Computers and Electronics in Agriculture 245 (2026) 111524

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Balancing accuracy and cost in precision agriculture: a few-shot learning approach for efficient weed – crop segmentation☆

Nico Catalano [a,*], Sofia Matilde Luglio [b,*], Agnese Chiatti [a], Mino Sportelli [b,d], Christian Frasconi [b], Davide Facchinetti [c], Matteo Matteucci [a]

[a] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Via Giuseppe Ponzio, 25/P1, 20133 Milano, Italy
[b] Department of Agriculture, Food and Environment, University of Pisa, Via del Borghetto 80, 56124 Pisa, Italy
[c] Department of Agriculture and Environmental Sciences, University of Milan, Via Celoria 2, 20133 Milano, Italy
[d] National Research Council (CNR), Institute of Information Science and Technologies (ISTI), Via G. Moruzzi 1, 56124 Pisa, Italy

ARTICLE INFO

Keywords:
Precision Agriculture
Artificial Intelligence
Computer Vision
Few Shot Segmentation
Weed Control

ABSTRACT

Autonomous weeding, a task requiring expertise at the intersection of Computer Vision and Agronomy depends on accurate segmentation of crops and weeds from robot-collected images. Traditional segmentation models (i.e. YOLO) require large, densely annotated datasets, whose creation is costly and labor-intensive. In contrast, Few-Shot Learning (FSL) methods can learn from minimal annotated examples and significantly reduce the costs of dataset creation.

This study evaluates the ability of a FSL architecture, HDMNet, to perform crop and weed segmentation using only a single annotated support image. Its performance retains 73–80% of the accuracy compared with widely used, annotation-intensive detectors designed for large datasets such as YOLOv5 and YOLOv8 when detecting bean and corn plants.

Because reliable estimates of annotation effort are lacking in agriculture, we provide a quantitative assessment of the labor required to produce pixel-level labels. Preparing the 2,069-images 'Early' dataset required **approximately 181 h,** while 102-images 'Refined' dataset still required **approximately 186 h.** Labeling accounted for **approximately 25** and **30 h**, respectively. These findings show that increasing annotation granularity sharply raises effort without proportional accuracy gains, making **dataset scale** more beneficial than mask detail for YOLO-based models. In contrast, few-shot methods achieve competitive performance while **eliminating most annotation labor**.

The study is further supported by the release of a new dataset from the 2023 ACRE field competition, including the 'Early' and 'Refined' versions.

Overall, the findings offer practical guidance for designing efficient datasets for agricultural image analysis and demonstrate that FSL can substantially reduce autonomous weeding systems deployment costs.

1. Introduction

As the global population is expected to grow to 9–10 billion of people by 2050, improved solutions for sustainable agriculture at scale are urgently needed to maximize crop production through large and uniform cultivation areas and consistent field treatment strategies (Karunathilake et al., 2023a). However, traditional large-scale agricultural methods, while boosting food production, often lead to inefficient resource utilization without considering nature conservation, environmental sustainability and ecosystem services (Tittonell et al., 2020).

Most recently, the widespread adoption of smart technologies in Artificial Intelligence (AI) and Robotics has transformed the agricultural sector, aiding the transition towards green solutions and supporting the objectives of the Green Deal (Finger et al., 2019). Broadly speaking, all applications that can provide farmers with precise information for a more efficient resource utilization and mitigate the environmental

☆ This article is part of a special issue entitled: 'Intelligent Machinery and robot' published in Computers and Electronics in Agriculture.

* Corresponding authors.
E-mail addresses: nico.catalano@polimi.it (N. Catalano), sofiamatilde.luglio@phd.unipi.it (S.M. Luglio), agnese.chiatti@polimi.it (A. Chiatti), mino.sportelli@isti.cnr.it (M. Sportelli), christian.frasconi@unipi.it (C. Frasconi), davide.facchinetti@unimi.it (D. Facchinetti), matteo.matteucci@polimi.it (M. Matteucci).

https://doi.org/10.1016/j.compag.2026.111524
Received 29 December 2023; Received in revised form 4 December 2025; Accepted 1 February 2026
Available online 11 February 2026

impact of agricultural practices contribute to the growing field of *Precision Agriculture* (Trang et al., 2022). Among these, Sharma et al. (2022) have identified six main areas of application: crop yield and price prediction, intelligent spraying, agricultural robot activity monitoring, crop development predictive insights, crop and soil monitoring, crop disease diagnosis.

Indeed, the adoption of AI and Autonomous Robotics technologies for precision farming has the potential to significantly reduce energy and resource consumption in agriculture, as prescribed in the European Commission Farm to Fork strategy,[1] automating and making more efficient labor-intensive tasks such as plant spraying, weeding, and harvesting. Crucially, the deployment of sustainable farming practices through agricultural robots directly contributes to meeting the UN's Strategic Development Goals (SDGs), by i) increasing the proportion of agricultural area under productive and sustainable farming (SDG Indicator 2.4.1), while also ii) drastically reducing water waste (Indicator 6.4.1). By monitoring plants more precisely, robots will also reduce the use of Nitrogen fertilisers, whose global cost has doubled in 2021. On-farm studies have estimated yields rise by 1.75%, with a reduction of 7–13 $/acre in energy input cost and an 8% reduction of irrigation resulting from the introduction of robotic phenotyping techniques (Pearson et al., 2022).

While initial applications in Precision Agriculture hold promise, the gap between the available data and the real-world adoption of these technologies remains vast. According to Blasch et al. (2022) the low adoption of precision faming technologies in Europe is due to the farmers' lack of knowledge and financial means to invest in cutting-edge solutions. Moreover, Blasch et al. (2022) also hypothesize that the small field size, unfit for most roboticized equipment currently available, may be another contributing factor.

Despite the obstacles and uncertainties surrounding the adoption of precision farming techniques, the efficacy of these methods is supported by scientific evidence. Karunathilake et al. (2023b) have shown how these punctual practices can contribute to a more efficient use of inputs while indirectly reducing the energy consumption footprint and costs. According to Nurcahyo et al. (2023) intelligent support systems which integrate different sensors and AI techniques can support data analysis and site-specific operations in weed management scenarios. Furthermore, precise crop and weed identification through autonomous solutions can reduce the chemical input and reduce crop injury, as discussed in (Tataridas et al., 2022, Gallo et al., 2023). A compelling case for the automatization of Precision Agriculture tasks is exemplified by the application of autonomous weeding robots depicted in the flow-chart in Fig. 1. These robots can effectively reduce chemical usage by delivering pesticides only where needed or by employing targeted mechanical destruction. To be successfully deployed in autonomous weeding scenarios, robots ought to be equipped with advanced perceptual capabilities, to autonomously comprehend the information collected through their sensors (Fig. 2).

One crucial pre-requisite to autonomous weeding in real-world environments is the segmentation of weed and crop regions from robot-collected images. Semantic segmentation, a task involving the prediction of category labels at the pixel level, as presented by (Milioto et al., 2018) becomes instrumental in achieving this goal. However, the successful application of traditional semantic segmentation models like Mask R-CNN (He et al., 2017), PSPNet (Zhao et al., 2016), and YOLO (Redmon, et al., 2016), face several limitations when applied in agricultural settings.

One of the primary challenges is the reliance of these models on large, well-annotated datasets, which are often expensive and time-consuming to produce. Field images in agriculture are complex, with high weed occlusion, variable environmental conditions and diverse crop types, making the process of manual annotation both labor-intensive and prone to inaccuracies (Ragu and Teo, 2023).

Annotating such images at the pixel level, particularly for dense vegetation and overlapping plants, requires significant expertise and effort (Krizhevsky et al., 2017). As a result, many farmers opt to bypass automated weed control in favor of expert-driven manual methods, underscoring the inefficiency of current deep learning approaches for practical use in agriculture.

Moreover, once trained, CNN-based models can recognize a fixed number of classes, thus limiting their flexibility. To expand the number of classes adding weed species or crops, the models must undergo fine-tuning with additional thousands of images and annotation work required (Chen et al.,2019), further exacerbating the data dependency issue. Transfer learning have been introduced to mitigate some of these limitations by adjusting pre-trained models to new data. This reduces the need for large annotated datasets and in some cases achieved high accuracy (Abdalla et al., 2019). While transfer learning provides benefits in terms of data efficiency performance, Gidaris and Komodakis (2018) highlighted that this method still requires adequate labeled data to prevent overfitting. This data-hungry paradigm creates an entry barrier to the development of improved autonomous weeding solutions for Precision Agriculture.

Unlike traditional segmentation methods, Few Shot Segmentation (FSS) models can learn a new class from a minimal set of examples, ranging from a single instance to a handful (e.g., one to five labeled examples). As such, these methods can drastically reduce the costs and resources required for bootstrapping autonomous segmentation solutions. Moreover, relying on models that can learn from limited data also simplifies the update of the model under changing environmental conditions. This feature is particularly important in real-world agricultural settings, which are characterized by marked weather, lighting and season variations. Hence, in principle, few-shot methods are prime candidates for the automatization of weeding tasks in Precision Agriculture. However, their performance is still to be fully compared to that of the widely adopted YOLO methods, which, albeit data hungry, exhibit impressive inference accuracy and speed. In addition, while FSS models have exhibited promising performance on widely adopted general datasets such as PASCAL VOC and MS-COCO, their efficacy in agricultural applications remains to be evaluated.

In this paper, we provide a comparative analysis between YOLO-based segmentation methods and HDMNet (Peng et al., 2023), a novel few-shot learning model designed for segmentation tasks. Our investigation centers on the practical implications of these models in the context of crop and weed segmentation for corn and beans cultivation. Our evaluation focuses on both the performance metrics of the segmentation models and human hours needed for dataset collection and labeling, analyzing the technical aspects and cost-effectiveness of YOLO and HDMNet in agricultural image analysis.

In summary, our paper makes the following contributions:

- Two comprehensive segmentation datasets for beans and corn cultivation: an expansive 'Early Dataset' (2069 images) and a meticulously annotated 'Refined Dataset' (102 images) with precise labels even for the smallest leaves.[2]
- A quantification of the human labor involved in the dataset creation phases, providing a valuable benchmark for estimating the efforts required to curate new agricultural datasets.
- A comprehensive set of experiments to assess the utility and practical applicability of few-shot segmentation models on real agricultural datasets of bean and corn cultivations, that introduce unique challenges compared to controlled benchmarks in Computer Vision (e.g., PASCAL-VOC and MS COCO).

[1] Available online at: https://food.ec.europa.eu/horizontal-topics/farm-fork-strategy_en.

[2] Available online at: https://data.mendeley.com/preview/yxy7drms8k?a=f62e8f04-ed09-4174-b751-6effa23cbb7f.

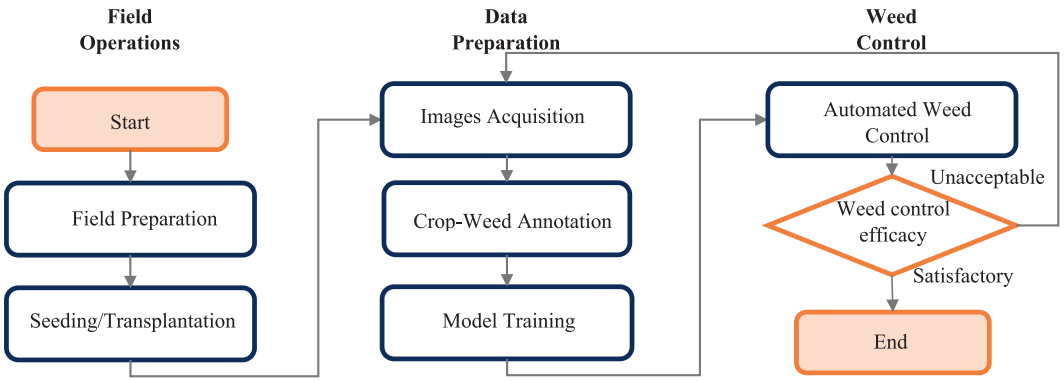


Fig. 1. Schematic representation of autonomous weed management workflow.

Fig. 2. Component of the main configuration employed during the trial: (A) router, (B) Smart switches, (C) on-board PC, (D) Intel® RealSense™ D435.

2. Literature review

Having access to diverse training datasets is essential to adapt existing weed and crop segmentation models to visual domain shifts, as highlighted by Weyler et al. (2023). However, constructing such labeled datasets is inherently resource-demanding and time-intensive. This challenge significantly hinders the application of deep learning-based models for weed and crop segmentation. To address this issue, several strategies have been proposed in the literature that are tailored to the needs of weed and crop segmentation and that aim to counteract data scarcity. These methods fall under three main groups: data augmentation, transfer learning and domain adaptation techniques.

Data augmentation, a technique commonly used in Computer Vision, involves generating synthetic data from the original images by applying various transformations. Zou et al. (2021) and Su et al. (2021) have shown that incorporating synthetic images during training enhances the segmentation model's robustness against domain shifts. Similarly, Nesteruk et al. (2021) and Guldenring et al. (2021) have leveraged a limited number of labeled images to create large synthetic datasets for training segmentation models.

Another approach, domain adaptation, focuses on transforming images from a source style to a target style. Bertoglio et al. (2023) and Magistri et al. (2023) have explored various domain adaptation methods for weed and crop segmentation by adapting new images to the style of a segmentation model's training set, thereby avoiding the need to retrain the segmentation model. Similarly, Chiatti et al. (2022) applied selective fine-tuning to individual network layers, a technique also known as surgical fine-tuning, to handle visual domain shifts concerning the plant varieties, camera viewpoints, and occlusions caused by foliage.

While these approaches have shown promise, they still require collecting and labeling new data or adjust the base models. On the other hand, the Few-Shot Learning (FSL) literature propose models capable of similar objectives but with drastically lower number of training examples. In this field emerge FSS aiming at learning a new semantic class with just handful of examples, making a compelling case for the smart agriculture sector. The inherent difference between traditional and FSS methods can be substantial and is summarized in Table 1.

