in the pseudo-ground truth arises from intra-row weeds being misclassified as crops. To evaluate the model's ability to handle this issue, we calculated two F1-scores: one for the pixels of plants within the rows, termed the *intra-row F1-score*, and another for the pixels of plants outside the rows, termed the *inter-row F1-score*. The results, averaged over the five folds test set, are presented in Table 2. Given the definitions of these two metrics, the inter-row F1-score for crops is 0, as the model cannot detect crops outside the rows. Similarly, the intra-row F1-score for weeds is 0 since the model fails to identify weeds within the rows. These scores, therefore, reflect the model's ability to learn from noisy data, indicating its noise resiliency.

Table 3 presents the computational requirements of the analyzed methods. The Hough+SLIC+ResNet50 method uses a CNN to classify detected crop rows and requires significant computational resources, with an inference time of 1620ms. In contrast, our method achieves an inference time of 7ms, making it

Method	Params (M)	GMACs	Inference time (ms)
Hough+SLIC+ResNet50 [3]	23.51	21.58*	1620
Hough+CC	/	/	84
Hough+SLIC	/	/	377
RoWeeder (SegFormer)	3.71	7.84	**7.07**
RoWeeder (Pyramid)	3.90	3.91	7.79
RoWeeder (Flat)	**3.60**	**3.68**	7.16

Table 3: Computational cost comparison between RoWeeder and competitors. * indicates that GMACs are calculated for the deep learning model only, and the method needs to rerun the deep learning model for each plant in the image, which is variable.

suitable for real-time applications through drones. Inference time was calculated on a single NVIDIA RTX 4090, a consumer-grade GPU, which is more powerful than a typical edge device but can still reflect the difference in computational cost between the methods. The RoWeeder model is lightweight, with only 3.6M parameters and 3.68 GMACs, making it suitable for edge-device deployment. Future work will focus on testing the model on edge devices to evaluate its performance in real-world scenarios.

Figure 4 presents qualitative results from the RoWeeder (flat) model. The first two rows highlight the primary limitation of the Hough baselines. In the first example, a line of weeds is misidentified as a crop row, while in the second, some crop rows go undetected. These errors stem from the Hough transform's sensitivity to parameters, such as the threshold value. Applying uniform parameters across an entire dataset can lead to such inaccuracies. Despite these ground truth errors, RoWeeder demonstrates robust performance, successfully identifying most weeds in the first example and most crops in the second.

4.4   Ablation Study

Our comprehensive ablation studies examined different components and configurations of the RoWeeder decoders. These experiments were performed using field 003 as the test set. Table 4 shows the results for different fusion and up-sampling methods across the two decoder settings. The results indicate that the flat decoder with sum as the fusion method achieved the best performance, with an F1 score of 77.2. The pyramid decoder with concatenation as the fusion method was the second best, with an F1 score of 76.8. Thus, for each decoder, a different fusion method proved more suitable. Additionally, upsampling through interpolation yielded better results than deconvolution.

Another component evaluated was the decoder's 3×3 spatial convolution. An ablation study over field 003, removing the spatial convolution from the flat and pyramid decoders, showed that spatial convolution is crucial for good performance. The F1 score dropped from 77.2 to 75.4 for the flat decoder and from 76.8 to 74.0 for the pyramid decoder when the spatial convolution was removed. Lastly, we evaluated the impact of the number of blocks in the model. Table 5 shows that removing blocks from the encoder and decoder decreased the