Recent reviews, such as those by Yang et al. (2022), Sun et al. (2024) and Nie et al. (2024), have underscored the growing importance of FSL in addressing data scarcity within agricultural segmentation tasks. Significant progress has been made in fields like plant disease recognition, where studies by Liang (2021) used metric learning-based FSL to identify cotton leaf spots, and Argüeso et al. (2020) demonstrated the

Table 1
Schematic representation of autonomous weed management workflow.

Traditional Segmentation methods	Few-shot Segmentation methods
• Data hungry (hundreds to thousands of training examples per class) • Higher annotation effort • High accuracy on classes seen at training time • Lower accuracy and limited generalisation on Out Of Distribution (OOD) examples. • Brittleness: adapting to new classes requires re-training the model	• Learning from a small number of training examples (typically 1 to 5 per class) • Lower annotation effort • Accuracy often lower on seen classes than methods based on traditional supervision • Higher accuracy and better generalisation to OOD examples • Improved adaptability to new classes, provided general features can be learned that aid adaptation

effectiveness of FSL in recognizing 38 plant diseases in the PlantVillage (Hughes and Salathé, 2015) dataset Similarly, Zhong et al. (2020) applied a conditional adversarial autoencoder for the identification of citrus golden grape diseases, while Wang et al. (2021) proposed a multimodal representation learning approach to tackle vegetable disease identification using both image and text data.

In addition to plant disease recognition, FSL has also been applied to other agricultural tasks. Zhang et al. (2021) leveraged FSL to detect the position of crop seeds from UAV images, and Li et al., (2021b) introduced a Siamese Domain Transfer Network (SDTN) to detect corn residues. FSL has also proven useful in pest identification, with Li et al. (2020) utilizing FSL to classify cotton pests and Gui et al. (2021) combining FSL with hyperspectral imaging to detect soybean heartworm.

Liu et al. (2024) introduced HSDNet, a new semantic segmentation model designed for poultry farm monitoring. It addresses the limitations of traditional models using HSDNet models showing the potential of FSL in adapting to new settings with just one image, reducing annotation costs and deployment time and outperforming state-of-the-art methods.

Nuthalapati and Tunga (2021) used a FSL approach for automatically classifying pests, healthy plants and plant diseases using minimal labeled data. Authors performed extensive tests (42 experiments across multiple datasets) shows the model significantly outperforms existing methods. Wang et al. (2022) improved weed and crop detection method using a transfer learning model and algorithms and optimization strategies to improve performance on a small dataset. RGB and multispectral images of three weed and three crop types were processed through image augmentation and produced high accuracy for classification and plant weed estimation.

However, despite these advances, Yang et al. (2022) emphasize a notable gap in the application of FSL to weed detection. This gap represents a critical challenge, as effective weed segmentation is vital for precision agriculture. In our study, we aim to address this gap by evaluating the FSS model HDMNet (Peng et al., 2023) for weed and crop segmentation. By comparing its performance with that of the widely used YOLO model, we assess both the accuracy and the labeling effort required for training. Our work seeks to explore the potential of FSL techniques in weed segmentation, an area that remains underdeveloped in current research.

3. Materials and methods

The following section outlines the materials and methodologies employed in conducting the study. As our study focuses on estimating human labor in data curation for autonomous weeding and evaluating the FSS model against the traditional segmentation model, we first set up a testbed for data acquisition. This field preparation process involved creating two distinct cultivation patches, one for *Zea Mays* L. (corn) and another for *Phaseolus vulgaris* L (bean), each containing both crops and weeds. The Field preparation section provides a detailed account of this step. Then, the Data acquisition section covers the robotic platform used for collecting the new dataset and the briefing describing the already available ACRE 2019 dataset. The Preprocessing and labelling section describes how the new Early and Refined datasets have been labelled and curated. Finally, in the Background Section, we cover the principles of FSS and introduce a popular training scheme in the field: Episodic Learning. We conclude with an overview of the specific models we have selected for the experiments in this paper: HDMNet and YOLO (in its two variants: YOLOv5 and YOLOv8).

3.1. Field preparation

The data acquisition was conducted during the ACRE Field Campaign 2023 at the Department of Agricultural and Environmental Sciences of University of Milan “Ciro Menozzi – Baciocca” experimental farm (45° 30’ 09’’ N, 9° 01’ 01’’ E). The testbed area was part of a broader area covering 2.8 ha of medium mixture soil with a high number of stones. Further details on the experimental design of this field campaign are available in Luglio et al. (2023). The machinery used for field preparation is reported in Table 2, where the listed operation time, fuel consumption, and power demand are indicative values referred to one working element and to an average working speed following the reference value in Peruzzi and Sartori, 1997.

Images were collected in a defined area of 600 m2 and it was divided into 4 plots (3 × 50 m). As the entire field, the plots were first mowed to remove the weeds not involved in the competition, then the soil was ploughed at 25 cm of depth, then the main part of the stones was buried at 15 cm of depth by a stone burier. Harrowing at 20 cm of depth was also conducted. This smaller area included 18 rows of corn and 16 rows of bean. *To ensure a balanced acquisition of weed examples in terms of weed density* we partitioned this area in three different zones corresponding to low, medium and high frequency of weed presence. The corn was sown at 75 cm of row distance and 14 cm between each plant, the bean was manually transplanted at 37,5 cm between rows and 7 cm between each plant. *In terms of ensuring variety in the crop and weed types being investigated, th*ree different weed plant species were manually transplanted and placed at random throughout the field area: Sinapis arvensis L. (wild mustard), Matricaria chamomilla L. (chamomile) and *Lolium perenne* L. (ryegrass). Weeds were placed in the central part of the rows, leaving 10 m at the beginning and at the end of each row. Plants were transplanted to obtain a high weed incidence in the central 5 m and a low distribution in the first and final 10 m. Weed varieties in the acquired set presented different shapes and visual patterns, although they were of comparable dimension due to being captured at the same growth stage**.** To reduce the incidence of marginal and background noise in the collected data due to the presence of scattered plants lying outside of the target area of intervention, plants were transplanted to obtain a high weed incidence in the central 5 m and a low distribution in the first and final 10 m. It is worth noting that the robot acquired images only under natural lighting and that images were taken from two opposite directions. This step ensures to include images with and without the shadow of the robot. The field was prepared so as to ensure that different weed and crop varieties, presenting different shapes and visual patterns, are represented in the acquired set. However, weed examples in the acquired set were of comparable dimension due to being captured at the same growth stage. To ensure optimal plant health and homogenous soil moisture conditions at data acquisition time, in the last step of field preparation, the target area was irrigated with 2,5 L m-2 twice a day for two days.

3.2. Datasets acquisition

The robotic platform used for data collection was the commercial mobile rover Scout 2.0 (0.93 m × 0.699 m × 0.349 m) by AgileX Robotics, which was equipped with different components and sensors. The rover consists of a skid steering 4WD chassis and, during this trial, it was

Table 2
List of operations carried out during the ACRE 2023 field preparation.

Field operation	Tractor	Implement	Operation time h ha^{-1}	Hourly fuel consumption kg h^{-1}	Power demand kW
Ploughing	Deutz-Fahr 105 hp	Two-share mounted plough	4.3–5.9	4.0–7.0	10.0–16.0
Stone removal	Same Solaris 25 hp	Stone burier	0.8–1.7	2.0–4.0	4.0–7.0
Harrowing	Deutz-Fahr 105 hp	Spring-tooth harrow	1.7–3.3	5.0–13.0	12–28
Seeding	Same Solaris 25 hp	Seed drill	2.0–2.9	3.0–5.0	8.0–13.0

equipped with a shuttle PC (with Ubuntu 18.04 LTS), a Wi-Fi router, and an RGB camera. A stable Wi-Fi connection was established to ensure full coverage of the large area of work. Realtime data acquisition was enabled via Secure Shell connection to the robot's shuttle PC, running he Robot Operating System (ROS1). The Intel® RealSense™ RGBD camera was mounted at a 50 cm height from the ground. The rover was teleoperated through a remote control, maintaining an average speed of 1.44 km h^{-1} throughout data acquisition. Images were automatically collected as a video, and they were stored in the ROS bag standard format for robot logging, to be used for the following data preparation and analysis steps.

3.3. Preprocessing and labelling

To annotate the acquired image set, we relied on the Robotflow API.[3] This tool also allowed us to convert the annotations to the standard COCO and YOLO image labelling formats.

The **2023 ACRE Competition Dataset** provides for a total of 2069 1280x720px images, featuring corn and bean crops, alongside instances of three weed species. Crops instances were individually labelled as 'Bean' and 'Corn', while the remaining plants in the scene not belonging to any crop class were generally labelled as 'Weed'. Thus, all dataset versions contain a total of three labels: 'Bean', 'Corn' and 'Weed'. To annotate the ground truth masks associated with object regions of interest, we resorted to the Smart Polygon function available in the Roboflow API. This feature integrates an annotation assistant, based on the Segment Anything Model (SAM) proposed by Kirillov et al. (2023), which streamlines the labelling process. By clicking the center of the target object, the Smart Polygon function applies an initial label based on SAM operating in the background. The model suggests a shape for the object, which the annotators can visually inspect and annotate with their textual label. Annotators also have the option to refine and correct the predicted polygon. Thanks to this semi-autonomous annotation pipeline, we could prepare a larger **Early set** of crop and weed annotations for the ACRE 2023 dataset. Then, we also produced a second set, which we will refer to, in the following, as the **Refined** dataset. For this second dataset, we focused on a smaller set of images sampled from ACRE 2023, non-overlapping with the Early set. Annotations for this set were produced through a more time-consuming and careful inspection of crop and weed regions output by the Smart Polygon function. In addition to manually refining object regions annotated automatically, we also manually labelled any regions that were missed by the annotation assistant. This is the case of small-sized weed leaves that are particularly difficult to spot against uneven terrain.

In sum, the **Early** dataset, 1326 images represent corn examples, and 743 images depict bean crops. In terms of individual plant instances, the Early dataset contains a total of 3155 instances of corn, 2918 instances of beans, and 622 instances of weed.

Notably, the **Refined** dataset was derived from a subset of 102 images chosen at random from the larger **2023 ACRE Competition Dataset**, to ensure a representative selection for all the studied classes. This refined collection consists of a total of 102 images, with 51 image examples for each crop type. Specifically, this set provides 3971 instances of weed, 269 instances of beans, and 133 instances of corn. Thus, while we sampled the original 2023 ACRE set through a stratified procedure across crop types, we did not intervene on the natural distribution of weed and crop instances within each individual image. This is important to ensure that the resulting dataset reflects the real-world conditions encountered by the weeding robot and provides an adequately challenging benchmark to pave the way for the future deployment of the solution.

Fig. 3 shows an excerpt of images and annotations from the two datasets. As can be noted from comparing the left-hand and right-hand sides of Fig. 3, the Refined set provides denser annotations of the weed and crop regions than the Early set, thanks to the more time-consuming annotation process adopted to draw mask contours precisely by hand. As such, compared to the Early set, this dataset provides a different perspective. It poses the focus on the capability to segment out weed instances, including small and scattered weed patches that are particularly costly to identify and annotate, while maintaining a balanced representation of corn and beans.

The **2019 ACRE Competition Dataset** (Bertoglio et al., 2021) was collected for benchmarking plant discrimination solutions developed by teams participating in the 2020 edition of the ACRE competition. Data were acquired in October 2020 at the experimental farm of the National Research Institute for Agriculture, Food and the Environment (INRAE), in Montoldre (France). The testbed area, consisting of sandy soil in the absence of any stones, was partitioned into four experimental plots, each measuring 2 m in width and 46.5 m in length. The data collected during this field campaign covers corn and bean crops. Additionally, several different weed varieties were included in the testbed arena and are part of this image collection: *L. perenne* (ryegrass), *S. arvensis* (wild mustard), *Chenopodium album* L. (Lamb's quarter) and *M. Chamomilla* (chamomile).

Thus, on the one hand, the ACRE 2019 set closely resembles the test conditions of the ACRE 2023 dataset, covering the same crop varieties as ACRE 2023 and a shared set of weed varieties. On the other hand, the ACRE 2019 set was collected in a different geographical region, under different soil conditions, following a different field preparation protocol and with different robotics platforms. These differences motivate testing whether models trained on the readily available ACRE 2019 set can also generalize to the case of the 2023 ACRE set, including the more challenging scenario of the Refined set, where even the smallest weed leaves are densely annotated.

Thus, together, these two datasets support the comprehensive analysis of crop and weed segmentation scenarios across different factors contributing to data acquisition, namely the geographical area, soil and lighting conditions, plant density and robotic platform. We will further illustrate the experimental setup we followed for conducting this test in Section 3.5.

3.4. Background

The semantic segmentation task involves predicting category labels describing the content of input images at the pixel level. Compared to object detection, where images are typically annotated with bounding boxes or polygonal masks representing the regions of interest, semantic segmentation produces finer-grained object masks where each pixel is assigned a distinct label.

In recent years, the development of semantic segmentation models has predominantly relied on the design principles of Convolutional Neural Networks (CNNs), following the widespread adoption of this paradigm across the field of Computer Vision (CV). CNN-based models certainly exhibit an impressive performance in terms of accuracy of the produced predictions. However, these methods are data hungry as their successful application relies on the availability of training datasets that adequately represent the distinguishing features of the target categories to be segmented. Collecting such densely annotated datasets is a costly and labor-intensive process.

To minimize these costs, a novel avenue of research known as FSL has emerged within the broader field of Machine Learning (ML). FSL seeks to mitigate the demand for large-scale training data by enabling models to generalize from only a small set of examples. The adaptation of FSL to the specific domain of semantic segmentation has sparked the research line of FSS.

In the remainder of this Section, we provide a few definitions to contextualize the experiments and contributions presented in this paper.

[3] Available online at: https://app.roboflow.com.

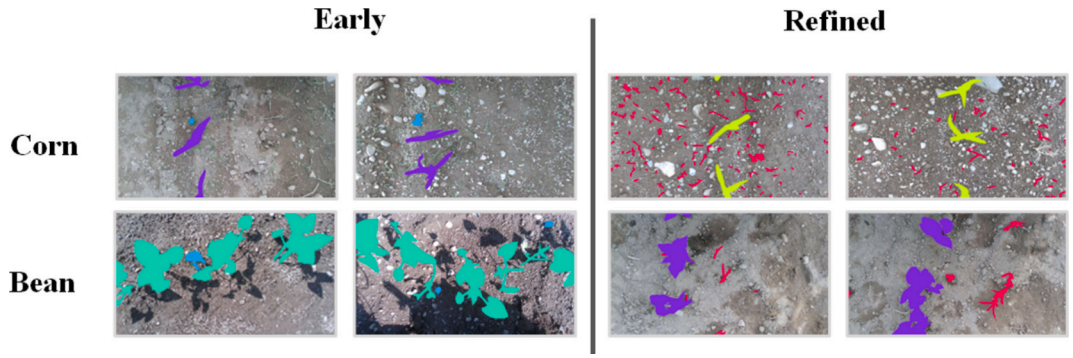


Fig. 3. Examples of the two datasets crops and weeds masks.

3.4.1. Few shot segmentation (FSS)

The FSS problem was originally defined by Shaban et al. (2017) as:

The FSS is the problem of predicting the semantic segmentation mask $\widehat{M}_q$ of a subject class l in a query image I_q given a support set S composed by k image mask couples.

More formally we define a query image I_q of shape $[H^q, W^q, 3]$, the support set of k image-label pairs, or k shots, as:

$$S(l) = \left\{ \left(I^i, M_l^i\right) \right\}_{i=1}^{k}.$$

Where I^i is an RGB image of shape $\left[H^i, W^i, 3\right]$ and M_l^i is a binary mask of shape $\left[H^i, W^i\right]$ segmenting the object of class l in I^i. The model is then the function f parameterised by θ that predicts the binary mask $\widehat{M}_q$ for the semantic class l in the query image I_q as

$$\widehat{M}_q = f_\theta\left(I_q, S(l)\right).$$

3.4.2. Episodic learning

In FSS, where labelled samples are scarce, conventional algorithms may fall short in training effectiveness. To overcome this limitation, a common approach in FSS studies (Rakelly et al., 2018, Zhang et al., 2019, Boudiaf et al., 2021, Tian et al., 2022, Wu et al., 2021, Nguyen and Todorovic, 2019, Li et al., 2021a, Lu et al., 2022, Siam and Oreshkin, 2019, Wang et al., 2019, Dong and Xing, 2018) is episodic training, a form of *meta*-learning, i.e., learning to learn.

Episodic learning involves training the model with simulated inference episodes. Originating from Vinyals et al. (2016) proposal, each inference episode in a Few Shot scheme comprises a support set, a query image, and its ground truth mask. During training, the model minimizes the loss between the predicted mask and the ground truth mask for each episode. The model performance is subsequently evaluated using *meta*-testing episodes on unseen classes.

Therefore, the labelled class set C of a dataset is split into two sets, C_{train} and C_{test}, such that $C_{train} \cap C_{test} = \varnothing$. An episode is then formed by randomly picking a label class c from C_{train}. For each class c, k unique images and masks segmenting class c constitute the support set S, along with one image query I_q. The training objective is:

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} E_{S,(I_q,M_q)}\left[\log P_\theta\left(M_q | I_q, S\right)\right].$$

That is, the model training corresponds at finding a set of parameters $\widehat{\theta}$ that maximises the expected value $\mathbf{E}_{S,(I_q,M_q)}$ of the posterior probability for the ground truth mask M_q given the query image I_q and the support set S for the distribution log P_θ. The aim here is, over multiple episodes, to guide the model learning to generalize despite the limited number of examples provided in the support set.

3.4.3. HDMNet

For the experiments in this paper, we chose to focus on a specific type of FSS architecture known as Hierarchically Decoupled Matching Network (HDMNet) (Peng et al., 2023) as depicted in Fig. 4. At the time of writing, HDMNet provides the state-of-the-art performance for FSS (one-shot and five-shot) on the MS COCO Computer Vision benchmark.

HDMNet (Peng et al., 2023) implements FSS by relying on a support set that comprises of both full images and object region masks. First, the model filters out the background of the support images based on the area of the provided masks. Then, it extracts features from both the support and the query images, using a pretrained backbone where weights are kept fixed. These support and query features undergo L stages of self-attention and downsampling operations. These processing steps are conceived for deriving a set of hierarchical features that facilitate the exploration of inter-scale correlations among features within the input images.

Let the query and support images be of dimensions $[H, W, 3]$. Each downsampling block then generates intermediate feature maps of spatial dimensions:

$$h_l^{q/s} = \frac{H^{q/s}}{2^{l+2}}, \quad w_l^{q/s} = \frac{W^{q/s}}{2^{l+2}}.$$

The resulting intermediate query and support feature maps, denoted as $\left\{F_l^q\right\}_{l=1}^{L}$ and $\left\{F_l^s\right\}_{l=1}^{L}$ from the L stages, are aggregated to compute: a correlation matrix $\left\{C_l \in R^{h_l^q w_l^q \times h_l^s w_l^s}\right\}_{l=1}^{L}$, as well as an enriched feature set $\left\{X_l \in R^{h_l^q w_l^q \times h_l^s w_l^s}\right\}_{l=1}^{L}$. These enriched features are fed to a decoder to generate the final prediction $\widehat{M}$.

The decoder employs an inverse process for each enriched map X'_{l+1}. Namely, it first applies bilinear interpolation to upsample the map so that it matches the spatial resolution of X'_l. Then, a Multi-Layer Perceptron (MLP) is applied to both feature maps, along with a residual connection layer. In mathematical terms, this operation can be expressed as:

$$X'_l = ReLU\left(MLP\left(X_l + \zeta_l\left(X'_{l+1}\right)\right)\right) + \zeta_l\left(X'_{l+1}\right).$$

Here, ζ_l represents the bilinear interpolation operation. Finally, the query mask $\widehat{M} \in R^{H \times W}$ is predicted by combining a 1×1 kernel size convolution to X'_1 with bilinear up sampling.

3.4.4. YOLO

YOLO was originally developed as an object detection algorithm by Redmon et al. (2016), although it has been gradually adapted to fit various CV tasks beyond object detection, such as classification, segmentation, tracking, and pose estimation. However, the core design principle of processing images in a single stage has remained unchanged from the original YOLO model (Redmon et al., 2016). Its key feature is that it performs object detection in a single pass through a single convolutional network, dividing the image into a grid and predicting bounding boxes and class probabilities for each cell in the grid.

This configuration simplifies the object detection pipeline, by

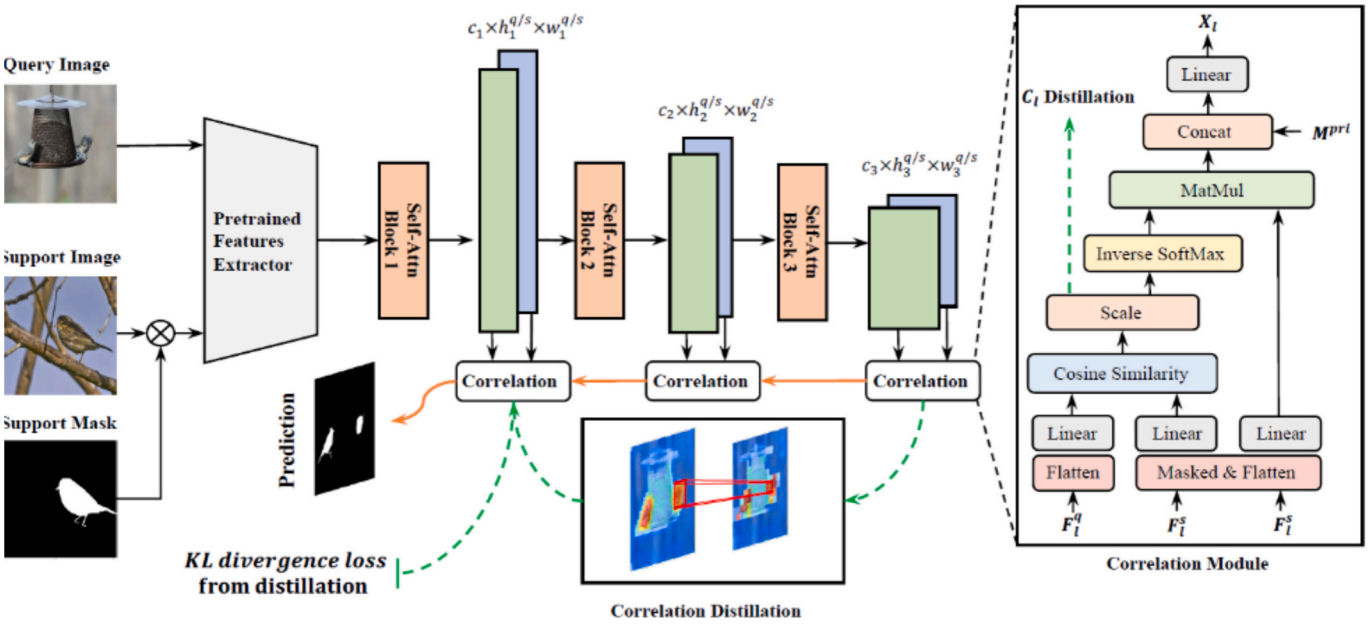


Fig. 4. Architecture of HDMNet as proposed by Peng et al. (. adapted from Peng et al., 2023)

framing the task as a single regression problem. As a result, the overall inference times are generally lower than those obtained with systems such as region-based detectors, where detection and classification are tackled in separate stages.

In this work, we present the results obtained with the YOLOv5 and YOLOv8 model versions, since their baseline performance, ease of use and off-the-shelf integration with the Roboflow data annotation and management system drastically simplify the bootstrapping of Computer Vision solutions in a new application scenario. Thus, evaluating YOLO-based methods on weed segmentation tasks provides a relevant representation of the de facto methodological choice that currently characterizes many applications across Field Robotics and Agricultural Robotics.

3.4.4.1. YOLOv5. The segmentation architecture of YOLOv5, introduced by Ultralytics LLC in May 2020 (depicted in Fig. 5), is organized into four key modules: i) the backbone Network, ii) the model neck, iii) the prototypical mask branch and iv) the prediction head.

The function of the backbone is extracting an initial set of features from input images through a pre-trained CNN. To reduce the number of training parameters and accelerate training, the backbone module implements a Cross-Stage Partial (CSP) Network (Wang et al., 2019) and a Focus module. The CSP network partitions the input feature map by adding cross-stage convolutions and lower-dimensionality (i.e., bottleneck) layers for feature refinement. The neck layer is conceived to extract feature pyramids from the backbone. Specifically, a Path Aggregation Network (PAN) is applied to aggregate the input features and produce a fixed-length, denser vector representation through a Spatial Pyramid Pooling (SPP) layer. The SPP layer separates crucial features and enhances the receptive field by subdividing the image into divisions and aggregating local features in each division (He et al., 2015). The refined features extracted through the neck[4] module are then passed in parallel to both the model head and a dedicated segmentation branch, similar to the Protonet structure described by Bolya et al. (2019). This branch computes image-sized, prototypical masks, i.e., masks that are independent from individual object instances.

Finally, the prediction head outputs the bounding box, object classes, and prediction scores respectively. In the case of segmentation tasks, this stage also produces a vector of k mask coefficients (one for each prototype). Non-maximum Suppression (NMS) is then applied to remove duplicate predictions and select the most relevant object proposals. Lastly, the remaining masks post NMS and after thresholding by confidence score are linearly combined with the prototypical masks, yielding the final segmentation results.

3.4.4.2. YOLOv8. YOLOv8, proposed by Ultralytics in May 2023 (Jocher et al., 2023), at the time of writing is the newest version in the YOLO model series. It builds upon YOLOv5, introducing a few key enhancements represented also in Fig. 6.

The first significant improvement lies in the neck layer, where YOLOv8 incorporates the PAFPN module to generate feature pyramids. PAFPN is another component of Path Aggregation Networks that implements bottom-up path augmentation. Bottom-up path augmentation shortens the path between top-level features and lower-level features in the pyramid structure, by adding skip connections. This configuration comes with the advantage of compacting the information path, mitigating the information loss from topmost to lower-level features.

Compared to YOLOv5, features at the neck level are concatenated together without forcing vectors to the same channel dimension, thus reducing the overall number of training parameters.

Another difference between YOLOv5 and YOLOv8 relates to relying on anchor boxes at training time. Indeed, previous YOLO versions (including version 5), predict object boxes as regions shifted by a certain offset with respect to a set of reference anchors. Because anchors implicitly encode the distribution of the pre-training set, they potentially hinder generalization. As such, the anchor-free detection pipeline in YOLOv8 provides, in principle, a more scalable solution than its YOLOv5 predecessor.

It is worth noting that both YOLOv5 and YOLOv8 provide the option to downsample object masks to increase the training speed. By default, in YOLOv8, masks are downsampled by a factor of 4. This design choice may impact the ability of the model to segment small objects, a common scenario when detecting weeds since their earliest growth stages. We will further discuss the implications of this downsampling strategy on the evaluation scores in the experiment section.

[4] Available online at: https://github.com/ultralytics/yolov5/issues/280.