Input image	Ground truth	Pseudo-ground truth	Prediction
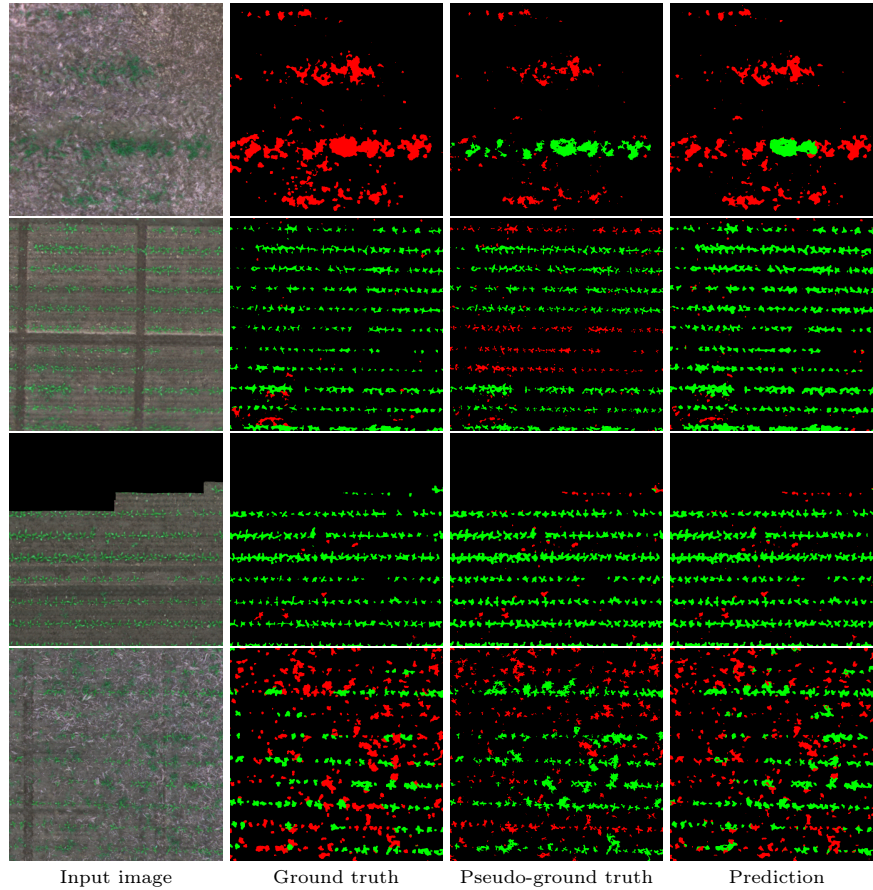
Fig. 4: Pseudo-ground truth is generated with the Hough+CC method and prediction by RoWeeder. The background is denoted in black, crops in green, and weeds in red.

model's performance for both the flat and pyramid decoders. However, GMACs linearly decrease with the number of blocks, while the F1 score drop is less pronounced. Therefore, this trade-off could be considered when the environment has limited computational resources.

We also trained our model in a supervised setting to establish a benchmark for the unsupervised approach. Using 5-fold cross-validation, we achieved an F1 score of 82.8. Although the unsupervised method results in a 7.5-point decrease in F1 score, it still delivers competitive performance.

5   Conclusion

In this study, we presented RoWeeder, an innovative framework for unsupervised weed mapping that combines crop-row detection with a noise-resilient deep

Upsample	Fusion	F1 score	
		Pyramid	Flat
Interpolation	Add	73.7	**77.2**
	Concat	**76.8**	75.3
Deconvolution	Add	73.1	73.1
	Concat	74.7	73.5

Table 4: Results of different settings of the decoders over field 003.

Method	# blocks	F1 Score	GMACs	Params
Pyramid	2	75.3	**2.00**	**0.69**
	3	76.3	3.10	1.67
	4	**76.8**	3.91	3.90
Flat	2	72.3	**2.00**	**0.69**
	3	74.8	2.97	1.62
	4	**77.2**	3.68	3.60

Table 5: Ablation study over the number of blocks in the encoder and decoder.

learning model. Our approach leverages crop-row information to create a pseudo-ground truth, which is then used to train a deep learning model. We evaluated our method on the WeedMap dataset, achieving an F1 score of 75.3, outperforming other methods. Our results demonstrate that the model can effectively distinguish between crops and weeds, even with noise in the pseudo-ground truth. Additionally, we conducted an ablation study on different components and configurations of the RoWeeder decoders, showing that the flat decoder with sum as the fusion method achieved the best performance.