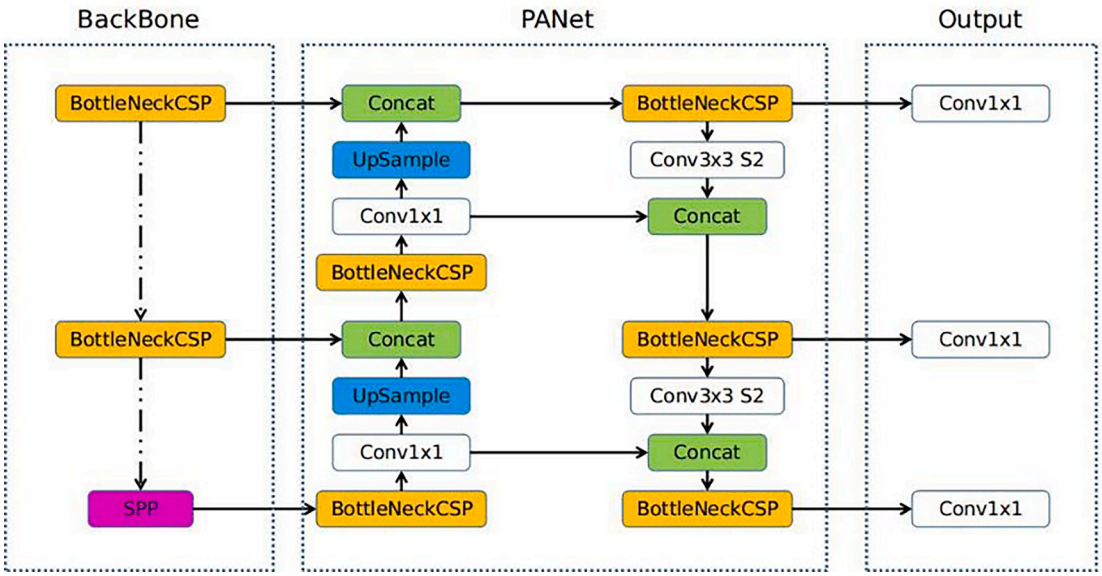


Fig. 5. Architecture of YoloV5, Image courtesy of Ultralytics.

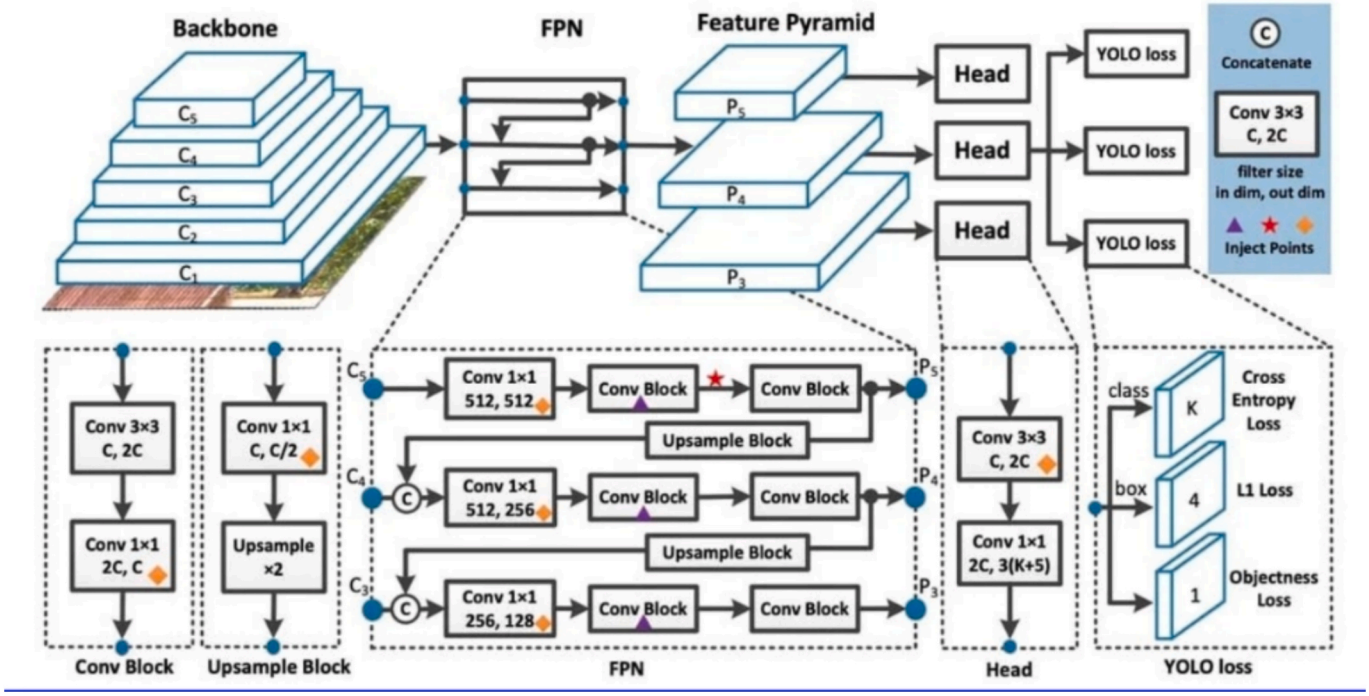


Fig. 6. YOLOv8 Architecture Original Image courtesy of Ultralytics(Available online at: https://yolov8.org/what-is-yolov8/).

3.4.4.3. Experiments. In this section, we illustrate the experimental design for evaluating the performance of YOLO (in its YOLOv5 and YOLOv8 versions) and HDMNet. Here we also specify the evaluation metrics adopted in our analysis. These metrics not only gauge the model segmentation performance when varying quantity and quality of training data is available, but also quantify the human labor involved in the different phases of dataset creation.

3.4.5. Experiments on YOLO

As anticipated in Section 3.3, we prepared two datasets for this experimentation (**Early** and **Refined**). This duality aims to assess the impact of changing the granularity of crop and weed region annotations on the performance of the YOLO family of methods.

Specifically, we seek to understand the impact on model performance of relying on training data that are larger in scale but noisier (i.e., the **Early** set), as opposed to capitalizing on fewer densely annotated and accurate annotations (i.e., the **Refined** set). To this aim, our experiments start from the scenario where models under comparison are trained on the early and refined sets respectively. Then, the performance for both sets of trials is opportunely compared.

Additionally, we are interested in studying the performance effects of training models on the simpler-to-construct **Early** dataset, while considering the more challenging **Refined** dataset as test set, where even smaller leaf structures are labelled.

Finally, we evaluate the generalization capability of the models under changing conditions by training them on the already available

2019 ACRE Competition Dataset and test them on both **Early** and **Refined** datasets.

To systematically address the research objectives, we conducted three experiments, organized as outlined in Table 3.

3.4.6. Experiments on HDMNet

With a similar objective, we conducted experiments with the HDMNet FSS model on the **Early** and **Refined** sets. To test the efficacy of applying FSS models with minimal supporting examples, we evaluated HDMNet using the weights from training on the MS COCO dataset and directly tested the model on the **Early** and **Refined** sets without any additional training.

We also evaluated a second testing scenario of intermediate difficulty. Namely, we trained HDMNet on the **ACRE 2019 Dataset**, which resembles the same domain as the target set, i.e., weed and crop examples collected in bean and corn cultivations, while introducing a few environmental differences (as highlighted in Section 3.3). In line with the FSS literature we conducted the experiments with both 1 and 5 support examples. The structure of experiments conducted in this phase is also summarized in Table 4.

3.5. Evaluation metrics

To assess the performance of the different models under comparison, we measure the segmentation accuracy across classes through the Intersection over Union (IoU) and Mean Intersection over Union (mIoU) metrics. Additionally, we introduce the human hours metric to gauge the manual effort involved in dataset preparation. Further details on these metrics and their formulations are provided in the following sections.

3.5.1. Model performance

First, we took the *IoU* metric as a reference to access and compare the performance of HDMNet and the YOLO models on crop and weed segmentation tasks. Moreover, since we tested models on 4 distinct classes (i.e., Corn, Bean, Weed Corn, and Weed Bean) we also used the *mIoU* as a summary metric of performance across experiments. We chose the *IoU* as a suitable metric of the segmentation accuracy since it evaluates the spatial overlap between predicted and ground truth masks.

In formal terms, let c be the subject class to segment, then the formulation of the *IoU* can be derived as follows:

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c},$$

where TP_c, FP_c and FN_c are the number of pixels that are respectively true positives, false positives and false negatives predicted to mask a given class c.

The *mIoU* is defined as the average of *iou* for all classes:

$$mIoU = \frac{1}{C}\sum_{c=i}^{i=C} IoU_c,$$

where C is the total number of classes.

3.5.2. Human effort assessment

To gauge the effort invested in constructing the experimental field, acquiring data, and labelling images, we tracked the human hours spent during all phases of data and field preparation.

Table 3
Table summarizing the experiment conducted on YOLOv5 and YOLOv8.

Training dataset	Test dataset
Early set	Early set
Refined set	Refined set
Early set	Refined set
Acre 2019	Early set
Acre 2019	Refined set

Table 4
Table summarizing the experiments conducted on HDMNet.

Training dataset	Test dataset
MS COCO	Early set
MS COCO	Refined set
Acre 2019	Early set
Acre 2019	Refined set

Significantly, this metric gains relevance in evaluating the labeling phase's impact on the trade-off between dataset richness and labor intensity in dataset preparation. Notably, the **Early** dataset, despite containing more labeled images, required considerably fewer human hours per picture for completing the annotation compared to the **Refined** dataset. This contrast reveals the intricate relationship between dataset richness and the labor intensity required for data preparation. Importantly, it provides a set of reference figures and valuable insights that can guide future efforts in the creation of improved benchmarks for Computer Vision in Precision Agriculture. To estimate the cost of developing a labeled dataset for image segmentation in agricultural contexts, we calculated the total cost by multiplying the hours spent on labeling by the corresponding hourly wage. The selected wages were based on official rates for professionals with the necessary expertise to produce the labels, ensuring an accurate representation of labor costs. Specifically, we utilized data from the National Collective Labor Agreement for Agricultural and Floricultural Workers and the University of Pisa's Costs and Wage Tables (see Table 8). These sources were chosen because they provide the most accurate wage estimates for the context of this study. They reflect Italian wage standards and include cost tables from the University where the larger portion of labeling occurred, ensuring that the cost estimation aligns closely with the conditions under which the experiments were conducted. By including this information, our study addresses a gap in the literature by offering a concrete reference point for estimating annotation costs in similar scenarios.

4. Implementation details

To implement and train the **YOLO models**, we used The Google Colaboratory platform (Colab), offered by Google. Colab, a Cloud-based service based on Jupyter Notebooks, provides a free single 12 GB NVIDIA Tesla K80 GPU. YOLOv5s model (YOLOv5s-seg summary: 255 layers, 7,413,608 parameters, 25.9 GFLOP) and YOLOv8s model (YOLOv8s-seg summary: 261 layers, 11,791,257 parameters 42.7 GFLOP) were trained on both Early and Refined datasets. In the case of both datasets, models were trained for 100 epochs and with a batch size of 16. Images sizes were resized to a 640 x 640 resolution and a Stochastic Gradient Descent (SGD) optimizer was used, with momentum set to 0.937. The learning rate was set to 0.01. We used the Ultralytics default implementation for both YOLO versions (v5, and v8). To improve the training success, the Refined dataset was augmented by applying Horizontal flipping, Vertical flipping, Rotation ($-15°$ and $+15°$), Hue ($-50°$ and $+50°$), Saturation (-50% and $+50\%$), and Noise (10% of pixels) transformations.

HDMNet was trained on an Nvidia Quadro RTX 6000 24 GB GPU, keeping the default configuration in terms of downsampling the training images to 633x633 pixels and the types of data augmentation techniques applied. Specifically, the augmentation process consists of the following operations: randomized scaling within the range of 0.8 to 1.25, randomized rotation between $-10°$ and $10°$, randomized Gaussian blur, and a randomized horizontal flip. The default training procedure ranged over 50 epochs, with a batch size of 6 and a total of 50,898,627 trainable parameters. As in the case of YOLO-based methods, we used SGD as optimizer, with a momentum of 0.9, a weight decay of 0.01, and Cross-Entropy Loss as optimization function. It is important to note that, due to

limited GPU memory, training HDMNet in a 5-shot configuration was not feasible. Consequently, only 1 shot at a time was fed to the model during training. Experiments involving the 5-shot configuration were conducted using weights from training on MS COCO by Peng et al. (2023), introducing the larger, 5-shot support set only at inference time. Among the four available splits of the broader COCO dataset, we always experimented using the subsample denoted as "split 0" in the official GitHub implementation of HDMNet.

4.1. Note on the YOLO evaluation pipeline

In the default implementation of YOLOv8 and YOLOv5 by Ultralytics LTD., the input images are downsampled to reduce the larger spatial dimension to 640 pixels. Given an input image of 1280x720px, the resulting prediction volume is reduced to a spatial resolution of 168*x*96. In the default evaluation script, ground truth masks are downsampled to match the spatial resolution of this prediction volume: i.e., 168*x*96. We initially adopted this default configuration to assess the off-the-shelf performance that can be achieved when widely adopted YOLO models are applied to real-world autonomous weeding scenarios. This approach, however, brings along an increased tendency to overlook the finer details in the image, as objects of modest size such as small leaves disappear from the prediction mask because of the downsampling. These overlooked regions introduce a bias in the IoU metrics detailed in Section 3.6.1, as downsampling by factor greater than 7 might let finer details and small instances disappear from the ground truth, leaving only the subjects of greater size to contribute to the IoU counts. Hence, the IoU scores may appear overly inflated as a result.

In contrast, HDMNet, in its default configuration, generates a prediction volume with a spatial resolution of 633*x*633. This prediction volume is subsequently up sampled to match the resolution of the input image before computing the IoU. This methodology ensures a more faithful representation of details, minimizing the loss of information.

To facilitate a fair comparison between these model families, we chose to manually set the input image resolution for YOLO during evaluation to the actual dataset image resolution, therefore avoiding any preprocessing downsampling, by specifying $imgsz = 1280$. Similarly, to bypass any downsampling of the ground truth we set $mask\_ratio = 1$. This decision results in a prediction space resolution of 184*x*328 and a padded ground truth of 736*x*1312*px*. Subsequently, we modified the evaluation code for both YOLOv5 and YOLOv8 to up sample the prediction volume to the resolution of the ground truth. The objective of the introduced modification is reverting the effects of downsampling. We will present the results obtained with this adjusted configuration in the following Section (Table 6).

5. Results

In this section, we present the results of our experiments, providing both quantitative and qualitative insights. The quantitative results concern the segmentation scores of the tested models, as well as the human hours invested in dataset creation. The qualitative results then further situate the explanation of results by showcasing selected examples of model predictions.

The model weights for both YOLO and HDMNet are based on the best performing weights. For YOLO those weights are selected according to a fitness function developed by Ultralytics as a weighted sum of mAP@0.5, mAP@0.5:0.95 with facto 0.1 and 0.9 respectively over the validation set. For HDMNet the best configuration is selected according to the highest mIoU score on the validation set. Further configuration details can be found in Section 4. Finally, in line with the FSS literature we tested HDMNet in both 1 and 5 shot. Here it is important to note that the HDMNet experiment on 1-shot included a training phase on ACRE 2019, while in 5-shot we could not replicate the training on agricultural data due to GPU memory constraints, and such we used a configuration pretrained on the MS COCO dataset.

5.1. Quantitative results

Table 5 reports the quantitative performance results (in terms of IoU scores) obtained where different versions of YOLO are used. Moreover, test results on the Early and Refined sets are also reported for HDMNet pre-trained on ACRE 2019.

As can be noted from Table 5, YOLOv5 achieved the top performance across different training configurations and class types (crop and weed). In particular, the highest results were reached in the scenario where YOLO models are trained on the Early set and tested of a different split of the same set. This result was expected and serves as a performance upper bound. Indeed, sampling training and test examples from the same distribution presents the least challenging scenario among the ones being tested. Moreover, in the early set only the larger and easier to predict leaves are labelled. At the same time, the early set provides the largest number of training examples.

The second best in terms of IoU scores for most classes is the case where YOLO models were trained on the ACRE 2019 set and tested on the Early set. One explanation we can provide for these results is the more forgiving nature of object masks in the Early set. Indeed, ground truth masks in the Early set were not labelled with the same granularity as the Refined set, partially inflating the IoU scores.

This explanation is further corroborated by the results obtained when models are trained and tested on two different subsamples taken from the Refined set. Despite the introduced simplification of sampling data from the same distribution, this case does not reach the upper bound performance. Indeed, segmenting refined masks is inherently more challenging.

We also examine a more challenging configuration when models are trained on the ACRE 2019 set (collected in Montoldre, France, Auvergne-Rhône-Alpes region by four different robots) and tested on different splits of the ACRE 2023 dataset (collected in Cornaredo, Italy, Lombardia region by a fifth new robot). Namely, this setup allows us to investigate the model adaptability to a different geographical region, soil conditions, field preparation protocols and different robotic platforms. Results show that, with as little as one training example per class, the FSL method HDMnet already manages to reach a high proportion of the top performance achieved with YOLO on both crop types, through a more costly training routine. Specifically,: $\sim$ 78% of the top performance was achieved on Corn crops; $\sim$ 80% on Bean crops when testing on the Early set; $\sim$ 74% and $\sim$ 73% for Corn and Bean crops when testing on the Refined set. Nevertheless, results are much more modest in the case of weed segmentation. The HDMNet IoU on the Early set only amounts to $\sim$ 17% of the top performance for Corn weeds and to $\sim$ 16% for Bean weeds. When testing on the Refined set, HDMNet reaches only $\sim$ 24% of the YOLO scores on Corn weeds. In the extreme case of Bean weeds, HDMNet only amounts to $\sim$ 2% of the top-performance scores obtained with YOLO. We can explain these results by considering that weed segmentation is an inherently more challenging task than crop segmentation, due to the small size and scattered configuration of weed masks.

Interesting insights can also be gathered from the results of the last configuration listed in Table 5: the scenario where models are trained on the Early set and tested on the Refined set. Despite the high scores achieved on the crop class, this setup led to the lowest performance for weed segmentation. Although weeds are generally more difficult to segment, a trend that is reflected in the scores from all trials, IoU metrics are particularly low in this case, with a near-zero performance on the corn weed species.

As anticipated in Section 4.1, YOLO downsamples masks when evaluating results in its default configuration. Therefore, to allow a fair comparison of the YOLO and HDMNet results across the different sets, we repeated the trials reported in Table 5 after customizing the evaluation procedure, following the method described in Section 4.1. Results from this set of experiments are listed in Table 6. One can note from comparing Table 5 with Table 6 that skipping the downsampling

Table 5
Table comparing the IoU score and number of training samples of YOLOv5, YOLOv8 and HDMNet trained and tested on different dataset with their default configuration. In bold are highlighted the best score for each category.

Training Dataset	#Training Samples		Test Dataset	Model	Crop		Weed		mean IoU
	corn	bean			Corn	Bean	Corn	Bean	
ACRE 2019	500	500	Early 2023	YOLOv5	**0.9663**	**0.9094**	**0.8015**	**0.8744**	**0.8879**
				YOLOv8	0.9318	0.8182	0.6934	0.794	0.8093
				HDMNet – 1 Shot	0.7515	0.7243	0.1454	0.1578	0.4447
ACRE 2019	500	500	Refined 2023	YOLOv5	**0.9854**	**0.8706**	0.6228	**0.8511**	**0.8324**
				YOLOv8	0.9426	0.7973	**0.6773**	0.8078	0.8062
				HDMNet – 1 Shot	0.7308	0.6481	0.1614	0.0211	0.3903
Early 2023	1326	743	Early 2023	YOLOv5	**0.9685**	**0.9755**	**0.8544**	**0.8363**	**0.9086**
				YOLOv8	0.9358	0.961	0.7494	0.7951	0.8603
Refined 2023	51	51	Refined 2023	YOLOv5	0.9091	0.8985	**0.7297**	**0.7988**	**0.834**
				YOLOv8	**0.9347**	**0.9378**	0.7145	0.7116	0.8246
Early 2023	1326	743	Refined 2023	YOLOv5	**0.9717**	**0.9302**	**0.0263**	**0.1663**	**0.5236**
				YOLOv8	0.9383	0.9111	0.0188	0.1364	0.5011

Table 6
Table comparing the IoU score and number of training samples of YOLOv5, YOLOv8 trained and tested on different datasets using a custom evaluation script. The scores are computed without donwsampling the ground truth, to align results with the default configuration of HDMNet. HDMNet results from Table 4 are also reported here to facilitate the comparison. Notably, the last row shows how HDMNet have comparable or better scores than YOLO on crop categories even without any training on the specific test categories. The best scores for each category are highlighted in bold.

Training	#Training Samples		Test Dataset	Model	Crop		Weed		
Dataset	corn	bean			Corn	Bean	Corn	Bean	mean IoU
ACRE 2019	500	500	Early 2023	YOLOv5	0.575	0.676	**0.4924**	**0.5686**	**0.578**
				YOLOv8	0.5394	0.6005	0.3849	0.4963	0.5053
				HDMNet – 1 Shot	**0.7515**	**0.7243**	0.1454	0.1578	0.4447
ACRE 2019	500	500	Refined 2023	YOLOv5	0.4935	0.6408	0.1859	0.4421	**0.4406**
				YOLOv8	0.4256	0.573	**0.2079**	**0.4557**	0.4156
				HDMNet – 1 Shot	**0.7308**	**0.6481**	0.1614	0.0211	0.3903
Early 2023	1326	743	Early 2023	YOLOv5	**0.5392**	**0.7836**	**0.5431**	**0.5441**	**0.6025**
				YOLOv8	0.5252	0.7758	0.4784	0.5241	0.5759
Refined 2023	51	51	Refined 2023	YOLOv5	0.37	0.6618	0.0815	0.2651	0.3446
				YOLOv8	**0.3835**	**0.7328**	**0.097**	**0.296**	**0.3773**
Early 2023	1326	743	Refined 2023	YOLOv5	**0.546**	**0.751**	**0.0103**	**0.0854**	**0.3485**
				YOLOv8	0.542	0.736	0.0094	0.0835	0.343
Pretrained on COCO	0	0	Early 2023	HDMNet – 5 shot	0.5565	0.6352	0.0042	0	0.299
			Refined 2023	HDMNet – 5 shot	0.527	0.5435	0.0312	0.008	0.2774

procedure led to a significant drop in the performance of both YOLOv5 and YOLOv8 in all segmentation classes (weed and crop types). With these newly-computed metrics, HDMNet achieved a comparable or increased performance on the crop classes compared to YOLO. Specifically, HDMNet achieved the top performance on crop types when pretrained on the ACRE.

2019 set and tested on either the refined or the early sets. In the case of corn and bean crops, the IoU obtained with HDMNet pre-trained on COCO, with access to only 5 shots from the target set, is even higher than some of the YOLO trials, particularly when YOLO is pretrained on ACRE 2019. Crucially, while, in the latter cases, YOLO was trained on a purposely built dataset, representative of the agricultural domain, HDMNet was simply pretrained on the general-purpose MS COCO dataset, i.e., on a different domain and task from weed and crop segmentation.

As can be noted from the last two rows in Table 7, however, results achieved with HDMNet only pre-trained on COCO are generally lower than those achieved with YOLO for bean crops and are also reached at the expense of weed predictions, which amount to the lowest IoU scores among all sets of trials in Table 6. Indeed, in this scenario, the model is only relying on one or five shots to learn from, as opposed to the larger training sets of previous trials.

However, by comparing the HDMNet results in one-shot and five-shot settings it can be noted that even adding a limited number of examples to the support set has the potential to drastically improve the model performance. Of particular note is the case of crop segmentation where relying on 5 shots partially bridges the gap towards achieving the high-performance scores of YOLO.

Table 8 presents a comprehensive assessment of the human effort invested in the development of the two studied datasets for ‘Bean’, ‘Corn’ and ‘Weed’ segmentation. This assessment focused on various phases, spanning from field operations to labeling processes. Results provide insights on the requirements of collecting image collections that

Table 7
IoU scores of HDMNet pretrained on the MS COCO Dataset and no additional training in1 and 5 shot configuration. The best scores for each category are highlighted in bold.

Training Dataset	Test Dataset	Model	Crop		Weed		mean IoU
			Corn	Bean	Corn	Bean	
Pretrained on COCO	Early 2023	HDMNet – 1 shot	0.4695	0.5668	**0.0145**	**0.0034**	0.2635
		HDMNet – 5 shot	**0.5565**	**0.6352**	0.0042	0.0000	**0.2990**
Pretrained on COCO	Refined 2023	HDMNet – 1 shot	0.4312	0.4899	0.0174	0.0068	0.2363
		HDMNet – 5 shot	**0.5270**	**0.5435**	**0.0312**	**0.0080**	**0.2774**

Table 8
Human effort assessment for Early and Refined dataset development. This Table outlines the time (min) invested in various stages, including field operations, seeding/transplantation, image acquisition, and labeling.

Field operation	**Time (min)**			**Operation total cost (€)**
Ploughing*	18.36			2.57
Stone burier*	9			1.26
Harrowing*	9			1.26
Seed drill*	8.82			1.24
Manual transplantation^{+}	6720			863.52
Manual weeding^{+}	2520			323.82
Images acquisition^{o}	**Time (min)**			**Operation total cost (€)**
Images	Average	Range	Total	
2069	0.033	(0.03–0.08)	95	15.96
Labelling^{o}	**Time (min)**			**Operation total cost (€)**
Dataset	Average	Range	Total	
Early (2069)	0.93	(0.25–1.5)	1507.13	253.20
Refined (102)	16.71	(4–38)	1778.35	298.76
Total Early	10887.31 (~181 h)			1462.83
Total Refined	11158.53 (~186 h)			1508.40

*Specialized agricultural worker: 8.41 € per hour (from the National Collective Labor Agreement for Agricultural and Floricultural Workers (Available online at: https://www.flai.it/wp-content/uploads/2023/11/CCNL-Operai-Agricoli-e-Florovivaisti-2022–2025.pdf.)).
$^{+}$Basic agricultural worker: 7.71 € per hour (from the National Collective Labor Agreement for Agricultural and Floricultural Workers4).
^{o}Average rate for postdoctoral researchers and PhD students: 10.08 € per hour (from the University of Pisa Costs and Wage Tables (Available online at: https://www.unipi.it/index.php/costi-e-tabelle-retributive/item/2141-importi-assegni-di-ricerca%OB.)).

fit the agricultural domain, while increasing awareness on the human effort required in each stage of dataset creation.

The human effort analysis revealed that apart from the necessary time for the field preparation, the more striking differences between datasets of different annotation (mask) granularity emerge at the labelling stage. The dataset acquisition and labeling process required significant labor, especially for manual tasks and image labeling. The overall time commitment for the dataset preparation, including field operations, manual tasks, image acquisition, and labeling, exceeded 360 h, with a total estimated cost of approximately 1460 € for the early dataset and 1500 € for the refined dataset. The labeling cost was approximately 250 € for the Early dataset and 300 € for the Refined dataset. Despite having fewer images, the Refined dataset required 5 additional hours for labeling compared to the Early dataset. This increase in time was due to the need for thorough visual inspection and correction of object masks generated by the Segment Anything Model, especially given the high number of weed instances and the need for precise annotation of small regions. Additionally, this process demands a deeper understanding of plant morphology to accurately distinguish between plants, particularly in their specific phenological stages, ensuring precise differentiation between weeds and crops.

Overall, the labelling time was 5 h higher for the Refined dataset compared to the Early set, even though the Refined set consists of a lower number of images. This resulted from the longer times required for visually inspecting and correcting the object masks output by the SAM, due to the high number of weed instances to be annotated, including small-sized regions to be accurately labelled.