We envision a broad spectrum of future research directions to advance our method. These include exploring different architectures for the deep learning model, investigating the impact of various crop-row detection methods, and extending the framework to other crops and datasets. We also plan to develop a new decoder more resilient to noise by leveraging contrastive learning to build a class prototype for each class and then use this prototype to classify each pixel.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (eds.): SLIC Superpixels. EPFL (2010)

2. Agarwal, R., Hariharan, S., Nagabhushana Rao, M., Agarwal, A.: Weed Identification using K-Means Clustering with Color Spaces Features in Multi-Spectral Images Taken by UAV. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. pp. 7047–7050 (Jul 2021), iSSN: 2153-7003

3. Bah, M.D., Hafiane, A., Canals, R.: Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. Remote Sensing **10**(11), 1690 (Nov 2018), number: 11 Publisher: Multidisciplinary Digital Publishing Institute

4. Bah, M.D., Hafiane, A., Canals, R., Emile, B.: Deep features and One-class classification with unsupervised data for weed detection in UAV images. In: 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–5 (Nov 2019), iSSN: 2154-512X

5. Bah, M.D., Hafiane, A., Canals, R.: CRowNet: Deep Network for Crop Row Detection in UAV Images. IEEE Access **8**, 5189–5200 (2020), conference Name: IEEE Access

6. Bah, M.D., Hafiane, A., Canals, R.: Hierarchical graph representation for unsupervised crop row detection in images. Expert Systems with Applications **216**, 119478 (Apr 2023)

7. Castellano, G., De Marinis, P., Vessio, G.: Applying Knowledge Distillation to Improve Weed Mapping with Drones. In: 2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS). pp. 393–400 (2023)

8. Castellano, G., De Marinis, P., Vessio, G.: Weed mapping in multispectral drone imagery using lightweight vision transformers. Neurocomputing **562**, 126914 (2023)

9. Chicchón Apaza, M.Á., Monzón, H.M.B., Alcarria, R.: Semantic Segmentation of Weeds and Crops in Multispectral Images by Using a Convolutional Neural Networks Based on U-Net. In: International Conference on Applied Technologies. pp. 473–485. Springer (2019)

10. dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H., Folhes, M.T.: Weed Detection in Soybean Crops Using ConvNets. Computers and Electronics in Agriculture **143**, 314–324 (2017)

11. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM **15**(1), 11–15 (Jan 1972)

12. Gašparović, M., Zrinjski, M., Barković, D., Radočaj, D.: An automatic method for weed mapping in oat fields based on UAV imagery. Computers and Electronics in Agriculture **173**, 105385 (Jun 2020)

13. Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., He, K.: Accurate, large minibatch sgd: Training imagenet in 1 hour. arXiv preprint arXiv:1706.02677 (2017)

14. Guerrero, J.M., Guijarro, M., Montalvo, M., Romeo, J., Emmi, L., Ribeiro, A., Pajares, G.: Automatic expert system based on images for accuracy crop row detection in maize fields. Expert Systems with Applications **40**(2), 656–664 (Feb 2013)

15. Hendrycks, D., Gimpel, K.: Gaussian Error Linear Units (GELUs) (Jun 2023), arXiv:1606.08415 [cs]

16. Ji, R., Qi, L.: Crop-row detection algorithm based on Random Hough Transformation. Mathematical and Computer Modelling **54**(3), 1016–1020 (Aug 2011)

17. Khan, S., Tufail, M., Khan, M.T., Khan, Z.A., Iqbal, J., Alam, M.: A novel semi-supervised framework for UAV based crop/weed classification. PLOS ONE **16**(5), e0251008 (May 2021), publisher: Public Library of Science