5.2. Qualitative results

This section illustrates a set of selected predictions gathered from the YOLOv5 and HDMNet results.

In Fig. 7, we present examples of weed and crop segmentation from the bean and corn cultivations examined in our study. Training the model on the early dataset is insufficient for accurately detecting weeds, as evidenced by the absence of weed predictions in pictures a) and c) of Fig. 7. Conversely, the model trained on the Refined dataset demonstrates improved weed identification capabilities as evident in b) and d). Importantly, both models exhibit high-quality crop segmentation with similar predictions a) and b). Nonetheless, while comparing these two predictions, it must be noted that YOLO trained on the Early dataset is correctly segmenting the leaves in the right bottom corner of the image, but mistakes weed leaves for bean crops.

5.2.1. YOLOv5

Fig. 8 showcases the qualitative outcomes achieved through HDMNet when tasked to segment a) bean crops, b) bean weeds, c) corn crops and d) corn weeds in a one-shot configuration. In each subpicture we show the results employing both a model pre-trained on the COCO dataset (third column) and a model trained specifically on the ACRE 2019 dataset (fourth column).

Across all the four queries, the size of the detected subject emerges as one key success factor. The prediction of bean crops is already

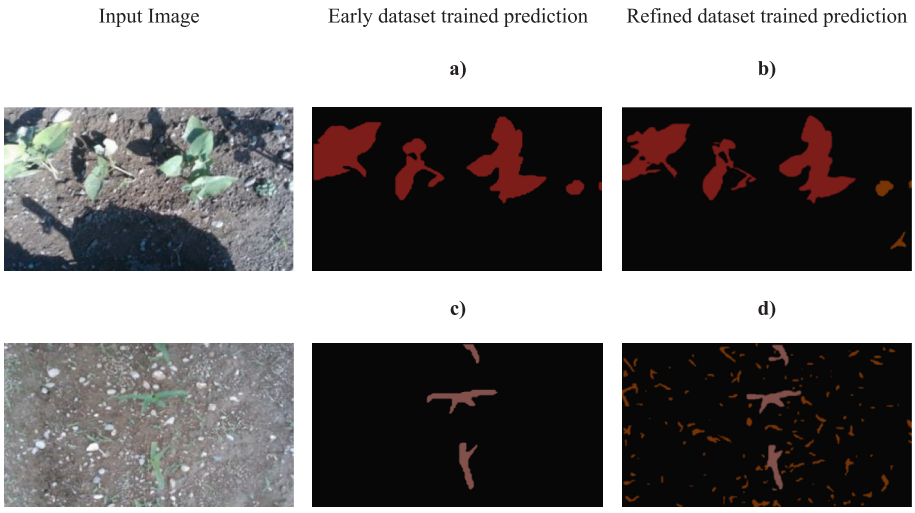


Fig. 7. Qualitative examples on weed and crop segmentation on bean cultivation (first row) and corn cultivation (second row) using YOLOv5. The second column displays the predictions when the model has been trained on the Early dataset, while the third column shows the prediction when the Refined dataset is used for training. The bean crop is masked in red, the corn crop is masked in salmon, the weeds for both cultivations are masked in orange.

satisfactory with the COCO pretrained model. Pre-training on ACRE 2019, in the case of bean crops, only improves the model ability to accurately segment edges – Fig. 8a). On the other hand, the model could only partially segment the relatively smaller corn crop leaves, which were instead more accurately segmented after pre-training on ACRE 2019. Finally, weeds are the most difficult subjects to detect. In particular, the model pretrained on COCO misclassifies a portion of the bean crop regions as weed, as can be seen in Fig. 8b). The model pretrained on ACRE 2019, while avoiding misclassification, still misses out a large portion of the weed leaf area. The same model (i.e., pre-trained on ACRE 2019) exhibits a marginally improved ability to segment corn weeds, Fig. 8d), while the model pretrained on COCO still fails to detect the target weed regions completely.

5.2.2. HDMNet

5.3. Failure cases

In this section, we analyze failure cases observed for HDMNet in two different training configurations: pretrained on MS COCO and trained on ACRE 2019.

Fig. 9 showcases failure cases for HDMNet trained on the MS COCO dataset. In the first block, the model attempts to segment a small weed leaf in the bottom right corner. Based on the failure examples, the small size of the weed and differences in soil, lighting, and rotation between the query and support images appear to lead HDMNet to erroneously classify parts of the robot wheel and bean leaves as weeds. The second block further illustrates these difficulties, with the model misclassifying a large area of the ground as weeds while failing to detect actual weeds in the bottom left of the image. Although this error is less severe than the first, as corn leaves are not misclassified, the implications for practical applications such as weeding robots are significant, as these errors may lead to crop damage.

Fig. 10 presents failure cases for HDMNet trained on the ACRE 2019

Fig. 8. Qualitative example of crop and weed segmentation on bean and corn cultivation using HDMNet. In each block the first row depicts the support set, the second the query image and the ground truth. The third column illustrates the prediction of HDMNet on the query image when no additional training is performed beside the pretraining on COCO, and the last column when additional training on the ACRE 2019 dataset is performed.

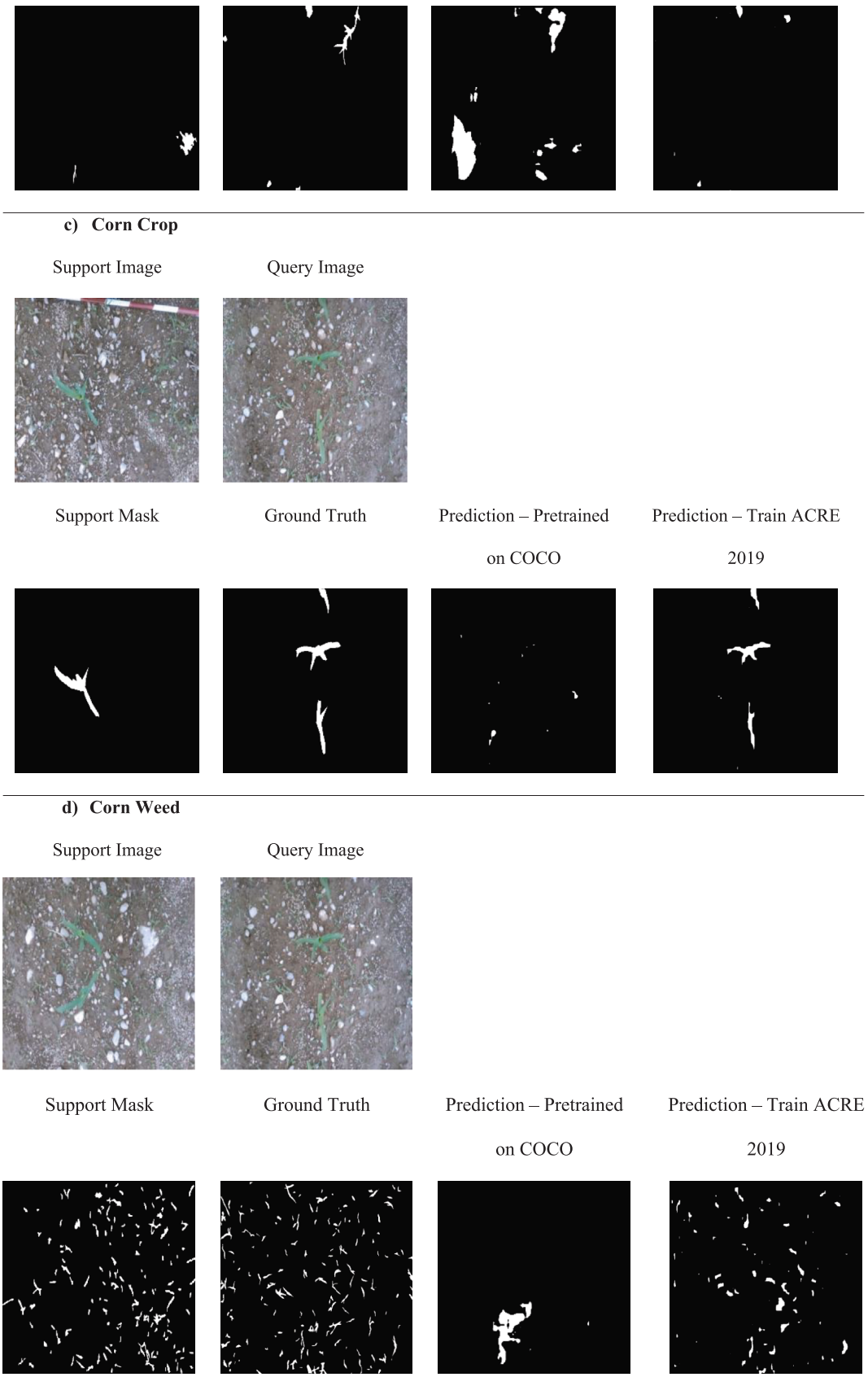


Fig. 8. (*continued*).

dataset, which includes images of corn, beans, and weeds from a different geographical location, robot platform, and year, as discussed in Section 3.3. Training on ACRE 2019 data reduces the error rate produced by HDMNet significantly compared to training on MS COCO

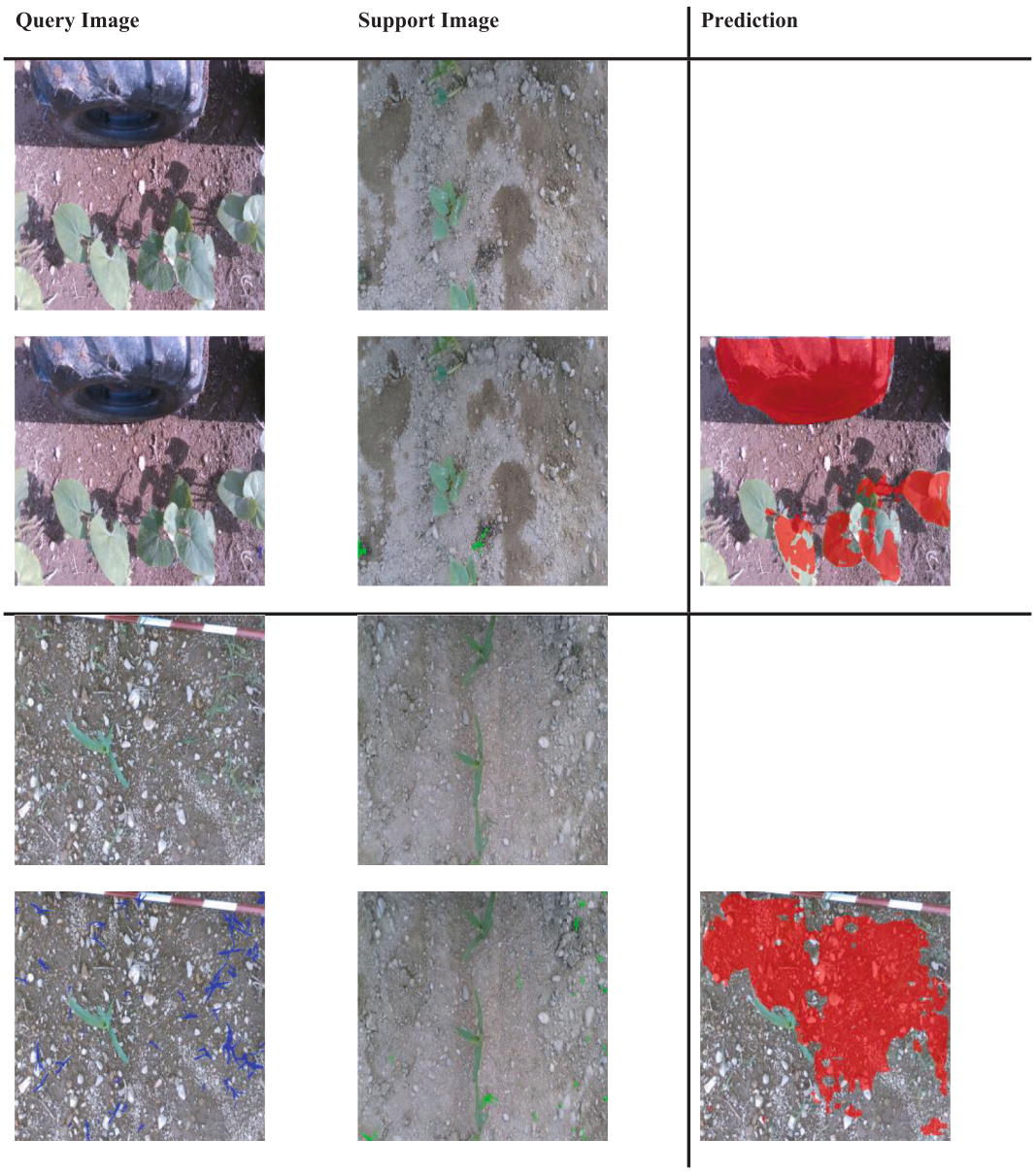


Fig. 9. Failure examples of HDMNet pretrained only on the split 0 of COCO on images of beans (top group) and corn (bottom group) cultivation. For both the support and query image we report the ground truth and label in the second row of teach group.

alone. In the first block, the model under-segments bean leaves; this under-segmentation may be due to differences in ground conditions and lighting, resulting in false negatives but no false positives. In the second block, although small weeds remain challenging, the model's error rate is lower, demonstrating improved performance over the MS COCO training configuration.

Fig. 11 present a collection of failure cases of YoloV5 and YoloV8, revealing critical insights into the impact of training dataset on this family of models. Models trained on the ACRE 2019 exhibit significant misclassification errors, such as YoloV5 misclassifying rocks as bean leaves and YoloV8 over-segmenting background areas as weeds while partially failing to detect crops. The domain shift between the training set and the data collected for these experiments can be identified as the most important factor leading to the observed errors, posing significant questions about the generalization capabilities of these models.

On the other hand, training on datasets collected during this study reduces domain shifts but highlights a tradeoff between dataset quantity and quality. The Early 2023 dataset, with fewer weed instances, minimizes crop misclassification but fails to adequately train the models to detect weeds. Conversely, the smaller yet more balanced Refined 2023 dataset, which includes labels for smaller weed leaves, improves weed predictions but introduces false positives and crop confusion, such as rocks misclassified as weeds by YoloV8 and crop species confusion by YoloV5.

In sum, the presented failure cases highlight how dataset selection plays a critical role in the performance of Yolo models. Training on readily available dataset, avoids labeling effort but can hinder performance on images from different fields, different time or different robot. Custom datasets can mitigate these effects but present tradeoffs: the Early 2023 dataset, with more images but fewer weed labels, reduces crop misclassification but struggles with weed detection. Conversely, the smaller, balanced Refined 2023 dataset improves weed segmentation, including smaller leaves, at the cost of increased false positives and crop confusion.

In contrast, FSS models like HDMNet, as also highlighted by Catalano et al. (2024), struggles with small, scattered weeds. HDMNet performance is further affected by variations in soil, lighting, and rotation between query and support images. Pretraining on the MS COCO generic

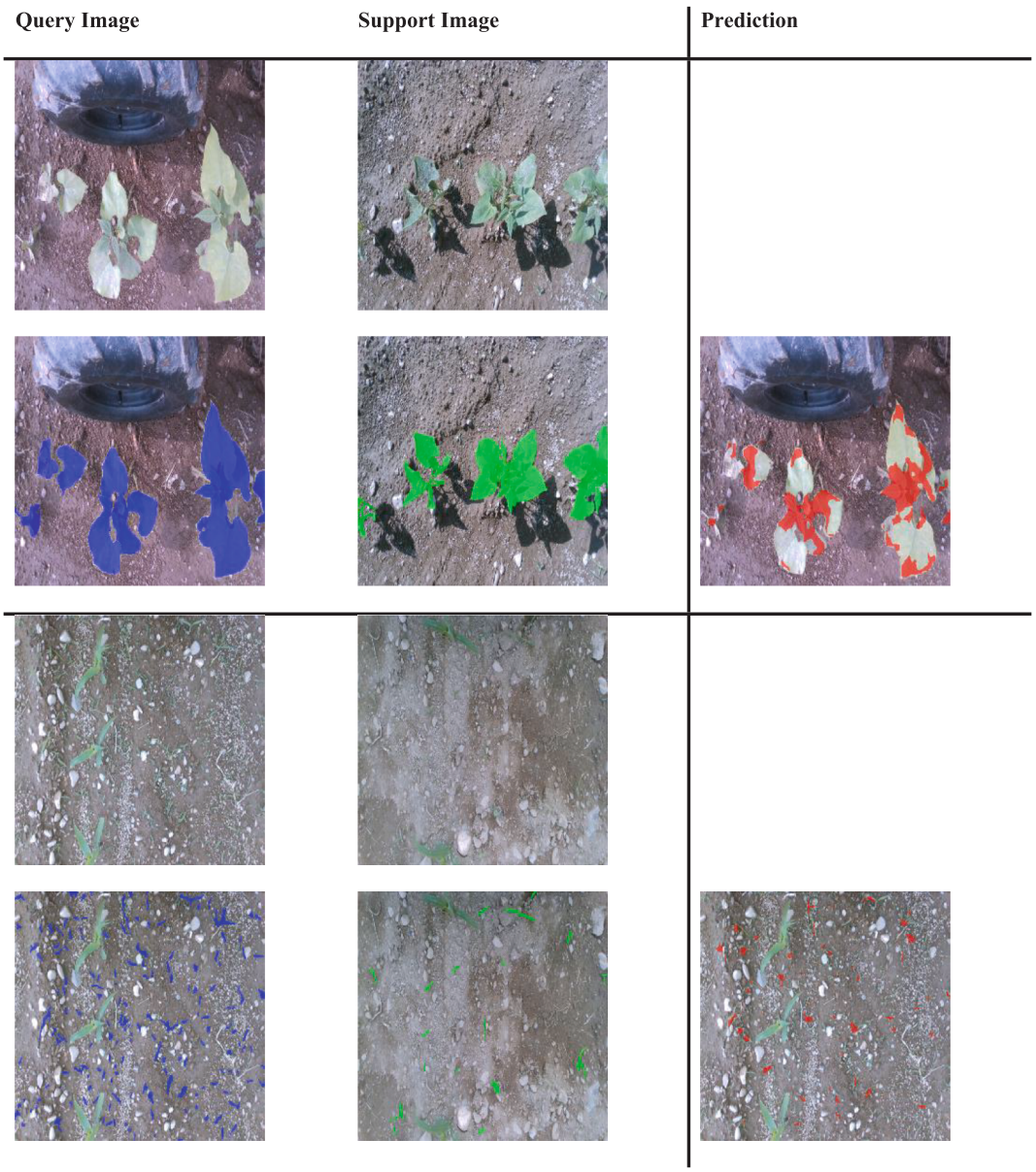


Fig. 10. Failure examples of HDMNet trained on ACRE 2019 on images of beans (top group) and corn (bottom group) cultivation. For both the support and query image we report the ground truth and label in the second row of teach group.

datset results in high error rates, while pretraining on the agricultural ACRE 2019 dataset partially mitigates these issues, reducing false positives and improving crop segmentation.

Overall, YoloV5 and YoloV8 require tailored datasets to optimize performance, while HDMNet benefits significantly from faster adaptation capabilities but requiring the support example to closely resemble the query image.

6. Discussion

In the remainder of this section, we provide a nuanced review of the experimental results that have been just presented, drawing connections between quantitative and qualitative outcomes from the performance of different instance segmentation models. Building upon the insights presented in the previous sections, our analysis focuses on key findings highlighted in Tables 5, 6, 7 and 8 that are indicative of the interplay between dataset characteristics, human labor investment, model efficacy, and pragmatic aspects that are required for the development of the autonomous weeding solutions in field.

In Table 5, we present the performance comparison between YOLO and HDMNet on two distinct datasets: the coarser yet larger Early dataset and the more finely annotated Refined dataset. One initial observation that can be drawn is the consistently higher IoU scores achieved with YOLO, particularly in the case where the model is trained on the Early dataset and tested specifically on the detection of crop categories. However, after analyzing more thoroughly the default evaluation procedure provided with YOLO models, we have identified a potential limitation in the ability of these models of capturing small objects, such as weed leaves, due to downsampling (Section 3.1). Namely, the downsampling process may lead to small object masks disappearing from the ground truth, introducing bias in IoU computation. Indeed, Table 6 demonstrates that refraining from downsampling the ground truth, by introducing the custom evaluation procedure detailed in Section 4.1, overall, generated lower IoU metrics for the YOLO predictions. This performance drop is especially noticeable in datasets that include a higher number of small leaf annotations, like the Refined dataset.

Under the fairer comparison enabled by this custom evaluation

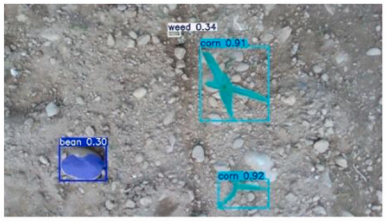
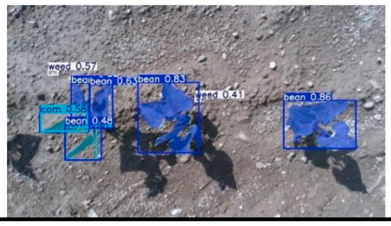
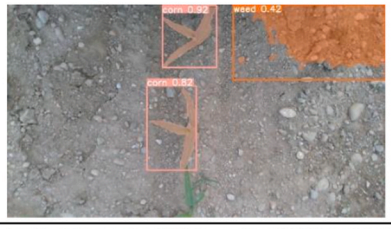
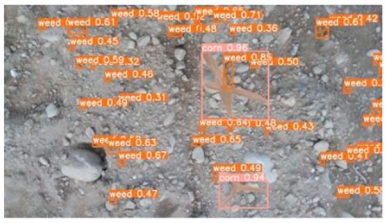
Model	Training Set	Prediction
YoloV5	ACRE2019	
YoloV5	Early 2023	
YoloV5	Refined 2023	
YoloV8	ACRE 2019	
YoloV8	Early 2023	
YoloV8	Refined 2023	

Fig. 11. Failure examples of YoloV5 and YoloV8 trained on ACRE 2019, Early 2023 and Refined 2023.

procedure, the FSS model HDMNet achieves a comparable segmentation performance to YOLO for crop classes, as evidenced by both the quantitative (Section 5.1) and qualitative results (Section 5.2). Notably, even without class-specific training, pre-training HDMNet on COCO was still sufficient to learn to segment large structures like bean leaves with as little as one or five labeled examples provided as support set. After tailoring the training on weed and crop segmentation for bean and corn cultivations, HDMNet even surpasses YOLO in performance on the same dataset. This result emphasizes the importance of domain-specific, targeted training for optimal outcomes, even in the lack of abundant training examples.

Moreover, our results highlight the difficulty both models face in predicting the weed category, regardless of training on the Early or Refined dataset. As shown in Table 8, curating a more densely annotated dataset while also maintaining a comparable manual effort in terms of human hours drastically reduces the total number of labeled images that can be produced. Reducing the size of the training set, overall, can negatively impact the performance of the models without introducing any appreciable improvement on the segmentation of small-sized regions. Consequently, our findings suggest that, when preparing datasets for training YOLO-based models on a new application domain, one should prioritize the dataset size over the annotation granularity, particularly if the application does not specifically focus on small-sized subjects. This is an important contribution for guiding optimal resource

allocation in future dataset curation efforts in this field. YOLO models necessitate extensive human intervention and are subjected to substantial costs, regardless of whether the dataset is extensive or highly refined. As indicated by Table 8 data, the process of refining a dataset notably increases the time required for manual labeling, consequently elevating the associated operational costs. Specifically, YOLO requires approximately 180 h for dataset preparation, and this time could increase further depending on the level of detail required.

Conversely, FSS models offers a more efficient alternative, as it significantly reduces the need for extensive manual labeling, thereby requiring only a few labeled pictures for the process. This not only minimizes the allocation of human resources but also eliminates the dataset preparation and training time altogether. Although FSS may exhibit slightly lower performance compared to conventional models like YOLO, its reduced dependency on large labeled datasets results in significantly lower operational costs. This positions FSS as a cost-effective approach for image segmentation, particularly in scenarios where resource optimization and budget constraints are critical considerations.

The inherent challenges of detecting small weeds with deep learning models is a known problem in the literature (López-Correa et al., 2022). However, the clear recognition of small weeds is often times crucial for farming success and crop yield enhancement (Corceiro et al., 2023). Indeed, detecting the presence of weeds since their earliest growth stages enables the application of essential treatments, targeted at reducing crop competition, at seed reduction and at minimizing herbicide use, thus producing higher yields and longer-term economic benefits. Ideally, autonomous weed control systems should then enable a prompt intervention on target weeds, and low-cost, small form-factor robots are especially suited for this task, as emphasized in a few related works (Wu et al., 2020, Gerhards et al., 2021).

Nonetheless is important to keep in mind that, although the early identification and treatment of weed is an effective method for ensuring crop yield, a later control may be preferable or more practical in certain situations. Weed growth habits, weather conditions and equipment availability can influence the timing of weed management activities. These practical constraints can considerably reduce the negative impact of the inability to detect small weed patches on the overall weed management process. Therefore, models like those tested in this work, despite their lower performance on smaller weeds, can still provide a valuable and cost-effective solution for bootstrapping autonomous weeding systems in field. This not only lowers labor costs but also enables scaling up labor-intensive crop production areas.

FSS methods like HDMNet are particularly promising candidates for enabling the shift towards autonomous weeding systems. The minimal effort required for the adaptation of FSS models, as proven in this paper, is particularly crucial in the agricultural context, which is characterized by a high variability due to the presence of diverse crops, weed species, camera angles, and plant growth stages. The baseline performance showcased by HDMNet, even with minimal target examples to learn from, makes these models a compelling alternative to more traditional approaches like YOLO, which require expensive training procedures to be updated.

7. Conclusions

In this paper, we have presented a comprehensive analysis of different YOLO models and the HDMNet model, which at the time of writing provides the state-of-the-art for FSS on the MS COCO dataset. Our results show that finely annotated datasets are preferable for precise small-leaf segmentation, while larger but coarser datasets remain advantageous when small-leaf detection is less critical. We further demonstrated that HDMNet can match or surpass YOLO in scenarios with leaves of substantial object sizes, requiring only one to five labeled examples and thereby reducing annotation time, considered as human effort, by orders of magnitude. These findings highlight the potential of FSS for agricultural robotics, where frequent re-annotation is needed due to varying lighting, terrain, and plant species. By enabling rapid adaptation to new field conditions, FSS models offer a promising path toward more flexible and cost-effective weed and crop segmentation systems.

Our experiments on the 2019 and 2023 ACRE datasets, diverse in regions, species, platforms, and field conditions, captured weed samples at comparable growth stages. A valuable direction for future work is to track how segmentation models generalize over time and across different weed growth stages. Beyond longitudinal monitoring, expanding the analysis to a broader range of crop and weed species would further strengthen our understanding of model robustness. Overall, autonomous weeding remains a multifaceted challenge, requiring not only advances in perception but also the integration of broader robotic capabilities to achieve precise and scalable weed control in real-world agricultural environments.

CRediT authorship contribution statement

Nico Catalano: Writing – original draft, Methodology, Investigation. **Sofia Matilde Luglio:** Writing – original draft, Methodology, Investigation. **Agnese Chiatti:** Writing – original draft, Methodology, Investigation. **Mino Sportelli:** Writing – original draft, Methodology, Investigation. **Christian Frasconi:** Supervision, Funding acquisition, Conceptualization. **Davide Facchinetti:** Resources, Funding acquisition, Conceptualization. **Matteo Matteucci:** Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

AgrifoodTEF: This project has received co-funding from the European Union's Digital Europe Programme under grant agreement N° 101100622.