18. Lee, D.J., Lee, J.Y., Shon, H., Yi, E., Park, Y.H., Cho, S.S., Kim, J.: Lightweight Monocular Depth Estimation via Token-Sharing Transformer. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 4895–4901 (May 2023). `https://doi.org/10.1109/ICRA48891.2023.10160566`, `https://ieeexplore.ieee.org/abstract/document/10160566?casa_token=HGmhtCEuYT8AAAAA:ugWvbKdGAz_3b7xgRVVOQbzdCICdVwEvLJOAdVMORlAnweXlsoO9fwk-JyOPXtfkbfv6OXQeLkE`

19. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)

20. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)

21. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)

22. Lottes, P., Behley, J., Chebrolu, N., Milioto, A., Stachniss, C.: Joint Stem Detection and Crop-Weed Classification for Plant-Specific Treatment in Precision Farming. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 8233–8238. IEEE (2018)

23. Nong, C., Fan, X., Wang, J.: Semi-supervised Learning for Weed and Crop Segmentation Using UAV Imagery. Frontiers in Plant Science **13** (Jul 2022), publisher: Frontiers

24. Peña, J.M., Torres-Sánchez, J., Castro, A.I.d., Kelly, M., López-Granados, F.: Weed Mapping in Early-Season Maize Fields Using Object-Based Analysis of Unmanned Aerial Vehicle (UAV) Images. PLOS ONE **8**(10), e77151 (Oct 2013), publisher: Public Library of Science

25. Peerbhay, K., Mutanga, O., Lottering, R., Agjee, N., Ismail, R.: Improving the unsupervised mapping of riparian bugweed in commercial forest plantations using hyperspectral data and LiDAR. Geocarto International (Feb 2021), publisher: Taylor & Francis

26. Pérez-Ortiz, M., Peña, J., Gutiérrez, P.A., Torres-Sánchez, J., Hervás-Martínez, C., López-Granados, F.: A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. Applied Soft Computing **37**, 533–544 (2015), publisher: Elsevier

27. Pérez-Ortiz, M., Peña, J.M., Gutiérrez, P.A., Torres-Sánchez, J., Hervás-Martínez, C., López-Granados, F.: Selecting patterns and features for between- and within-crop-row weed mapping using UAV-imagery. Expert Systems with Applications **47**, 85–94 (Apr 2016)

28. Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R.: Weednet: Dense Semantic Weed Classification Using Multispectral Images and Mav for Smart Farming. IEEE robotics and automation letters **3**(1), 588–595 (2017)

29. Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R.: WeedMap: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming. Remote Sensing **10**(9), 1423 (2018)

30. dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H., Folhes, M.T.: Unsupervised deep learning and semi-automatic data labeling in weed discrimination. Computers and Electronics in Agriculture **165**, 104963 (Oct 2019)

31. Shorewala, S., Ashfaque, A., Sidharth, R., Verma, U.: Weed Density and Distribution Estimation for Precision Agriculture Using Semi-Supervised Learning. IEEE Access **9**, 27971–27986 (2021), conference Name: IEEE Access

32. Stefanović, D., Antić, A., Otlokan, M., Ivošević, B., Marko, O., Crnojević, V., Panić, M.: Blueberry Row Detection Based on UAV Images for Inferring the Allowed UGV Path in the Field. In: Tardioli, D., Matellán, V., Heredia, G., Silva, M.F., Marques, L. (eds.) ROBOT2022: Fifth Iberian Robotics Conference. pp. 401–411. Springer International Publishing, Cham (2023). `https://doi.org/10.1007/978-3-031-21062-4_33`
33. Vidović, I., Cupec, R., Hocenski, v.: Crop row detection by global energy minimization. Pattern Recognition **55**, 68–86 (Jul 2016)
34. Vougioukas, S.G.: Agricultural Robotics. Annual Review of Control, Robotics, and Autonomous Systems **2**, 365–392 (2019)
35. Waqar, R., Grbovic, Z., Khan, M., Pajevic, N., Stefanovic, D., Filipovic, V., Panic, M., Djuric, N.: End-to-End Deep Learning Models for Gap Identification in Maize Fields
36. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. Advances in Neural Information Processing Systems **34** (2021)