Data availability

Data will be made available on request.

References

Karunathilake, E.M.B.M., Le, A.T., Heo, S., Chung, Y.S., Mansoor, S., 2023. The path to smart farming: innovations and opportunities in precision agriculture. In Agriculture (Switzerland) (Vol. 13, Issue 8). Multidisciplinary Digital Publishing Institute (MDPI). Doi: 10.3390/agriculture13081593.

Tittonell, P., Piñeiro, G., Garibaldi, L.A., Dogliotti, S., Olff, H., Jobbagy, E.G., 2020. Agroecology in large scale farming—a research agenda. Front. Sustainable Food Syst. 4. https://doi.org/10.3389/fsufs.2020.584605.

Finger, R., Swinton, S.M., el Benni, N., Walter, A., 2019. precision farming at the nexus of agricultural production and the environment. Doi: 10.1146/annurev-resource-100518.

Trang, B., Hoa, P.T., Kawarazuka, N., Schreinemacher, P., Liu, Y., 2022. Introducing an agricultural app to vegetable farmers: a pilot study in Lam Dong, Vietnam. Doi: 10.4160/9789290606376.

Sharma, A., Georgi, M., Tregubenko, M., Tselykh, A., Tselykh, A., 2022. Enabling smart agriculture by implementing artificial intelligence and embedded sensing. Comput. Ind. Eng. 165. https://doi.org/10.1016/j.cie.2022.107936.

Blasch, J., van der Kroon, B., van Beukering, P., Munster, R., Fabiani, S., Nino, P., Vanino, S., 2022. Farmer preferences for adopting precision farming technologies: a case study from Italy. Eur. Rev. Agric. Econ. 49 (1), 33–81. https://doi.org/10.1093/erae/jbaa031.

Ragu, N., Teo, J., 2023. Object detection and classification using few-shot learning in smart agriculture: a scoping mini review. Front. Sustain. Food Syst. 6, 1039299. https://doi.org/10.3389/fsufs.2022.1039299.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. ACM Trans. Math. Software 60 (6), 84–90. https://doi.org/10.1145/3065386.

Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B., 2019. A closer look at few-shot classification. arXiv preprint arXiv:1904.04232.
Abdalla, A., Cen, H., Wan, L., Rashid, R., Weng, H., Zhou, W., He, Y., 2019. Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure. Comput. Electron. Agric. 167, 105091. https://doi.org/10.1016/j.compag.2019.105091.
Gidaris, S., Komodakis, N., 2018. Dynamic few-shot visual learning without forgetting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4367–4375.
Karunathilake, E.M.B.M., Le, A.T., Heo, S., Chung, Y.S., Mansoor, S., 2023. The path to smart farming: innovations and opportunities in precision agriculture. In Agriculture (Switzerland) (Vol. 13, Issue 8). Multidisciplinary Digital Publishing Institute (MDPI). Doi: 10.3390/agriculture13081593.
Nurcahyo, A., Soeparno, H., Gaol, F.L., Arifin, Y., 2023. Developing smart precision farming using big data and cloud-based intelligent decision support system. 1–6. Doi: 10.1109/iciss59129.2023.10291960.
Tataridas, A., Kanatas, P., Chatzigeorgiou, A., Zannopoulos, S., Travlos, I., 2022. Sustainable crop and weed management in the era of the EU green deal: a survival guide. Agronomy, 12(3). Doi: 10.3390/agronomy12030589.
Gallo, I., Rehman, A.U., Dehkordi, R.H., Landro, N., la Grassa, R., Boschetti, M., 2023. Deep object detection of crop weeds: performance of YOLOv7 on a real case dataset from UAV images. Remote Sens. (Basel) 15 (2). https://doi.org/10.3390/rs15020539.
Miloto A., Lottes P., Stachniss C., 2018. Institute of Electrical and Electronics Engineers. Proceedings of IEEE International Conference on Robotics and Automation (ICRA): May 21-25, Brisbane, Australia.
He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. http://arxiv.org/abs/1703.06870.
Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2016. Pyramid scene parsing network. http://arxiv.org/abs/1612.01105.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified. Real-Time Object Detection. http://pjreddie.com/yolo/.
Peng, B., Tian, Z., Wu, X., Wang, C., Liu, S., Su, J., Jia, J., 2023. Hierarchical dense correlation distillation for few-shot segmentation. .
Weyler, J., Läbe, T., Magistri, F., Behley, J., Stachniss, C., 2023. Towards domain generalization in crop and weed segmentation for precision farming robots. https://github.com/PRBonn/DG-CWS.
Zou, K., Chen, X., Wang, Y., Zhang, C., Zhang, F., 2021. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. Comput. Electr. Agric. 187. https://doi.org/10.1016/j.compag.2021.106242.
Su, D., Kong, H., Qiao, Y., Sukkarieh, S., 2021. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. Comput. Electr. Agric. 190. https://doi.org/10.1016/j.compag.2021.106418.
Nesteruk, S., Shadrin, D., Pukalchik, M., 2021. Image augmentation for multitask few-shot learning: agricultural domain use-case. http://arxiv.org/abs/2102.12295.
Guldenring, R., Boukas, E., Ravn, O., Nalpantidis, L., 2021. Few-leaf learning: weed segmentation in grasslands. IEEE Int. Conf. Intell. Robots Syst. 3248–3254. https://doi.org/10.1109/IROS51168.2021.9636770.
Bertoglio, R., Mazzucchelli, A., Catalano, N., Matteucci, M., 2023. A comparative study of Fourier transform and CycleGAN as domain adaptation techniques for weed segmentation. Smart Agric. Technol. 4. https://doi.org/10.1016/j.atech.2023.100188.
Magistri, F., Weyler, J., Gogoll, D., Lottes, P., Behley, J., Petrinic, N., Stachniss, C., 2023. From one field to another—Unsupervised domain adaptation for semantic segmentation in agricultural robotics. Comput. Electr. Agric. 212. https://doi.org/10.1016/j.compag.2023.108114.
Chiatti, A., Bertoglio, R., Catalano, N., Gatti, M., Matteucci, M., 2022. Surgical fine-tuning for grape bunch segmentation under visual domain shifts. Doi: 10.5281/zenodo.7866442.
Luglio S.M., Sportelli M., Frasconi C., Fontanelli M., Matteucci M., Fontana G., Piazza E., Facchinetti D., 2023. Proceedings of 2023 IEEE International workshop on metrology for agriculture and forestry, Pisa.
Peruzzi, A., Sartori, L., 1997. Guida Alla Scelta ed All'impiego Delle Attrezzature per la Lavorazione del Terreno (Guide for Selecting and Using Soil Tillage Tools); Edagricole: Bologna, Italy, pp. 1–236.
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything. http://arxiv.org/abs/2304.02643.
Bertoglio, R., Fontana, G., Matteucci, M., Facchinetti, D., Berducat, M., Boffety, D., 2021. On the design of the agri-food competition for robot evaluation (ACRE). In: 2021 IEEE International Conference on Autonomous Robot Systems and Competitions. https://doi.org/10.1109/ICARSC52212.2021.9429792.
Shaban, A., Bansal, S., Liu, Z., Essa, I., Boots, B., 2017. One-shot learning for semantic segmentation. British Machine Vision Conference 2017, BMVC 2017. Doi: 10.5244/c.31.167.
Rakelly, K., Shelhamer, E., Darrell, T., Efros, A., Levine, S., 2018. Workshop track-ICLR 2018 conditional networks for few-shot semantic segmentation.
Zhang, C., Lin, G., Liu, F., Yao, R., Shen, C., 2019. CANET: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2019.00536.
Boudiaf, M., Kervadec, H., Masud, Z.I., Piantanida, P., Ayed, I. ben, Dolz, J., Hoel, M., Kervadeć, K., Montreal, K., Imtiaz, Z., Masud, M., Montreal, M., Ben, I., Ayed, A., Montreal, A., Dolź, J.D., Montreal, D., 2021. Few-shot segmentation without meta-learning: a good transductive inference is all you need. 13974–13983. Doi: 10.1109/CVPR46437.2021.01376ï.
Tian, Z., Zhao, H., Shu, M., Yang, Z., Li, R., Jia, J., 2022. Prior guided feature enrichment network for few-shot segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 44 (2), 1050–1065. https://doi.org/10.1109/TPAMI.2020.3013717.
Wu, Z., Shi, X., Lin, G., Cai, J., 2021. Learning Meta-class Memory for Few-Shot Semantic Segmentation.
Nguyen, K., Todorovic, S., 2019. Feature weighting and boosting for few-shot segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2019.00071.
Li, G., Jampani, V., Sevilla-Lara, L., Sun, D., Kim, J., Kim, J., 2021a. Adaptive prototype learning and allocation for few-shot segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 8330–8339. https://doi.org/10.1109/CVPR46437.2021.00823.
Lu, Z., He, S., Zhu, X., Zhang, L., Song, Y.-Z., Xiang, T., 2022. Simpler is better: few-shot semantic segmentation with classifier weight transformer. 8721–8730. Doi: 10.1109/iccv48922.2021.00862.
Siam, M., Oreshkin, B., 2019. Adaptive masked weight imprinting for few-shot segmentation. 2019 the International Conference on Learning Representations.
Wang, K., Liew, J. H., Zou, Y., Zhou, D., Feng, J., 2019. PANet: Few-shot image semantic segmentation with prototype alignment.
Dong, N., Xing, E.P., 2018. Few-shot semantic segmentation with prototype learning.
Vinyals, O., Deepmind, G., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D., 2016. Matching networks for one shot learning. 2016 Conference on Neural Information Processing Systems.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.
Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2019. YOLACT: real-time instance segmentation. .
Jocher, G., Chaurasia, A., Qiu, J., 2023 YOLO by Ultralytics. https://github.com/ultralytics/ultralytics.
López-Correa, J.M., Moreno, H., Ribeiro, A., Andújar, D., 2022. Intelligent weed management based on object detection neural networks in tomato crops. Agronomy 12 (12). https://doi.org/10.3390/agronomy12122953.
Corceiro, A., Alibabaei, K., Assunção, E., Gaspar, P.D., Pereira, N., 2023. Methods for detecting and classifying weeds, diseases and fruits using ai to improve the sustainability of agricultural crops: a review. Processes 11 (4), MDPI. https://doi.org/10.3390/pr11041263.
Wu, X., Aravecchia, S., Lottes, P., Stachniss, C., Pradalier, C., 2020. Robotic weed control using automated weed and crop classification. J. Field Robot. https://doi.org/10.1002/rob.21938.
Gerhards, R., Späth, M., Sökefeld, M., Peteinatos, G.G., Nabout, A., Rueda Ayala, V., 2021. Automatic adjustment of harrowing intensity in cereals using digital image analysis. Weed Res. 61 (1), 68–77. https://doi.org/10.1111/wre.12458.
Yang, J., Guo, X., Li, Y., et al., 2022. A survey of few-shot learning in smart agriculture: developments, applications, and challenges. Plant Methods 18, 28. https://doi.org/10.1186/s13007-022-00866-2.
Sun, J., et al., 2024. Few-shot learning for plant disease recognition: a review. Agron. J. 116 (3), 1204–1216. https://doi.org/10.1002/agj2.21285.
Liang, X., 2021. Few-shot cotton leaf spots disease classification based on metric learning. Plant Methods 17, 114. https://doi.org/10.1186/s13007-021-00813-7.
Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M.G., Bereciartua, A., Alvarez-Gila, A., 2020. Few-Shot Learn-ing approach for plant disease classification using images taken inthe field. Comput. Electr. Griculture 175, 105542. https://doi.org/10.1016/j.compag.2020.105542.
Zhong, F., et al., 2020. Zero-and few-shot learning for diseases recognition of *Citrus aurantium* L. using conditional adversarial autoencoders. In: Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2020.105828.
Wang, C., et al., 2021. Few-shot vegetable disease recognition model based on image text collaborative representation learning. Comput. Electr. Agric., 184 (2021): 106098. Doi: 10.1016/j.compag.2021.106098.
Hughes, D., Salathé, M., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv: 1511.08060, Doi: 10.48550/arXiv.1511.08060.
Zhang, D.i., et al., 2021. Seeding crop detection framework using prototypical network method in UAV images. Agriculture 12 (1), 26. https://doi.org/10.3390/agriculture12010026.
Li, L., et al., 2021. Maize residue segmentation using Siamese domain transfer network. Comput. Electr. Agric., 187, 106261. Doi: 10.1016/j.compag.2021.106261.
Pearson, S., Camacho-Villa, T.C., Valluru, R., et al., 2022. Robotics and autonomous systems for net zero agriculture. Curr. Robot Rep. 3, 57–64. https://doi.org/10.1007/s43154-022-00077-6.
Nie, J., Yuan, Y., Li, Y., Wang, H., Li, J., Wang, Y., Song, K., Ercisli, S., 2024. Few-shot learning in intelligent agriculture: a review of methods and applications. J. Agr. Sci.-Tarim Bili. 30 (2), 216–228. https://doi.org/10.15832/ankutbd.1339516.
Liu, D., Wang, B., Peng, L., Wang, H., Wang, Y., Pan, Y., 2024. HSDNet: a poultry farming model based on few-shot semantic segmentation addressing non-smooth and unbalanced convergence. PeerJ Computer Sci. 10, e2080.
Nuthalapati, S.V., Tunga, A., 2021. Multi-domain few-shot learning and dataset for agricultural applications, 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, pp. 1399-1408, Doi: 10.1109/ICCVW54120.2021.00161.
Wang, S., Han, Y., Chen, J., He, X., Zhang, Z., Liu, X., Zhang, K., 2022. Weed density extraction based on few-shot learning through UAV remote sensing RGB and multispectral images in ecological irrigation area. Front. Plant Sci. 12, 735230. https://doi.org/10.3389/fpls.2021.735230.